IMPACT OF FOREIGN MARKET AND CURRENCY EXCHANGE RATE ON INDIAN IT STOCKS

by

Rakesh Kumar Sharma

DISSERTATION

Presented to the Swiss School of Business and Management Geneva

In Partial Fulfillment

Of the Requirements

For the Degree

DOCTOR OF BUSINESS ADMINISTRATION

SWISS SCHOOL OF BUSINESS AND MANAGEMENT GENEVA

MARCH 2023

IMPACT OF FOREIGN MARKET AND CURRENCY EXCHANGE RATE ON INDIAN IT STOCKS

by

Rakesh Kumar Sharma

APPROVED BY

____

<Chair’s Name, Degree>, Chair

Dr. David Annan PhD.

____

<Member’s Name, Degree>, Committee Member

Dr. Olesya Meskina PhD.

____

<Member’s Name, Degree>, Committee Member

RECEIVED/APPROVED BY:

____

SSBM Representative

Acknowledgements

I would like to express my heartfelt gratitude to all those who have contributed to the successful completion of this thesis.

First and foremost, I would like to thank my advisor, PROF DAVID ANNAN, for his unwavering support, encouragement, and guidance throughout my research journey. His expertise, knowledge, and insights have been invaluable in shaping the direction and quality of my work.

I am grateful to my family and friends for their unwavering support, love, and encouragement throughout my academic journey. Their belief in me and their constant encouragement have been a source of strength and motivation.

Finally, I would like to thank UPGRAD for providing me with the resources, facilities, and opportunities to pursue my academic interests and cultivate my intellectual curiosity. I am proud to be a part of such a distinguished institution and grateful for the opportunities it has provided me.

To all those who have contributed in one way or another to the completion of this thesis, I extend my heartfelt thanks and appreciation. Your support, encouragement, and guidance have been instrumental in shaping my academic journey and the completion of this thesis.

ABSTRACT

IMPACT OF FOREIGN MARKET AND CURRENCY EXCHANGE RATE ON INDIAN IT STOCKS

The stock market plays a crucial role in the economy of every nation. With globalization, the world's economies have become interlinked through trade, foreign investment, capital flow, and technological advancements. Many studies have examined the influence of foreign stock exchanges, currency exchange rates, and economic articles on the Indian stock market using only the Granger causality method without the use of different approaches such as machine learning.

Hence, this study takes a more robust approach by incorporating ensemble methods of machine-learning, deep-learning-based, and statistical methods and using majority voting to examine the impact of foreign IT stock markets, currency exchange rates, and economic articles on the Indian IT stock market.

The results indicate that foreign IT stock markets such as France, the United States of America, Britain, and Germany have a positive impact on the Indian IT stock market. As such, an increase in the USD_INR, JPY_INR, and SGD_INR currency exchange rates increases in the Indian IT stock market, whereas an increase in the MUR_INR rate corresponds with a decrease. Positive sentiments reflected in articles from The Hindu and The Financial Times correspond with an increase in the Indian IT stock market, while other representations of economic articles correspond with a decrease.

The study's findings could be valuable for financial markets as they might fill the gap in understanding strategies to increase their IT stock indices and provide an-in-depth insights

into developing different trajectories in the stock market. Moreover, the study will also assist portfolio managers and investment management organizations in developing more profitable investment strategies.

TABLE OF CONTENTS

List of Tables viii

List of Figures x

List of EQUATIONS xiii

CHAPTER I: INTRODUCTION 1

1.1 Introduction 1
1.2 Problem statement 3
1.3 Purpose of Research 4
1.4 Significance of Research 5
1.5 Research Question/Hypothesis 6
1.6 Summary 7

CHAPTER II: REVIEW OF LITERATURE 8

2.1 Introduction 8
2.2 Statistical causality detection algorithms 9
2.3 Deep Learning-based causality detection algorithms 12
2.4 Statistical methods to deep-learning methods 15
2.5 Impact of the foreign stock exchange on the Indian stock market 16
2.6 Impact of the News Articles on the Indian Stock Market 23
2.7 Impact of the Currency Exchange Rate on the Indian Stock Market 26
2.8 Miscellaneous 29
2.9 Conclusion 31
2.10 Summary 32

CHAPTER III: METHODOLOGY 34

3.1 Introduction 34
3.2 Methodology approach 34
3.3 Research Philosophy 36
3.4 Research Approach 37
3.5 Research Strategy 38
3.6 Data Collection 38
3.7 Methods of data analysis 50
3.8 Methods 57
3.9 Justification for methodology 95
3.10 Validity 96
3.11 Reliability 97
3.12 Research design limitation 97

3.13 Summary 98

CHAPTER IV: RESULTS.. 99

4.1 Introduction.. 99
4.2 Sample for analysis 99
4.3 Bivariate analysis 99
4.4 Pearson correlation.. 114
4.5 Spearman correlation 118
4.6 NonLinCausality 121
4.7 DoWhy.. 138
4.8 Temporal Causal Discovery Framework 151
4.9 Granger Causality 153
4.10 Summary 161

CHAPTER V: DISCUSSION.. 163

5.1 Introduction.. 163
5.2 Evaluating the Research Objective. 163
5.3 Research objective 1 – Examine the impact of the foreign IT stock market on the Indian IT stock market using ensemble methods. 164
5.4 Research objective 2 – Examine the impact of significant currency exchange rates on the Indian IT stock market using ensemble methods. 172
5.5 Research objective 3 – Examine the impact of text features and sentiments of economic articles on the Indian IT stock market using ensemble methods. 180
5.6 Limitation and Future work 198
5.7 Conclusion 200

APPENDIX A DATASET AND CODE USED FOR STUDY.. 202

REFERENCES 203

LIST OF TABLES

Table 3.1 The data source for foreign IT indices 42
Table 3.2 Data sources for currency exchange rates 44
Table 3.3 Data sources for economic articles 45
Table 3.4 Data description for foreign IT indices 47
Table 3.5 Data description for the currency exchange rate 47
Table 3.7 Data description for named entities. 49
Table 3.8 Independent variable that represents the economic articles. 52
Table 4.1 Data description for the currency exchange rate. 115
Table 4.2 Data description for the currency exchange rate. 118
Table 4.3 Most impacted variables considering ten days lag. 122
Table 4.4 Most impacted variables considering 20 days lag. 124
Table 4.5 Most impacted variables considering 30 days lag. 125
Table 4.6 Most impacted variables summary for GRU nonLinCausality 128
Table 4.7 Most impacted variables considering ten days lag for LSTM nonLinCausality. 130
Table 4.8 Most impacted variables considering 20 days lag for LSTM nonLinCausality. 132
Table 4.9 Most impacted variables considering 30 days lag for LSTM nonLinCausality. 134
Table 4.10 Most impacted variables summary for LSTM nonLinCausality 137
Table 4.11 RCC refutation results 139
Table 4.12 PTR refutation results 142
Table 4.13 DSR refutation results 146
Table 4.14 Summary of all refutation methods 148
Table 4.15 Results of TCDF with time-lag 152
Table 4.16 Most impacted variables considering ten days lag for Granger Causality 154
Table 4.17 Most impacted variables considering 20 days lag for Granger Causality 155
Table 4.18 Most impacted variables considering 30 days lag for Granger Causality 157

Table 4.19 Most impacted variables summary for Granger Causality 160
Table 5.1 Most impacted foreign IT variables using DoWhy 165
Table 5.2 Most impacted foreign IT variables using nonLinCausality (GRU) 167
Table 5.3 Most impacted foreign IT variables using nonLinCausality (LSTM) 168
Table 5.4 Most impacted foreign IT variables using Granger Causality Test (GRU) 169
Table 5.5 Most impacted foreign IT variables using ensemble methods (GRU) 171
Table 5.6 Most impacted currency exchange variables using DoWhy........................... 174
Table 5.7 Most impacted currency exchange variables using nonLinCausality (GRU). 175
Table 5.8 Most impacted currency exchange variables using Granger Causality. 177
Table 5.9 Most impacted currency exchange variables using ensemble methods.......... 179
Table 5.10 Most impacted economic article variables using doWhy. 181
Table 5.11 Most impacted economic article variables using nonLinCausality(GRU)... 185
Table 5.12 Most impacted economic article variables using nonLinCausality (LSTM). 187
Table 5.13 Most impacted economic article variables using Granger Causality............ 190
Table 5.14 Most impacted economic article variables using ensemble methods. 194

LIST OF FIGURES

Figure 2.1 Granger causality visualization 10
Figure 3.1 Causal Analysis Framework 35
Figure 3.2 Missing value imputation 54
Figure 3.3 Scaler method 55
Figure 3.4 Perceptron 58
Figure 3.5. Step function 59
Figure 3.6 MLP architecture 60
Figure 3.7 MLP implementation using nonlincausality framework. 60
Figure 3.8 LSTM architecture 62
Figure 3.9 LSTM implementation using nonlincausality framework. 65
Figure 3.10 GRU architecture 67
Figure 3.11 GRU implementation using nonlincausality framework. 68
Figure 3.12 Binary step function 71
Figure 3.13 Linear activation function 72
Figure 3.14 Sigmoid activation function 73
Figure 3.15 TanH activation function 74
Figure 3.16 ReLu activation function 75
Figure 3.17 LeakyReLu activation function 77
Figure 3.18 Parametric ReLu activation function 78
Figure 3.19 DoWhy Library framework 81
Figure 3.20 DoWhy implementation 87
Figure 3.21 Convolution operation 90
Figure 3.22 Average pooling 91
Figure 3.23 Max pooling 91
Figure 3.24 Flatten Layer 92
Figure 3.25 Attention mechanism 93
Figure 3.26 TCDF implementation 93
Figure 3.27 Granger Causality implementation 95
Figure 4.1 Foreign stock market vs. Indian IT stock 101

Figure 4.2 Currency exchange rate vs. Indian IT stock 103
Figure 4.3 Money control named entities vs. Indian IT stock. 105
Figure 4.4 Money control sentiments vs. Indian IT stock 106
Figure 4.5 The Hindu named entities vs. Indian IT stock. 109
Figure 4.6 The Hindu sentiments vs. Indian IT stock 110
Figure 4.7 Financial news sentiments vs. Indian IT stock 112
Figure 4.8 Financial named entities vs. Indian IT stock. 113
Figure 4.9 Independent Variables vs. Vote count for nonLinCausality GRU 129
Figure 4.10 Independent Variables vs. Vote count for nonLinCausality LSTM 138
Figure 4.11 Independent Variables vs. Vote count for DoWhy's refutation methods 151
Figure 4.12 Independent Variables vs. Vote count for Granger Causality methods 161
Figure 5.1 Most impacted foreign IT variables vs. Vote count for DoWhy. 166
Figure 5.2 Most impacted foreign IT variables vs. Vote count for nonLinCausality (GRU). 168
Figure 5.3 Most impacted foreign IT variables vs. Vote count for nonLinCausality (LSTM). 168
Figure 5.4 Most impacted foreign IT variables vs. Vote count for Granger Causality. 170
Figure 5.5 Most impacted foreign IT variables vs. Vote count for ensemble methods. 172
Figure 5.6 Most impacted currency exchange variables vs. Vote count for doWhy. 174
Figure 5.7 Most impacted currency exchange variables vs. Vote count for nonLinCausality (GRU). 176
Figure 5.8 Most impacted currency exchange variables vs. Vote count for Granger causality. 177
Figure 5.9 Most impacted currency exchange variables vs. Vote count for ensemble methods. 179
Figure 5.10 Most impacted economic articles variables vs. Vote count for doWhy. 183
Figure 5.11 Most impacted economic articles variables vs. Vote count for nonLinCausality (GRU). 187
Figure 5.12 Most impacted economic articles variables vs. Vote count for nonLinCausality (LSTM) 189

Figure 5.13 Most impacted economic articles variables vs. Vote count for Granger Causality. 192

Figure 5.14 Most impacted economic articles variables vs. Vote count for ensemble methods. 196

LIST OF EQUATIONS

Equation 3.1 Step function 59
Equation 3.2 LSTM's input gate 63
Equation 3.3 LSTM's output of the input gate 63
Equation 3.4 LSTM's input gate element-wise multiplication 64
Equation 3.5 LSTMs need to remember gate output. 64
Equation 3.6. LSTMs forget gate and input gate multiplication. 64
Equation 3.7 LSTM’s output gate 64
Equation 3.8 LSTM’s final output 65
Equation 3.9 GRU’s update gate. 67
Equation 3.10 GRU’s reset gate. 68
Equation 3.11 GRU’s candidate state 68
Equation 3.12 GRU’s final output 68
Equation 3.13 Binary step function 70
Equation 3.14 Linear activation function. 71
Equation 3.15 Sigmoid activation function 73
Equation 3.16 tanH activation function 74
Equation 3.17. ReLu activation function 75
Equation 3.18 Leaky ReLu activation function 76
Equation 3.19 Parametric ReLu activation function 77
Equation 3.20 Softmax activation function 78

CHAPTER I:

INTRODUCTION

1.1 Introduction

A stock market is a collection of markets and exchanges where shares of publicly traded companies are frequently bought, sold, and issued (CHEN, 2021). The Stock market plays a prominent role in every nation's economy and is no longer dependent upon national factors. With globalization, the world's economies have become increasingly interdependent. International economies integrate with national economies through trade, foreign direct investment, capital flow, and technological advancement. As a result, changes in the foreign stock market have ripple effects on the national stock market, and India is no exception (Goel & Gupta, 2011).

Indian IT companies are also sensitive to currency exchange rates due to the Leading IT companies maintaining various hedging policies, which is one of the critical factors in determining gain from the falling Rupee (Indian currency). The other factor is offshoring – more offshoring generates excess profit from currency depreciation (BusinessToday, 2018). Hence, whenever the rupee weakens against foreign currency, these companies gain extra, and vice-versa.

Kwatra (2018) indicated that the stock market experiences short-term volatility (rise and fall) driven by news articles in foreign markets. Positive news triggers the rise, and negative news triggers the fall in the stock market. The effect of foreign markets (foreign indices and currency exchange rates) and news are mainly studied using Granger causality detection methods (Kishor & Singh, 2017; A. Kumar, 2019). A Granger causality test is a statistical test determining if one data series can predict another (Granger, 1969). Data stationarity is a prerequisite for Granger causality. It also finds it challenging to detect causal relationships in multivariate time series under non-linear dynamics. Thus, these limitations of Granger causality encourage researchers to envision a better alternative based on state-of-the-art deep-learning networks. Hence, Tank et al. (2021) proposed deep learning-based neural Granger causality methods that efficiently capture the long-term dependencies between time series and can work on non-linear datasets. The method outperformed the state-of-the-art Granger causality test on non-linear gene expression data.

Similarly, Rosoł et al. (2022) proposed a causality analysis package based on deep learning. To perform causality tests, the authors posited that it supports deep-learning architectures such as Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), and Multilayer Perceptron (MLP). The study's metadata demonstrated through various experiments that the proposed package is more expressive than Granger causality approaches at detecting non-linear causal relationships.

Many studies have examined the impact of the international stock market, currency exchange rate, and economic news articles on the national stock exchange (Goel & Gupta, 2011; Kishor & Singh, 2017; A. Kumar, 2019). Most of this research has utilized the Granger causality test, which could better detect causality in non-linear datasets. Furthermore, studies such as Rosoł et al. (2022) and Tank et al. (2021) stated that deep-learning-based approaches outperformed the classic granger causality test. Thus, this study examined the impact of foreign stock markets, currency exchange rates, and economic articles on the Indian IT stock market using various techniques. These techniques include statistical-based methods, machine-learning-based methods, and deep-learning-based methods. This study aims to increase confidence in global stock exchange market analysis and its results using multiple techniques.

1.2 Problem statement

The stock market is highly volatile due to the global pandemic and economic shocks of nations engaging in conflicts and the continuous rises in product markets for goods and services. Identifying the factors that influence the stock market has long been a challenge. Various research authors, including Kishor & Singh (2017), Kumar (2019), and Srivastava & Sharma (2016), have conducted studies to identify the impact of foreign exchange on the Indian stock market. They have tried to uncover the influence of factors such as foreign stock market performance, currency exchange rates, and news article sentiments on the Indian IT stock market.

Most of these studies are too broad and have used statistical-based Granger causality tests to analyze the influence of a single factor on the Indian stock market. Marcinkevičs & Vogt (2021) and Tank et al. (2021) have shown that deep-learning-based causality detection methods efficiently infer multivariate causality in non-linear dynamics and long-term dependencies between time series, respectively. Hence, they are better alternatives to statistical-based Granger causality tests, and the study researcher will uncover these alternatives to ensure currency efficiency in the IT Stock market.

However, while conducting a preliminary literature review, not enough research has been encountered that has used deep-learning-based causality detection methods for impact analysis on the Indian stock market. Research indicates that fewer studies have analyzed the impact using statistical-based, machine-learning-based, and deep-learning-based algorithms. Hence, there is a substantial gap in reviewing the impact of factors such as the foreign market, currency exchange rate, and economic articles on the Indian stock market using statistical-based, machine-learning-based, and deep-learning-based causality methods. Moreover, India has positioned itself as the hub of information technology for the rest of the world. Thus, narrowing down the research to the information technology sector and considering named entities (a real-world object with a name, such as people, organizations, locations, medical codes, time expressions, quantities, monetary values, percentages, and so on (Honnibal, 2015)) and sentiments of economic articles may also yield better results.

As a result, this study will contribute knowledge to the field of research in academia and foreign market industries by examining the impact of foreign IT indices, exchange rates, and economic articles using multiple causality detection methods.

1.3 Purpose of Research

This study determines to analyze the impact of foreign IT indices, currency exchange rates, and text features and sentiments of economic articles using ensemble methods (ensemble of statistical-based, machine-learning-based, and deep-learning-based causality detection methods). The study's implication is to assist investors in identifying the most influencing factors on Indian IT stocks.

The research focus on and narrows its scope to address the following research objectives.

To examine data for foreign IT indices and analyze their impact on Indian IT stocks using ensemble methods.

To Identify data for effective currency exchange rates and analyze their impact on Indian IT stocks using ensemble methods.

To analyze the impact of text features and sentiments of economic indices on Indian IT stocks using ensemble methods.

1.4 Significance of Research

The extant research has yet to address the challenge of identifying the factors affecting the stock market. Several authors, including Joshi (2013); Srivastava and Sharma (2016); Kishor and Singh (2017); Kumar (2019); Deo and Prakash (2017); Paramanik and Singhal (2020); Kumar et al. (2020); Mohith and Sangeetha (2019); and Manu and Bhaskar (2018), have examined the impact of foreign indices, news articles, and exchange rates on the stock market.

Their research concludes that there is no clear way to single out one factor that impacts the stock market. Despite this, most of these studies have analyzed the impact of the above factors on the national stock market. Some studies like those (Sharma, 2020) have decently predicted the stock prices of Indian IT stocks using foreign indices, news articles, exchange rates, and technical indicators. Therefore, it is also worth narrowing the research down to a specific industry (e.g., IT) and using ensemble methods (ensemble of statistical-based, machine-learning-based, and deep-learning-based causality detection methods) for impact analysis. Therefore, this study analyzes the impact of foreign indices, economic items, and exchange rates on the Indian IT stock market using ensemble methods (ensemble of statistical-based, machine-learning-based, and deep-learning-based causality detection methods).

Identifying the most impactful characteristics of the Indian IT stock market will help individual investors invest their money wisely, which will bring more profit from the Indian IT stock market investment. Furthermore, the study will also assist portfolio managers and investment management organizations in developing more profitable investment strategies.

In addition, the idea of focusing the study on a specific sector, e.g., the IT sector, combined with the use of using ensemble methods (ensemble of statistical-based, machine-learning-based, and deep-learning-based causality detection methods) for causality detection, provides a basis for further investigations in other sectors of the stock market. These investigations would start by identifying the characteristics that could impact the particular sector with the help of a domain expert. Hence, reuse the techniques described in this study to identify the most impactful traits. For example, researchers in the pharmaceutical field could use their domain expertise to list all foreign market factors that may affect the pharmaceutical field. Then run the statistical-based, machine-learning-based, and deep learning-based approaches to confirm the most impactful features from the listed features.

1.5 Research Question/Hypothesis

The research aims to examine the impact of foreign IT indices, economic articles, and currency exchange rates on the Indian IT stock market using ensemble techniques (ensemble of statistical-based, machine-learning-based, and deep-learning-based causality detection methods). The following are the research questions and hypotheses addressed in this study.

Research question: How do foreign IT indices and currency exchange rates affect the Indian IT stock market?

Hypotheses: The rise/fall in foreign IT indices and currency exchange rates will also lead to a rise/fall in the Indian IT stock market and vice-versa.

Research question: How do the economic articles and their sentiment affect the Indian IT stock market?

Hypotheses: The study hypothesizes that sentiments of economic activities lead to a rise/fall in the Indian stock market.

1.6 Summary

The first section of the chapter introduced the study's main goal to examine the impact of foreign stock markets, economic news, and currency exchange rates on the Indian IT stock market. The problem statement section states that most existing studies, such as Kumar (2019); Srivastava & Sharma (2016), used statistical-based Granger causality tests to analyze the influence of foreign factors on the Indian stock market.

The research explains that to obtain more confident results, we can use ensemble methods, an ensemble of statistical-based, machine-learning-based, and deep-learning-based causality detection methods, for identifying the factors affecting the stock market. The research then outlines its purpose and significance before concluding the chapter by discussing the research question and hypothesis. The next chapter introduces a literature review and conceptual framework of the study.

CHAPTER II:

REVIEW OF LITERATURE

2.1 Introduction

In the recent past, several studies have been conducted to evaluate the impact of various factors, such as foreign stock market performance, currency exchange rates, and news sentiments, on the Indian stock market. Several of these studies, such as Kishor and Singh (2017); Kumar (2019); Srivastava and Sharma (2016) have used the Granger causality test and the Johansen co-integration test. A Granger causality test determines if one time series can be used to forecast another (Granger, 1969), and the Johansen co-integration test identifies long-run and short-run integration among variables (Johansen, 1991).

However, while this topic has received considerable attention, existing research has yet to be narrowed down to a specific sector (i.e., information technology). It has yet to keep up with advances in causality detection using deep learning. Even extant studies did not study the impact of foreign stock market performance, currency exchange rates, and news sentiments on the Indian IT stock market. As a result, in addition to examining the impact of the factors mentioned above on the Indian IT stock market, this research will also look at ensemble methods (ensemble of statistical-based, machine-learning-based, and deep-learning-based causality detection methods).

The following sections will cover the literature review of related studies that analyzed the impact of various foreign factors on the Indian stock market as well as the deep learning-based causality detection methods. Sections are organized below in the following order:

- Statistical causality detection algorithms.
- Deep Learning-based causality detection algorithms.
- Statistical methods to deep-learning methods.
- Impact of the foreign stock exchange on the Indian stock market.
- Impact of the news articles on the Indian stock market.
- Impact of the currency exchange rate on the Indian stock market.
- Miscellaneous
- Conclusion
- Summary

2.2 Statistical causality detection algorithms

The section gives an overview of the traditional algorithms for detecting causality and co-integration, namely the Granger causality test and Johansen's co-integration test. Both tests have been used for decades by scholars (Kumar (2019); Rao (2019)) and practitioners to study causality and co-integration between any two variables. Therefore, they are also the most popular choice for analyzing the impact of the foreign stock market, exchange rates, and news articles on the domestic stock market. This study also analyzed the impact of the above variables on the Indian stock market. Therefore, it has become crucial to review these traditional causality detection methods and compare them to the newer deep learning-based causality detection methods. The Granger causality and Johansen’s integration tests are described in detail in the following sections.

2.2.1 Granger Causality

The Granger causality test is a statistical hypothesis test for determining whether one-time series helps forecast another. As per the Granger causality, a signal X Granger causes a signal Y when the combination of past values of Y and X has a more significant impact on the value of Y than the past value of Y (Granger, 1969). Due to its computational simplicity, Granger causality has been a popular method for the causal analysis of time series data for decades. As a result, studies such as Rao (2019); Kumar (2019); Samadder and Bhunia (2018); Kishor and Singh (2017); and Srivastava and Sharma (2016) preferred Granger causality to examine the effect of foreign stock markets on the Indian stock market.

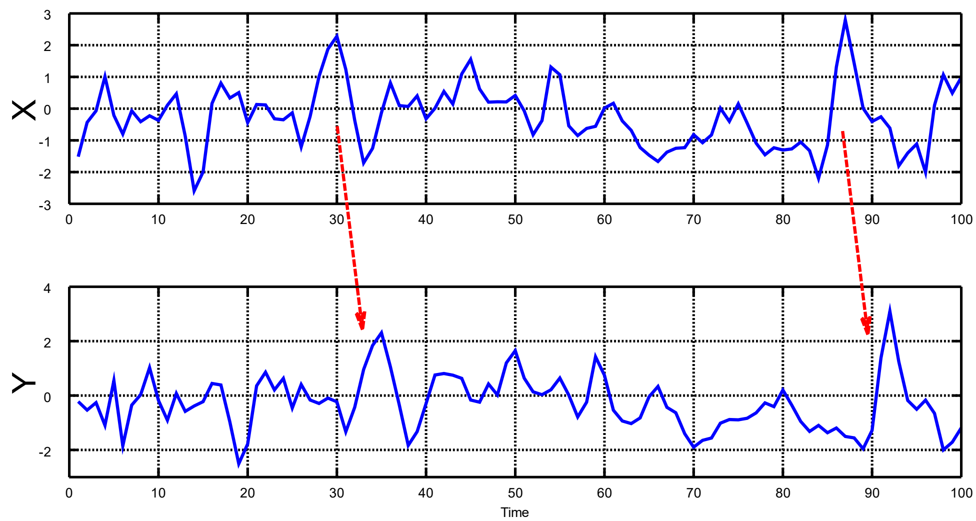

Figure 2.1 Granger causality visualization Source: (BiObserver, 2014)

Despite being so popular, Granger causality needs to be revised, such as: 1) Granger causality cannot be applied to non-stationary data. 2) Granger causality assumes linearity in the dataset for causality detection. 3) Granger causality still needs to provide more insight to understand the relationship between the variables.

As mentioned above, one of the limitations of Granger causality detection methods is that they need to work better with non-linear datasets. Text data is an example of non-linear data. This study also analyzed the impact of economic articles (text data) on the Indian stock market. Therefore, Granger causality was not the best choice for causality detection for this research. As a result, this research considered alternative approaches that showed all the strengths of Granger causality minus its limitation.

2.2.2 Johansen's Co-integration Test:

Johansen's test is a tool to identify the co-integration between three or more time series. The test validates the co-integration relationship using the maximum likelihood estimation (MLE) approach (Johansen, 1991). There are two ways to conduct Johansen's test: with trace or with eigenvalue. Both tests are used to identify the co-integration between multiple time series. For both tests, the null hypothesis states that there is no co-integration between the time series. The alternative hypothesis for the trace is to have at least one co-integration relationship between multiple time series. However, the alternative hypothesis for maximum eigenvalue states that only one combination of the non-stationary variables gives a stationary process. As a result, studies such as Deo and Prakash (2017); Mohanty and Pathak (2017); Kishor and Singh (2017); and Singh (2015) have used Johansen's co-integration test to confirm the relationship between the foreign stock market and the Indian stock market.

Johansen's co-integration works best with large sample size. However, due to the small sample size, this leads to erroneous results. Additionally, this research considered the sample data from January 1, 2018, to December 31, 2021, which is a small number of records. Therefore, there may be better choices than Johansen's co-integration test for this research.

2.3 Deep Learning-based causality detection algorithms

For the last few decades, the Granger causality test has been the most popular choice to determine the causality between foreign stock markets and the Indian stock market. However, there are a couple of limitations associated with Granger causality, which is as follows. 1) Granger causality can only be applied to stationary data. 2) Granger causality struggles to detect causality in non-linear multivariate time series data (Marcinkevičs & Vogt, 2021). These limitations motivate researchers to seek better alternatives to detect causality in non-linear data. These alternatives come in the form of deep-learning networks. Due to their ability to capture non-linearity in data (Schmidhuber, 2015), the following section discusses the deep-learning framework for causality detection.

Rosoł et al. (2022) recently introduced a python package with a neural network-based causality analysis approach. It supports Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), and Multilayer Perceptron (MLP) to perform causality tests. The neural network-based approach is more suitable for capturing patterns in non-linear settings. Even the experiments performed by the authors demonstrated that the proposed package is superior at detecting non-linear causal relationships, which are not detectable by the Granger causality method.

Hence, readily available implementation and comprehensive support for deep learning-based methods make Rosoł et al. (2022) a better alternative for causality tests on non-linear dataset settings.

In another study, Marcinkevičs and Vogt (2021) proposed a novel framework for inferring multivariate Granger causality in non-linear dynamics using self-explaining neural networks. This framework happens to be more expressive than other neural networks for inferring causality. Unlike Granger causality tests, it detects both signs and their variations over time, as well as a rational inference. Therefore, the framework facilitates a more comprehensive and insightful exploratory analysis.

This study analyzed the impact of foreign stock market variables (such as foreign IT indices, currency exchange rates, and economic sentiments) on Indian IT stock. Additionally, these variables may not have a linear relationship with the Indian IT stock. Due to the support for non-linearity for multiple variables, Rosoł et al. (2022) and Marcinkevičs and Vogt (2021) are excellent candidates for this research.

Moreover, the relationship between the foreign market and the Indian stock market can be short-term or long-term. Tank et al. (2021) introduced a deep-learning-based casualty detection method for long-term dependencies. In contrast to traditional Granger causality, this framework captures the long-term dependencies between time series efficiently and can also work on non-linear datasets. The experiments indicated that neural Granger causality methods outperformed Granger causality tests on a non-linear gene expression dataset. As a result, Tank et al. (2021) is a viable option for efficiently detecting causality due to their ability to capture the long-term dependencies between the foreign stock market and the Indian IT stock market.

In addition to advanced deep learning methodologies, specifying and testing the assumptions about observed data using a statistical estimator yields comparable results.

Therefore, Sharma and Kiciman (2020) developed an open-source library named DoWhy. It is based on the framework of causal graphs, which can be used to identify and test the causes of observed data. Most of the prior causal inference libraries consist primarily of statistical estimators. However, the successful application of causal inference requires specifying assumptions about the underlying observed data and validating those assumptions with statistical estimators. Consequently, the "DoWhy" library provides an application programming interface (API) with four common steps for causal analysis: 1) Modeling data using a causal graph and structural assumptions. 2) Using a graphical model to identify desired causal effects. 3) estimating the effect using statistical estimators, and 4) refuting the obtained estimate using robustness checks and sensitivity analyses. In summary, the library focuses on devising the correct causal model and testing its assumptions. That makes the implementation more robust and accurate for causality detection. Thus, the "DoWhy" library seems to be a suitable candidate for determining the causal impact of the foreign stock market on Indian IT stock.

Huang et al. (2020) proposed a machine learning-based causality method called Reservoir Computing Causality (RCC). Differing from the standard Granger causality method, the RCC framework has better robustness to noise and improved computational efficiency.

In recent advances in deep learning, the attention-based mechanism has proven to be a fundamental building block because it captures the long-term dependency very well (Bahdanau et al., 2015). Hence, Nauta et al. (2019) presented the Temporal Causal Discovery Framework (TCDF), an attention-based convolutional neural network combined with a causal validation step. This framework constructs a temporal causal graph that shows the relationships between time series and their corresponding time delays. Moreover, TCDF claims state-of-the-art performance in discovering causal relationships in continuous time series data on financial and neuroscientific benchmarks. Consequently, this study also considers the outcome of the TCDF framework to determine the causal relationship between the foreign stock market and Indian IT stocks.

Chivukula et al. (2018) proposed deep learning networks to discover the Granger cause from multivariate temporal data in the financial market. The network discovers Granger-Causal features for bivariate regression. Several tests, such as F-tests and T-tests, have been used to select features and compare models. Moreover, the results of these tests were favorable.

Most of the deep learning-based causality detection methods discussed above are claimed to outperform traditional Granger approaches. As a result, all these methods appear to be viable options for detecting causality between the foreign stock market and the Indian IT stock market. However, to arrive at more robust and convincing conclusions, the study has used the representation of statistical-method, machine-learning-based, and deep-learning-based methods. Following that, the majority vote from these representations will be used to demonstrate the impact of the foreign IT stock market, currency exchange rate, and economic articles from different sources on the Indian IT stock market.

2.4 Statistical Methods to Deep-Learning

Granger causality is the most widely used tool for identifying causal relationships between two-time series. Johansen's co-integration is employed for defining long-term and short-term integration between two or more time series. However, both techniques are limited to non-linear data and small sample sizes.

Rosoł et al. (2022); Marcinkevičs and Vogt (2021); Tank et al. (2021); and Nauta et al. (2019) introduced deep-learning-based causality detection approaches to support non-linear data and efficiently discover the long-term and short-term integration between two or more time-series.

Furthermore, it is expected that the careful selection of a few representations from the representation of ensemble methods (statistical, machine-learning-based, and deep-learning-based methods) approaches and their votes to identify causality will outperform standard causality detection techniques. Therefore, this research has the majority vote from ensemble methods (statistical, machine-learning-based, and deep-learning-based) to confirm the impact of foreign stock markets, exchange rates, and economic items on the Indian stock market.

2.5 Impact of the Foreign Exchange on Indian Stock market

Kumar (2019) examined the co-integration between the Nifty (50) Future Index and the Future Stock Market from advanced economies. The list of these advanced economies is as follows: Nikkei 225(Japan), NASDAQ 100 Futures (USA), Dow Jones 30 (USA), SSEC (China), Hang Seng Future (Hong Kong), and FTSE 100 (London). The analysis was based on the monthly stock prices from April 2008 to March 2018. The interdependencies between the Indian and foreign stock markets have been measured through the Granger causality test. The result of the study reveals that the Nifty 50 future can describe the future stock market of Nikkei (Japan) and SSEC (China). Moreover, the Hang Seng future (Hong Kong) and Nifty 50 futures can explain each other.

Similarly, Rao (2019) investigated the stock market integration between the important foreign stock markets: China, Japan, India, European Union, the UK, and the USA. The study was based on the stock price data from April 2013 to May 2018. The result of descriptive analysis reveals that there is a long-run relationship among all the six stock exchanges. However, the Granger causality test confirmed that the Indian stock market does not impact the foreign stock market and vice versa.

Gulzar et al. (2019) examined the financial co-integration and spillover effect of the global financial crisis on emerging Asian financial markets (India, China, Pakistan, Malaysia, Russia, and Korea). The study has used daily stock returns and divided them into three different periods, namely, the pre-, during-, and post-crisis periods, from July 1, 2005, to June 30, 2015. Moreover, the study has used the Johansen's co-integration test, the vector error correction model (V.E.CM.), and GARCH to investigate the integration between the global financial markets and emerging Asian markets. The study has found long-term co-integration between the US market and the emerging stock market, with the level of co-integration increasing post-crisis.

A stock market can have a long-term or short-term impact on another stock market. The relationship between foreign and Indian stock markets is no exception. Thus, Samadder and Bhunia (2018) have examined the short-run and long-run relationships between the Indian stock market and selected developed stock markets, including Australia, Canada, France, Germany, the UK, and the USA. Their study considered time-series data from 2001 to 2016. The Granger causality test found that the Indian stock market and the USA stock market are associated in the long run. However, Germany and France are associated with the Indian stock market in the short run.

The Bombay Stock Exchange (BSE) and the National Stock Exchange (NSE) are India's two major stock exchanges (NSE). As a result, most of the research used the BSE or NSE to represent the Indian stock market. Deo and Prakash (2017) and Singh (2015), for example, examined the impact of the foreign stock exchange on the NSE (the Indian stock exchange). However, Manu and Menda (2017), Mohanty and Pathak (2017), and Patel and Patel (2012) investigated the impact of foreign stock exchanges on the BSE (the Indian stock market representative). The specifics of these studies are as follows.

Deo and Prakash (2017) examined the interlinkage between NSE and leading international stock markets from 2006 to 2015. They have used the Johansen's co-integration test to confirm the long-term relationship between NSE and major foreign stock exchange indices. Moreover, Granger causality has been used to establish the bidirectional and uni-directional relationship. It confirmed the bidirectional relation between India (NSE), Japan (NIKKIE 225), and China (SHANGHAI) and the uni-directional relation between India (NSE) and Australia (Australia Stock Exchange), Germany (Deutsche Stock Exchange), Europe (Euro Nxt), Hong Kong (Hang Sang), US (Nasdaq), US (Nyse), China (Shenzen Stock Exchange).

Manu and Menda (2017) explored the dynamic connection between Indian stock markets, such as Nifty 50, Nifty 100, Nifty 200, Nifty Midcap 50, and BSE SENSEX, and foreign stock markets, including Japan's Nikkei 225, China's Shanghai Composite and

Taiwan Weighted, the US's NASDAQ, the UK's FTSE 100, Germany's DAX, South Korea's KOSPI, and France's CAC 40. They conducted the study based on daily returns from April 2004 to March 2015 and found that the Indian stock market positively correlated with the foreign stock market. Additionally, Panda and Nanda (2017) studied the relationship between South American and Central American stock markets using weekly returns from 1995 to 2015. They applied the Granger causality test and discovered strong co-movement among Venezuela, Chile, and Peru. Although their study was not focused on the Indian stock market, their method of using weekly returns for impact analysis is valuable to consider.

Mohanty and Pathak (2017)investigated the interlinkages among the selected indices, viz. NYSE (USA), Nikkei 225 (Japan), Shanghai Composite Index (China), Hang Seng (Hong Kong), Kospi (Korea) and BSE30 (India). The study covers the data from July 1997 to December 2014. The study used Johansen's co-integration test to detect short-term and long-term integration and the Granger causality test to understand the lead-lag effect among stock markets. The result of the study conveyed the bi-directional causality between BSE and KOSPI, BSE and SHANGHAI, NYSE, and KOSPI, KOSPI, and HENG SENG, and NIKKEI and HANG SENG.

Kishor and Singh (2017) examined the relationship between the BRICS stock exchanges, i.e., Brazil, Russia, India, China, and South Africa. Their study was carried out over daily index returns from 2007 to 2014. It has used the Johansen’s co-integration test to show long-run association and the Granger causality test to find the cause-and-effect relation. The result of the study shows that Nifty is positively correlated with other BRICS indices. Specifically, Uni-directional (one-way) causality is observed in Nifty with South African and Chinese Stock Indices. Moreover, bi-directional (two-way) causality was present among Indian, Brazilian, and Russian Stock Indices.

Similarly, Srivastava and Sharma (2016) studied the inter-linkage among the daily returns from stock markets in Hong Kong, South Korea, Japan, India, China, and Pakistan. Their study used the daily returns from 2003 to 2013. The result of the Granger causality test confirmed that all six stock exchanges have a cause-effect relationship.

A study by Abdullahi (2020) focused on the co-movements between the Nigerian and Dubai stock exchanges. The study covers the monthly share index from September 2009 to August 2019. The research has adopted the Generalized Method of Moment for co-integration detection. The result of his study confirmed the short-run causality between the Dubai financial market and the Nigerian stock exchange.

Singh (2015) examined the interdependency between the Tokyo Stock Exchange (TSE), Hong Kong Stock Exchange (HSE), Bombay Stock Exchange (BSE), and National Stock Exchanges (NSE). The study used daily closing data from March 2007 to August 2014. The result of the Granger causality test revealed that the Indian stock market influences the Japanese and Hong-Kong stock markets. The Johansen's co-integration conveyed that the influence lasted for a short period.

Ncube and Mingiri (2015) conducted a study to identify the link between African and world stock markets. The study covered the African stock markets (South Africa (All Share Index), Botswana (All Company Index), Namibia (NSX overall Index), Mauritius (SEMDEX Index), and Nigeria (NGSE Index)) and the world stock markets (Germany

(DAX index), Japan (NIKKEI 225 index), and the USA's S&P 500 index). It used monthly data from 2000–2008. The results of Johansen's co-integration and the Granger causality tests confirmed that the international stock market affects the African stock market.

Patel and Patel (2012) examined the co-movement between India (BSE), the UK (FTSE), Hong Kong (Hangseng), Jakarta (JKSE), Japan (NIKKEI), Sri Lanka (CSE), Switzerland (SMI), China (SSE) and Taiwan (TSEC). The study used data from January 2001 to December 2011. The methodologies used in the study were correlation and Granger causality. The results conveyed the correlation between BSE (Bombay stock exchange (India)), and FTSE is 96%, BSE and HANG SENG is 95%, BSE and JKSE is 95%, BSE and TSEC is 87%, BSE and CSE is 82%, and BSE and SSE is 73%. The Granger causality test found that BSE is unaffected by any selected markets. However, BSE affects FTSE, Hangseng, JKSE, CSE, and TSEC.

Gangadharan and Yoonus (2012) examined the impact of the global financial crisis on the degree of financial integration between the US and Indian stock markets. The study was carried out from March 2005 to November 2010. It has used the daily returns of the indices of the US (S&P 500) and Indian stock markets (CNX S&P). Using Johansen's co-integration analysis and the Vector Auto Regression (VAR) model. The study has examined if there is co-integration and a dynamic relationship between the US and Indian indices during the pre-crisis, crisis, and post-crisis periods, as well as the period comprising all three durations. The article's findings revealed no co-integration between the two indices for the four periods. However, the previous day's returns in the US stock market impacted the Indian stock market. During the same period, no vice-versa influence was seen.

Tripathi and Sethi (2010) examined the integration between the major global markets such as the United States (S&P 500 and DJIA), United Kingdom (FTSE 100), Japan (NIKKEI 225), and China (SSE Composite) with the Indian stock market (NIFTY). The study has gathered daily stock prices over the period ranging from January 1, 1998, to October 31, 2008. After that, the study utilized Granger's causality test and Johansen's co-integration to examine the integration and impact, respectively. The study's findings indicated that, in the long run, the Indian stock market was not individually integrated with any of the major global stock markets, apart from the United States. The author of the paper hypothesized that one of the causes of the observed outcomes was the growing economic and financial linkages between India and the USA.

After analyzing the literature presented above, it has been observed that most of the research examined the impact of major foreign country indices on the Indian stock market. While reviewing the literature, it was discovered that no study had examined the effect of foreign IT indices on the Indian IT stock market. In contrast, the Indian IT sector is profoundly influenced by the global stock market, as more than 85% of its revenue is contributed by foreign countries (Thehindubusinessline, 2020). Therefore, narrowing down the foreign stock market and the Indian stock market to a specific sector, i.e., information technology (IT), is expected to yield better results. Thus, the present study has considered foreign IT indices as one of the factors for analyzing the impact on Indian IT stocks.

2.6 Impact of the News Articles on the Indian Stock Market

The global market indices and media articles influence the movement of a national stock market. The following section will examine the effect of news articles on the Indian stock market in detail.

Recently, Hicham and Salah-Ddine (2021) predicted stock movement based on financial news from different sources. It has used financial news from four Moroccan economic journals from September 2014 to February 2019. The methodologies used for the study are a combination of support vector machines and artificial neural networks. The results of this merged learning are more accurate than those produced by a simple news analysis based on only one source of information.

Paramanik and Singhal (2020) analyzed the impact of news articles on financial market volatility. The data for the study was collected from April 2007 to January 2020. The study used the GARCH model to analyze the impact of sentiments on the stock market. The GARCH model is a statistical modeling technique used to predict the volatility of returns on financial assets. The study's empirical findings conveyed the dominance of negative views over positive opinions on the stock market.

Financial news can also be represented in the form of psycho-linguistic features. The psycho-linguistic is the field of study in which researchers investigate the psychological processes involved in the use of language. The tools used for analyzing psychological features are Linguistic Inquiry and Word Count (LIWC) and the Tool for the Automatic Analysis of Lexical Sophistication (TAALES). Thus, Kumar et al. (2020) attempted to forecast the Indian stock market using the psycho-linguistic features of financial news. The study has extracted LIWC and TAALES from news articles from December 2015 to May 2016. After experimenting with various methodologies, the Group Method of Data Handling (GMDH) and General Regression Neural Network (GRNN) must be the best statistical techniques for predicting stock prices using the psycho-linguistic features of financial news.

Many researchers analyzed the sentiment of news articles on the stock market. However, only some researchers examine the effect of Google web searches on the stock market downfall. For instance, Nikkinen and Peltomäki (2020) examined the effect of news articles and web searches expressing crash fears on stock market returns. The study utilized weekly newspaper articles, and Google searches with the word "crash" for January 2016. The State-space model analyzes the dynamic interaction of the Google Search Volume Index (GSVI) and articles.

Further, the Distributed lag model has been used to examine the effect of GSVI and news articles on the stock market. The study results concluded that the impact of web searches is immediate on the stock market. However, the effect of news articles can last longer, up to 11 weeks.

Biswas et al. (2020) analyzed the impact of the sentiments of news articles on the economic downfall of the primary international stock market. The New York Times news articles and the Dow Jones stock market reports have been collected from September 2001 to December 2001 and from January 2020 to May 2020 to analyze the impact of previous events like terrorist attacks or epidemics (9/11, Covid-19). The study used the tool named VADER for sentiment analysis. The result of the study conveyed that, at the time of both events, there was more negative news than positive news. Hence, a Fall in the stock market has been observed.

Hwang and Kim (2019) analyzed the inter-dependency between news sentiments and stock prices. Their study was based on news articles about North Korea and Tesla stock between April 2019 and July 2019. Different machine learning models such as Gaussian processes (GPs), linear regression (LR), multilayer perceptron (MLP), support vector regression (SVR), and random forest (RF) were used to evaluate the relationship between news sentiments and stock prices. The study found that news content has a significant influence on stock prices.

Shah et al. (2019) studied the impact of news sentiment on pharmaceutical market sector stocks. This research's main contribution is using a sentiment analysis dictionary. The domain experts created the Sentiment Analysis Dictionary in this research specifically for a domain like pharmaceuticals. Due to the dictionary-based learning approach, the study claimed to achieve 70.59% directional accuracy in predicting the trends in short-term stock price movements using news sentiment.

Wong and Ko (2017) collected 24,763 articles on national business and politics in 2015. They proposed a model to use public emotions by predicting the movements of stock market indices. The study used DepecheMood (Staiano and Guerini (2014)) to extract emotional scores from news articles. The study's results confirmed the significant correlation between public emotions and news articles.

Verma et al. (2017) presented methodologies for accessing the impact of news articles on the Indian stock indices. The study used PESTEL (Politics, Economic, Social

Factors, Technology, Environment, and Legal) events from 2013 to 2016. The Granger causality test is used for impact analysis and LSTM (long-term, short-term memory) for prediction. The result of the study confirmed that PESTEL events significantly impact the specific sector of the stock market. Furthermore, because of its ability to learn long-term dependencies, the LSTM model based on the PESTEL framework outperforms the previous state-of-the-art techniques.

Lima et al. (2016) proposed a mechanism to predict stock prices using Twitter sentiments. The tweets were collected for the study between September 2015 and December 2015. The library called "Sentiment140" has been used for analyzing tweets and the SVM algorithms to predict dynamic behavior. The study's results showed a significant gain in correctly classifying instances using the SVM algorithm.

As discussed, many studies analyzed the impact of single-source news sentiments on the Indian stock market. However, very few studies, such as Hicham and Salah-Ddine (2021), analyzed the impact of multiple journal news sentiments on the Indian stock market. Moreover, while reviewing the existing literature, none of the studies has examined the impact of named entity recognition (NER) features and sentiments on the Indian stock market. A named entity is a real-world object with a name, such as people, organizations, locations, medical codes, time expressions, quantities, monetary values, percentages, and so on (Honnibal, 2015). Hence, this research looked at the impact of named entities and sentiments from various news sources on Indian IT stocks.

2.7 Impact of the Currency Exchange Rate on Indian Stock Market

The following section will review the existing studies that have analyzed the impact of currency exchange rates on the stock market.

Bhattacharjee and Das (2020) investigated the relationship between the USD-INR currency exchange rate and the Indian stock market. It was conducted over data from April 2005 to December 2019. Apart from the standard causality test, the study used the Dickey and Fuller test for non-stationarity testing. Dickey and Fuller (1979) describe a standard statistical test used to test whether a given time series is stationary. The results of this test have revealed that the currency exchange rate and the Indian IT stock market are non-stationary at the current level. Moreover, the Johansen's co-integration and the Granger causality test identified a long-run and bi-directional causal relationship between the USD-INR exchange rate and the Indian stock market.

Mohith and Sangeetha (2019) studied the impact of currency exchange rates on selected sectoral indices during specific events. The study used sectoral indices viz. Nifty Auto, Nifty Bank, Nifty Energy, Nifty Pharma, and Nifty FMCG, and the events considered for the study were the global financial crisis of 2008, "Demonetization, Brexit, and the 2018-rupee crisis". The results of the 2008 financial crisis reveal the uni-directional Granger causality between USD and Nifty Bank, USD and Nifty Energy, Nifty FMCG, and USD and Nifty Pharma. However, in the event of demonetization, a one-way impact has been observed only between USD and Nifty auto. In the case of Brexit, only GBP has been considered, and no causality has been observed between GBP rates and any sectoral indices. Moreover, for the 2018 crisis, only USD was considered, and uni-directional causality has been confirmed between USD and Nifty auto.

AbdulRahman Bala and Hassan (2018) investigated the interaction between the exchange rate and the Nigerian stock market using annual data from 1985 to 2015. The study has examined the relationship between the exchange rate, economic growth, money supply, and stock market using the autoregressive distributed lag (ARDL) model and Granger causality tests. According to the study's findings, the exchange rate and economic expansion have a favorable and statistically significant impact on the Nigerian stock market. However, the money supply has a detrimental and statistically significant impact on the Nigerian stock market. Based on the result of the study, it was advised that effective monetary policies be put into place to maintain a stable exchange rate and prevent structural breaks that would negatively impact the entire system, including the stock market.

Anuradha (2018) analyzed the dynamic linkage between the exchange rate and the Bombay stock exchange (An Indian stock market index). The study considered data from 2010 to 2016 for correlation analysis. The result of the regression algorithm confirmed the effect of the currency exchange rate on the Bombay stock exchange. It indicates that as the rupee depreciates against the dollar, the stock price increases and vice-versa.

Furthermore, Manu and Bhaskar (2018) examined the effect of currency exchange rate volatility on the Indian stock market performance. Their study has considered four exchange rates (EURO/INR, USD/INR, GBP/INR, and YEN/INR) and three stock market indices (BSE SENSEX, BSE 500, and Nifty 50) from April 2000 to March 2016. The GARCH methodology was used in the research. The result of the study indicates that the previous day's currency exchange rate will influence the current day's price of selected stock indices. Hence, currency exchange rate volatility significantly impacts the Indian stock indices.

From the above literature, most studies analyzed the impact of EUR, USD, GBP, and YEN on the Indian stock market. However, in addition to the USA, Great Britain, France, and Japan, Indian IT companies are heavily invested in Singapore, Mauritius, the Netherlands, and Germany (INDIA: FOREIGN INVESTMENT, undated). Therefore, it is worth considering the impact of these countries' exchange rates on Indian IT stocks. In addition to the foreign stock markets and economic articles, this study also looked at the influence of the exchange rates of the countries mentioned above on Indian IT stocks.

2.8 Miscellaneous

All the research discussed in the preceding section examined the impact of international stock markets, currency exchanges, and economic articles on the Indian stock market. However, a significant amount of research has been conducted on stock price prediction (SPP) using overseas stock markets, currency exchange rates, and global news such as (Sharma, 2020; Vargas et al., 2018). As a result, this section examines studies that have used the abovementioned characteristics to predict stock prices using deep learning techniques.

The combination of global indices, currency exchange rates, historical stock prices, world news, and technical indicators yields more stable results, which do equally well (with 60% accuracy) in predicting the increasing and decreasing trends of Indian IT stocks. The study used the ensemble of long-short term memory (LSTM) to predict the trends of Indian IT stocks (Sharma, 2020). The focus of the study is to predict the Indian IT stock price trends as accurately as possible. The study, however, fell short of discussing the causal impact of the variables mentioned above on the Indian IT stock market. As a result, this research aims to determine the causal impact of foreign stock markets, currency exchange, and economic events on the Indian IT stock market.

Stock price prediction has remained a topic of interest for the past few decades. The researchers applied statistical techniques to advanced deep-learning techniques to predict stock price trends. Recently, Gao et al. (2019) have used historical prices of the stock from the Taiwan Stock Exchange Corporation (TWSE) to predict the stock price trend. The study has proposed a Convolution Long-Short Term Memory (ConvLSTM) architecture. The LSTM component of the architecture claims to improve long-term dependency while also increasing forecast accuracy and stability.

Like (Sharma (2020); Vargas et al. (2018)) considered historical time-series data, technical indicators, and world news articles to predict the stock price. The study used a hybrid model to anticipate the stock price with a Convolution Neural Network (CNN) for news items and an LSTM for technical indicators. Li et al. (2016) also use extreme learning machines (ELM) to predict stock prices using intra-day data. The study claims to have produced a more accurate prediction using historical prices and news articles.

Furthermore, many researchers analyzed news articles and social media sentiments for stock price prediction using machine-learning and deep-learning techniques. Khan (2022) has employed machine-learning algorithms to analyze social media and financial news to determine the stock price for the next ten days. The study has meticulously removed spam tweets from the dataset to increase the performance and quality of the predictions. Additionally, the study has been conducted on the stock market, which is hard to anticipate and is heavily influenced by social media and financial news. Finally, the study has used an ensemble of deep-learning and machine-learning-based classifiers to achieve maximum prediction accuracy.

Similarly, Wang et al. (2019) proposed an enhanced learning-based neural network to predict the stock price based on news sentiments. The proposed network illustrates the improvement over the existing methods in reducing the Mean Square Error (MSE). Similarly, Pagolu et al. (2017) used the word-to vector (word2vec) and the N-gram technique to analyze sentiments from tweets. Following that, the study used supervised machine learning techniques to convey the strong correlation between the public sentiments on Twitter and the stock price. Joshi et al. (2016) represented the news articles regarding positive or negative sentiments. After that, Random Forest (RF) and Support Vector Machine (SVM) algorithms are used to predict the stock price trend. The study claims to achieve 80% accuracy in anticipating stock price movement.

Across all these studies, it is evident that deep-learning-based techniques achieved better results predicting stock prices. Hence, to have more robust and convincing results, the study examines the impact of foreign stock markets, currency exchange rates, and economic articles on the Indian IT stock market using ensemble methods (statistical method, machine-learning-based methods, and deep-learning-based methods).

2.9 Conclusion

This chapter has examined the empirical studies that have analyzed the impact of the foreign stock market, currency exchange rate, and news articles on the Indian stock market. Nevertheless, no significant studies have analyzed the impact of the variables on the specific sector (such as information technology (IT)) of the Indian stock market. Furthermore, the existing studies have employed the Granger causality test and the Johansen co-integration test to examine the impact of international stock markets, currency exchange, and news items on Indian stock markets. However, extant research still needs to consider ensemble model (incorporating statistical, machine learning, and deep-learning-based methods) methodologies for analyzing the impact, which is not limited by Granger's causality constraints and is more robust and convincing.

Moreover, the studies such as Paramanik and Singhal (2020); Biswas et al. (2020); and Shah et al. (2019) have examined the impact of news article sentiments on the Indian stock market. However, in the context of the Indian stock market, no significant studies have been reported that have expressed economic articles as named entity recognition (NER) and analyzed their impact on the Indian stock market. Therefore, extant research has a substantial gap in analyzing the impact of named entities from economic articles on the Indian IT stock market using deep-learning-based techniques.

Additionally, using multiple causality detection techniques would enhance the credibility of the findings of this study. Hence, the study analyzes the impact of foreign IT stock markets, currency exchange rates, and economic articles on the Indian IT stock market using ensemble models

2.10 Summary

The chapter began with a discussion of the existing statistical methods, such as Granger Causality and the Johansen Co-integration Test for identifying relations between two or more time series. Both methods are limited in terms of non-linear data and small sample sizes, respectively. Following that, deep-learning-based causality detection methods such as Rosoł et al. (2022); Marcinkevičs and Vogt (2021); and Nauta et al. (2019) have been discussed. Due to their underlying architectures, deep-learning-based approaches eliminate the restrictions of Granger causality and the Johansen's co-integration test. Hence, they perform better while detecting causality in non-linear data settings.

The later section of the chapter discussed the impact of various foreign factors on the Indian IT stock market. These foreign factors include foreign stock indices, currency exchange rates, and news articles. A comprehensive review of stock price prediction using overseas stock markets, currency exchange rates, and global news through deep-learning-based techniques was conducted. Finally, the chapter concludes by discussing the gap in the existing studies. These gaps include not investigating machine-learning and deep-learning-based techniques along with statistical-based techniques for detecting causality, not restricting studies to specific sectors such as information technology, and not expressing economic articles in terms of named entities. Therefore, this study analyzes the impact of foreign IT stock markets, currency exchange rates, and economic articles (named entities and sentiments) on the Indian IT stock market using ensemble models (incorporating statistical, machine learning, and deep-learning-based methods). The next chapter addresses the research methodology, data collection methods, data analysis, and validation in detail.

CHAPTER III:

METHODOLOGY

3.1 Introduction

This chapter explains the methodology adopted for this research. After that, the chapter will describe the different data sources, data collection, and selection methods. The data pre-processing and statistical-based, machine-learning, and deep-learning-based techniques for impact analysis will be thoroughly discussed. Finally, the chapter will establish the research's justification, validity, reliability, and limitations.

3.2 Methodology approach

The study was an exploratory study that adhered to the quantitative research principle. The study used secondary data to meet the study's objective. The data were obtained from secondary sources such as www.investing.com, www.moneycontrol.com, www.ft.com, www.thehindu.com, and Google finance Sheets.

In extant research, the Granger causality test is the most widely used approach for detecting causality between variables. However, this technique has limitations in terms of supporting non-linear data. On the other hand, deep-learning methods excel at capturing patterns on non-linear datasets. As a result, the study proposed to analyze the impact of foreign IT indices, name entities and sentiment features of economic articles, and currency exchange rates on Indian IT stocks using an ensemble model (incorporating statistical, machine learning, and deep-learning based methods) to attain robust and credible results.

The causality analysis framework is provided below, which details the proposed methodology for analyzing the impact of foreign IT indices, economic articles, and currency exchange rates on Indian IT stocks.

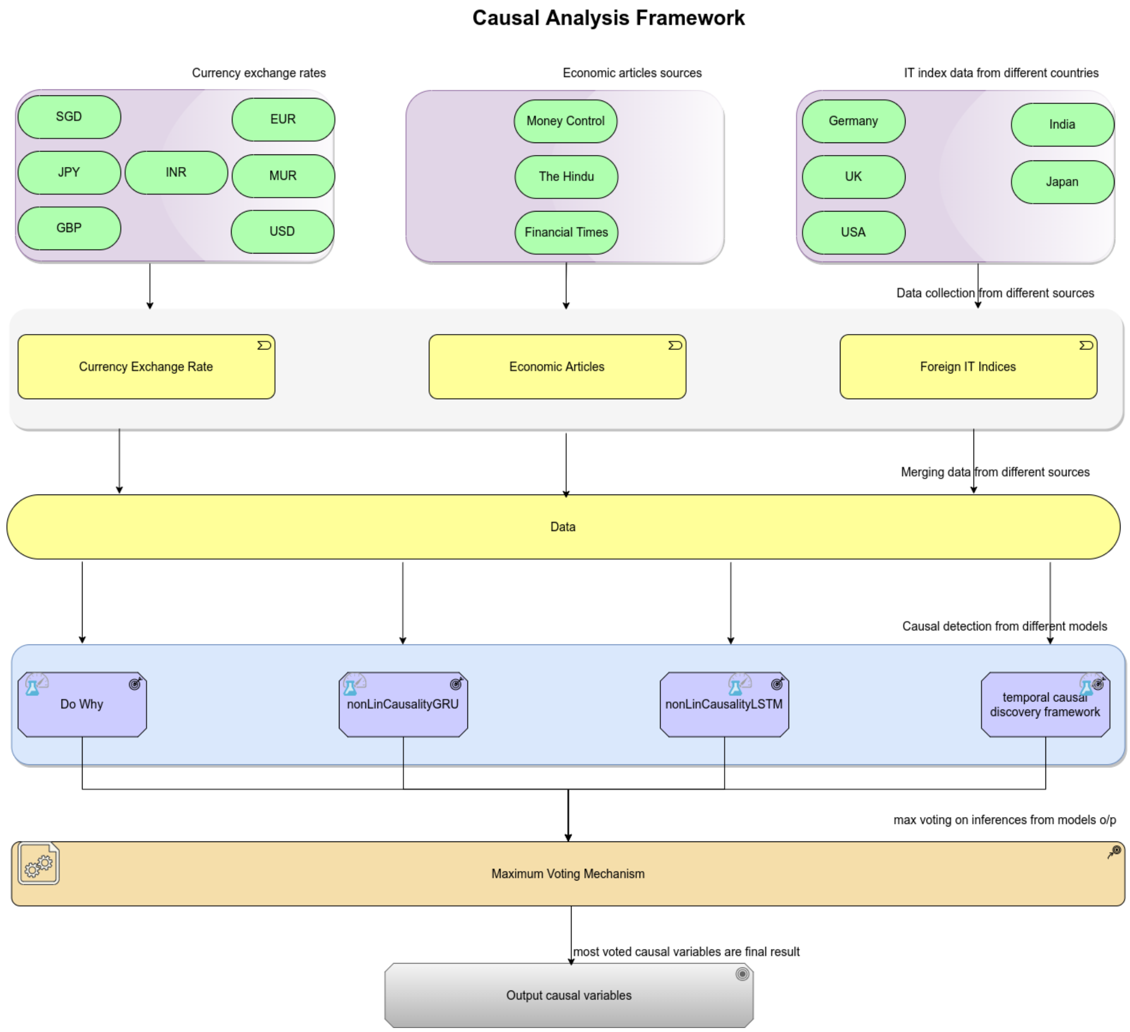

Figure 3.1 Causal Analysis Framework Source: (Kumar Sharma et al., 2022)

As shown in Figure 3.1, the currency exchange rates considered for this study are Singapore (SGD), Japan (JPY), the United Kingdom (GBP), Germany (EUR), Mauritius (MUR), and the United States (USD). Similarly, economic articles are sourced from Money

Control, The Hindu, and Financial Times, and foreign IT indices are drawn from Germany, the United Kingdom, the United States, France, and Japan.

Furthermore, data from various sources is cleaned, pre-processed, and combined. Data cleaning, pre-processing, and combining techniques will be covered in depth later in the chapter. The resulting data is then fed into four different causality detection algorithms, namely "DoWhy" (Sharma and Kiciman (2020)), nonLinCausality GRU (Rosoł et al. (2022)), nonLinCausality LSTM (Rosoł et al. (2022)), and a temporal causal discovery framework (Nauta et al. (2019)).

Finally, the output of each deep-learning causality detection, along with the output of the granger causality algorithm, is passed to the top voting mechanism layer. This layer would collect the causal variables that appeared as output for most of the algorithms.

3.3 Research Philosophy

Research philosophy refers to the beliefs that underpin how data should be collected, analyzed, and applied to a specific phenomenon (Kirongo et al., 2020). This study applies quantitative research to analyze the impact of foreign IT indices, named entities, and sentiments from economic articles and currency exchange rates on Indian IT stocks. Quantitative research is the process of collecting and analyzing numerical data. It emphasized finding the patterns, predicting, and testing causal relationships to generalize results to the wider population (Daniel, 2016). Moreover, this study hypothesizes that the change in foreign IT indices, currency exchange rates, named entities, and sentiments of economic activities leads to a rise/fall in the Indian IT stock market. This study has employed statistical, machine learning, and deep learning-based causality detection methodology to test the above hypothesis. The philosophy that develops hypotheses like the above based on quantifiable observations and tests them during the research process is called positivism (Park et al., 2020). Therefore, this study adheres to the positivist research philosophy.

It is essential to collect data from reliable sources to study the cause-and-effect relationship between variables. Hence, the study has collected data from trustworthy secondary resources such as www.investing.com, The Hindu, The Financial Times, and Money Control. Only after that the study analyzed the impact of collected data on the Indian IT stock market using statistical, deep-learning, and machine-learning-based causal detection techniques.

3.4 Research Approach

The research approach is a plan and procedure covering everything from rough assumptions to detailed data collection methods, analysis, and interpretation (Woiceshyn & Daellenbach, 2018). This study used a deductive research approach. A deductive approach goes from a broad to a more specific perspective. It starts with the theory, generates hypotheses based on it, identifies, and collects data, tests the hypotheses, and revises the theory. Similarly, this study narrowed down to the specific sector of the stock market and collected numerical data for foreign IT indices from secondary data sources. Likewise, the exchange rates of the countries that have invested the most in Indian IT companies have been carefully considered. Three primary business news sources also consider the entities and sentiment characteristics mentioned. Subsequently, to attain robust and credible results, this study applied an ensemble model (statistical, machine learning, and deep-learning-based methods) to examine how the above variables affect the Indian IT stock market.

3.5 Research Strategy

A research strategy is a step-by-step plan of action that guides a researcher's thinking. In addition, it allows the researcher to conduct research in a systematic and timely manner (Creswell & Creswell, 2018). The study used experimental research to establish causality between variables because it allows researchers to identify causal effects. The study collected numerical data for Indian IT stocks, global IT indices, exchange rates, named entities, and sentiment traits from various economic sources to analyze the impact on Indian IT stocks using statistical, machine learning, and deep learning methods.

3.6 Data Collection

The methods used to collect data to make better decisions are referred to as data collection procedures (Bhandari, 2021). There are basically two types of data collection procedures: one is primary data collection, and the other is secondary data collection. The primary data is collected from first-hand experience and has not been used in the past (Sherif, 2018). Usually, the primary data is collected for a specific research purpose through interviews or questionnaires. Hence, it is highly accurate. Whereas the secondary data has been captured by someone other than the user. The secondary data is readily available through government reports, business journals, the internet, etc. These secondary data sources are less expensive and time-consuming than primary data. Therefore, this study gathered information from secondary open data sources quickly. It gathered quantitative data from investing.com for Indian IT equities, global IT indexes, and currency exchange rates. While named entities and sentiment features for economic articles were derived from primary economic news sources such as Money Control, The Hindu, and The Financial Times.

The Indian IT stock market is represented by the Nifty IT (NIFTYIT) index. The following indices represent the foreign market: DAX Software (CXPSX) (Germany), Tokyo SE TOPIX17 IT & Services Stock Price (ITSV17.T) (Japan), CAC Technology (FRTEC) (France), S&P 500 Information Technology (SPLRCT) (USA), and FTSE 350 Software & Computer Services (FTNMX101010) (UK). Moreover, this study has considered the exchange rates of EUR-INR, JPY-INR, MUR-INR, SGD-INR, and USD-INR. Hence, after collecting the data from the above representation, this study has analyzed the impact on the Nifty IT (Indian IT stock market). The details of the data sources for the indices mentioned above, currency exchange rates, and economic articles have been discussed in a later section of the study.

3.6.1 Research Population and Sampling Method

The research population comprises individuals, groups, organizations, and other entities of interest (Casteel & Bridier, 2021). As already mentioned in the data collection, this study's population included foreign IT indices (DAX Software (CXPSX)(Germany), Tokyo SE TOPIX17 IT & Services Stock Price (ITSV17.T) (Japan), CAC Technology (FRTEC)(France), S&P 500 Information Technology (SPLRCT)(USA), and FTSE 350 Software & Computer Services (FTNMX101010) (UK), currency exchange rates (EUR-INR, JPY-INR, MUR-INR, SGD-INR, and USD-INR), and economic articles from Money Control, The Hindu, and Financial Times. This study has collected the daily observations for the features mentioned above from January 2018 to December 2021. This method of gathering and analyzing the observations of clearly defined data items recorded, monitored, down-sampled, and aggregated over time is called time-series data. Additionally, there are two types of time-series data: regular and irregular. Regular time-series data is gathered by specific sensors or software at predetermined intervals (every 10 minutes, for instance). In contrast, data collection in an irregular time series is triggered by either user activity or other external events. The average trade price of a stock every ten minutes throughout the day illustrates an irregular time series (Dix, 2021).

This study has collected the closing prices of foreign IT stock markets and the currency exchange rate, which are reported daily at a fixed time, thus being regular time-series data. However, the economic articles are published multiple times a day as per the external situations, leading to irregular time-series data. However, this work transformed irregular time-series data into regular time-series data by combining named entities and sentiments over the entire day. Hence, this study has used regular time-series data to assess the effects of the global IT stock market, currency exchange rate, and economic articles from Money Control, The Hindu, and The Financial Times on the Indian IT stock market.

3.6.2 Data Sources

The data sources describe the various aspects of the data in the study. Generally, a data source can be a database, a flat file, streaming data, or data extracted from a website. Identifying the required data sources from the most appropriate sources is very important to answering the research question. Hence, this research has extracted foreign IT indices and currency exchange rates from inversting.com (known for stock market data) and named entities and sentiments from Money Control, The Hindu, and The Financial Times.

The financial markets portal investing.com offers real-time daily worldwide stock market data, prices, charts, financial tools, breaking news, and analysis in 44 languages and across 250 exchanges. According to SimilarWeb and Alexa, Investing.com is among the top three financial websites in the world, with more than 46 million monthly visitors. Therefore, investing.com established itself as a trustworthy stock market data publisher over the year.

The Hindu is an English daily newspaper based in Chennai, Tamil Nadu, India. It was founded in 1878 and is owned by The Hindu Group. It is one of India's major and second most widely circulated English-language newspapers, and it publishes various sections, including a section on economics. The Financial Times is a British daily newspaper focusing on business and current economic affairs. It was initially launched as the London Financial Guide on January 10, 1887, and was later renamed the Financial Times on February 13 of the same year. According to the Global Capital Market Survey, the Financial Times is considered an essential business publication by more than 36% of the sample population. It is the most credible publication for reporting financial and economic issues among the worldwide professional investment community.

Similarly, the fourth data source for this study is Money Control, a financial portal launched in late 1999. It receives over 17 million visitors monthly across all its platforms (web, mobile, and tablet), making it India's largest online financial platform. Thus, this financial portal becomes India's biggest store of financial news that will help investors invest their money wisely. Hence, this study has considered all the renowned and trustworthy sources in stock prices and economic articles.

3.6.2.1 Data Sources for Foreign IT indices

The study has considered the information technology (IT) indices from Germany, the UK, the USA, France, India, and Japan. The following section will provide the specific indices and their data sources.

Table 3.1 The data source for foreign IT indices

Country	IT Indices	Data Source
Japan	Tokyo SE TOPIX17 IT & Services Stock Price (ITSV17.T)	https://www.investing.com
India	Nifty IT (NIFTYIT)	https://www.investing.com
France	CAC Technology (FRTEC)	https://www.investing.com
USA	S&P 500 Information Technology (SPLRCT)	https://www.investing.com
UK	FTSE 350 Software & Computer Services (FTNMX101010)	https://www.investing.com
Germany	DAX Software (CXPSX)	https://www.investing.com

As shown in Table 3.1, the first column lists the countries that were studied in this research. The next column contains the stock market indices representing the IT stock markets for each country, and the final column indicates the data sources from which the information was collected. The DAX Software index, which is listed in the second column, represents the IT sector of the German stock market. It comprises the 32 largest German software companies on the Frankfurt Stock Exchange. It is considered a leading stock index in Germany, serving as a proxy for the performance of the German economy. The Tokyo SE TOPIX17 IT & Services Stock Price index, also listed in the second column, represents the IT sector of the Japanese stock market. The Tokyo Price Index, or TOPIX, is published by the Tokyo Stock Exchange, and the TOPIX-17 series of indices is created by dividing the TOPIX constituents into 17 categories based on the 33 sectors defined by the Securities and Identification Code Committee (SICC). The TOPIX17 IT & Services index is one of these categories, and it consists of 510 components representing the information and communication, services, and other product sectors.

Additionally, the CAC Technology index represents the French IT stock market. CAC stands for "Cotation Assistée en Continu," which translates to "Continuous Assisted Trading" in English. It provides insight into the direction of Euronext Paris, the largest stock exchange in France. The CAC Technology (FRTEC) index specifically represents the IT sector of France and consists of the top 37 French technology companies.

Furthermore, this study used the daily closing price of the S&P 500 Information Technology index as a representation of the US IT stock market. The S&P 500 index measures the performance of approximately 500 companies in the US and includes 11 sectors to provide a comprehensive overview of the US economy. The Information Technology sector of the S&P 500, also known as the S&P 500 Information Technology index, consists of 58 companies classified as members of the GICS Information Technology sector. This study has also used the FTSE 350 Software & Computer Services index data to represent the UK IT stock market. The Financial Times Stock Exchange (FTSE) group is a financial organization that specializes in creating index offerings for global financial markets. The FTSE 350 index covers large and mid-cap stocks traded on the London Stock Exchange, and the FTSE 350 Software & Computer Services index specifically tracks the five major companies in the IT sector. Finally, this study used the Nifty IT index to represent the Indian IT stock market, which consists of the top 10 IT companies listed on the National Stock Exchange.

After collecting the closing prices for all these IT indices from Investing.com, this study analyzed the impact of the German, Japanese, French, US, and UK IT stock markets on the Indian IT stock market.

3.6.2.2 Data Sources for currency exchange rates

The study has examined the currency exchange rates of Singapore, Japan, the United Kingdom, Germany, France, Mauritius, and the United States. The data sources for the currency exchange rates are listed below.

Table 3.2 Data sources for currency exchange rates

Country	Exchange Rate	Data Source
France	EUR-INR	https://www.investing.com
Japan	JPY-INR	https://www.investing.com
Mauritius	MUR-INR	Google Finance
Singapore	SGD-INR	https://www.investing.com
UK	EUR-INR	https://www.investing.com
USA	USD-INR	https://www.investing.com

As shown in Table 3.2, the first column lists the countries that were studied in this research, while the second column contains the currency exchange rates for each of these countries, and the final column indicates the data sources from which the information was obtained. In this study, the focus was on the impact of foreign markets on the Indian IT stock market, so the currency exchange rates listed in Table 2 are for foreign currencies being converted to the Indian currency (INR) rather than the other way around. The countries included in this study are Germany (EUR), France (EUR), Japan (JPY), Mauritius (MUR), Singapore (SGD), the UK (EUR), and the US (USD). The reason for considering these countries is that their currencies (EUR, USD, GBP, and JPY) have an impact on the Indian stock market (Bhattacharjee & Das, 2020; Manu & Bhaskar, 2018), and Indian IT companies are also heavily invested in Singapore, Mauritius, the Netherlands, and Germany (INDIA: FOREIGN INVESTMENT, undated). Therefore, this study analyzed the impact of the currency exchange rates for Germany (EUR), France (EUR), Japan (JPY), Mauritius (MUR), Singapore (SGD), the UK (EUR), and the US (USD) to India (INR) on the Indian IT stock market.

3.6.2.3 Data Sources for economic articles

Subscribing to financial news from multiple journals produces more accurate stock price predictions than using a single source of financial news (Hicham & Salah-Ddine, 2021). As a result, this study also looked at economic articles from various sources. The following are the specifics of these sources:

Table 3.3 Data sources for economic articles

Source Name	Data Source
Financial Times	www.ft.com
Money Control	www.moneycontrol.com/news/business/economy
The Hindu	www.thehindu.com/business/Economy/

Table 3.3 shows the first column lists the economic sources studied in this research to assess the impact on the Indian IT stock market. The second column indicates the data sources from which the economic articles were obtained. This study considered three economic article sources: The Hindu, Financial Times, and Money Control. These economic articles were then represented in terms of named entities such as CARDINAL, DATE, EVENT, FAC, GPE, LANGUAGE, LAW, LOC, MONEY, NORP, ORDINAL,

ORG, PERCENT, PERSON, PRODUCT, QUANTITY, TIME, and WORK_OF_ART, as well as sentiments such as POSITIVE and NEGATIVE, using the Spacy library (Honnibal, 2015) and Flair (Akbik et al., 2019) respectively. The study has then analyzed the impact of these named entities and sentiments on the Indian IT stock market.

3.6.3 Data Description

The datasets usually contain data and additional information. The additional information enables researchers to gain more insight into the data. It typically contains the purpose, context, and meaning of the data. This type of information is known as metadata or data description (Rosli et al., 2016). This research has considered the data from foreign IT indices, currency exchange rates, and economic articles. As mentioned above, each of these data points has been collected from different sources and thus has a different structure. For instance, foreign IT indices are mainly considered the daily stock prices, so they have one type of structure, whereas currency exchange rates are the daily exchange rates. Thus, they have another type of structure. On the other hand, the economic articles are more textual data, thus they must be represented in a totally different way. The following section details the structure of this study's data points.

3.6.3.1 Foreign IT indices

Foreign IT indices are one of the data points that this study has analyzed. It has considered the information technology indices from Germany, the United Kingdom, the United States, France, India, and Japan. As mentioned above in the data source section, the data for these indices were obtained from investing.com. Due to the same data source, all indices mentioned above follow the same data structure.

Table 3.4 Data description for foreign IT indices

Data Field Name	Data Field Description
Price	The closing price of the index on that particular date
Open	The opening price of the index for that particular date
High	The highest price of the index for that particular date
Low	The lowest price of the index for that particular date
Vol.	The traded volume of the Index for that particular date
Change%	Change in the Index price from the previous day

Instead of considering all the fields, the research has only used the daily closing price of an IT index. The closing price of a stock is the last price anyone pays for a stock. However, the closing price of an index is the average of the last prices of each stock tracked in that index. Thus, the daily closing price gives us a good indication of how an index performed at the end of the day. Therefore, this research has analyzed the impact of the closing prices of foreign IT indices on the Indian IT index.

3.6.3.2 Currency exchange rate

The currency exchange rate is the second data point that this study has analyzed. As mentioned in the data source section, the study has considered currency exchange rates from Singapore, Japan, the United Kingdom, Germany, France, Mauritius, and the United States. The data source for all the currencies mentioned above exchanges is investing.com. Since the data source for all the currency exchange rates is the same, their data structures are also the same.

Table 3.5 Data description for the currency exchange rate

Field Name	Field Description
Date	Reported date for exchange rate as per INR on that date
Price	Closing Price of the exchange rate as per INR on that date

Open	The opening price of the exchange rate as per INR on that date
High	The highest price of the exchange rate as per INR on that date
Low	The lowest price of the exchange rate as per INR on that date
Change%	Change in the exchange rate price from the previous day

Like the foreign IT indices, this study has also considered the daily closing price of the currency exchange rate. As mentioned above, the daily closing price indicates the rise and fall in the foreign exchange rate for a given currency. Thus, along with the daily closing price of foreign IT indices, this study has also analyzed the impact of the daily closing price of the currency exchange rate on the Indian IT stock market.

3.6.3.3 Economic articles

The study obtained financial news from three economic sources: The Hindu, The Financial Times, and Money Control. Directly, the economic articles cannot be measured because they are unstructured and textual. Thus, the research has operationalized the economic articles. The process of transforming an abstract concept into measurable observations is known as operationalization (Blackstone, 2018). The study measured economic sentiments using the count of named entities and news sentiments.

Furthermore, named entities were extracted from the Spacy library (Honnibal, 2015), and article sentiments were obtained from Flair. Flair is a natural language processing (NLP) framework that enables researchers to use cutting-edge models for language modeling and text classification (Akbik et al., 2019). This study has considered the 18 named entities and an article's positive and negative sentiments. The data description details for all these named entities and news sentiments have been provided as follows:

Table 3.6 Data description for named entities.

Named Entity	Description
CARDINAL	Numerals do not fall into another entity
DATE	Absolute or relative dates
EVENT	Named battles, wars, sports, events, etc.
FAC	Buildings, airports, highways, bridges, etc.
GPE	Countries, cities, states
LANGUAGE	Any Named language
LAW	Named documents made into laws
LOC	Non-GPE locations, mountains, water bodies.
MONEY	Monetary value including units.
NORP	Nationalities or religious or political groups.
ORDINAL	First, second, etc.
ORG	Companies, agencies, institutions, etc.
PERCENT	Percentage, including %.
PERSON	People, including fictional.
PRODUCT	Objects, Vehicles, Foods, etc. (Not services)
QUANTITY	Measurement, as of weight or distance.
TIME	Time smaller than a day
WORK_OF_ART	Title of books, songs, etc.

Table 3.7 Data description for named entities.

Sentiment	Description
Positive	If an article conveys a positive feeling such as enthusiasm, happiness, or excitement.
Negative	If an article conveys a negative feeling such as anger, annoyance, or frustration.

The research data set was gathered for all the above variables between January 01, 2018, and December 31, 2021. The information was gathered daily. The daily closing price was considered for foreign IT indices and currency exchange rates. At the same time, 18 named entities and their positive and negative sentiments were retrieved for economic articles.

Moreover, the closing prices of indices and currency exchanges are not reported on weekends and public holidays. Additionally, there are a few articles where spacy-based models could not extract any of the 18 named entities. Hence, while intersecting the data daily, a few records have been dropped; thus, the data count has been reduced to 876 records. Therefore, the study used the resulting data to assess the impact of foreign IT indices, exchange rates, and economic events on Indian IT stocks.

3.7 Methods of data analysis

Data analysis is the process of using statistical and logical techniques to evaluate, clean, transform, and pre-process raw data and to draw inductive inferences from it (Bouikidis, 2018). There are two main methods for data analysis: qualitative data analysis and quantitative data analysis. Qualitative data analysis involves using techniques such as interviews, focus groups, and experiments to identify common response patterns and critically analyze them to achieve research aims and objectives. Quantitative data analysis, on the other hand, involves critically analyzing and interpreting numerical data using statistical or algorithmic techniques to understand the reasoning behind the main findings (Kucharavy et al., 2020).

The study used the quantitative data analysis method. After collecting the necessary data within the specified timeframe, the study pre-processed it to make it suitable for analysis. To obtain more convincing and robust results, The study has employed machine-learning and deep-learning algorithms to detect the causal relationship between foreign stock markets, currency exchange rates, economic articles, and the Indian IT stock market. The following sections describe the pre-processing and data analysis techniques in more detail.

3.7.1 Data Pre-Processing

As mentioned previously, this study applied statistical, machine learning, and deep learning algorithms to analyze the impact of various factors on the Indian IT stock market. These algorithms are susceptible to the given data and can produce inaccurate results if the data is irrelevant, redundant, noisy, or unreliable (Kotsiantis & Kanellopoulos, 2006). Therefore, it is essential to devote time to the processing and preparing the data for statistical, machine-learning, and deep-learning algorithms. Data processing involves techniques such as data cleaning, normalization, transformation, feature extraction, and selection; these techniques are referred to as "data pre-processing techniques" because they are applied before the data is fed to the model.

Since this study involved analyzing the impact of foreign IT stock markets, currency exchange rates, and economic articles on the Indian IT stock market, data pre-processing was applied to these factors to convert them into numerical form. The foreign IT stock market and currency exchange rate data were already in numerical form, so the pre-processing was only applied to the economic articles, which were then merged with the foreign IT stock market and currency exchange rate data. The details of the pre-processing techniques are discussed in the following sections.

3.7.1.1 Economic-articles pre-processing

This study collected economic articles from three different sources: Money Control, The Hindu, and The Financial Times. The economic articles were represented as the combination of the article's title and headline and were transformed into positive and negative sentiments using the Flair library (Akbik et al., 2019). Additionally, named entities - real-world objects with a name - were identified using the Spacy library

(Honnibal, 2015). As shown in Table 8, this study considered 18 named entities and two sentiments (i.e., positive and negative) to represent the economic articles from the three news sources. Since multiple articles were published on a single day for Money Control, The Hindu, and The Financial Times, this study calculated the cumulative sum of all 20 variables (18 named entities and two sentiments) for all articles published on these news sources per day. These variables were then used to represent the economic articles, and their impact on the Indian IT stock market was analyzed.

Table 3.8 Independent variable that represents the economic articles.

Named Entities and Sentiment Representation of economic articles		
Money Control	The Hindu	Financial Times
CARDINAL_MoneyControl	CARDINAL_TheHindu	CARDINAL
DATE_MoneyControl	DATE_TheHindu	DATE
EVENT_MoneyControl	EVENT_TheHindu	EVENT
FAC_MoneyControl	FAC_TheHindu	FAC
GPE_MoneyControl	GPE_TheHindu	GPE
LANGUAGE_MoneyControl	LANGUAGE_TheHindu	LANGUAGE
LAW_MoneyControl	LAW_TheHindu	LAW
LOC_MoneyControl	LOC_TheHindu	LOC
MONEY_MoneyControl	MONEY_TheHindu	MONEY
NORP_MoneyControl	NORP_TheHindu	NORP
ORDINAL_MoneyControl	ORDINAL_TheHindu	ORDINAL
ORG_MoneyControl	ORG_TheHindu	ORG
PERCENT_MoneyControl	PERCENT_TheHindu	PERCENT
PERSON_MoneyControl	PERSON_TheHindu	PERSON
PRODUCT_MoneyControl	PRODUCT_TheHindu	PRODUCT
QUANTITY_MoneyControl	QUANTITY_TheHindu	QUANTITY
TIME_MoneyControl	TIME_TheHindu	TIME
WORK_OF_ART_MoneyControl	WORK_OF_ART_TheHindu	WORK_OF_ART
POSITIVE_MoneyControl	POSITIVE_TheHindu	POSITIVE
NEGATIVE_MoneyControl	NEGATIVE_TheHindu	NEGATIVE

As shown in Table 3.8, this study has differentiated the named entities and sentiments for all three economic article sources by adding suffixes indicating the source of the articles. Thus, all 18 named entities and two sentiments captured from articles published on Money Control have been suffixed with _MoneyControl. Similarly, the 18 named entities and two sentiments captured from The Hindu have been suffixed with _The Hindu. On the other hand, the named entities and sentiments captured from the Financial Times have not been suffixed with anything.

3.7.1.2 Merging data from different sources

Once the data for foreign IT indices, currency exchange rates, and economic articles was collected, the study merged the data based on date. The study considered data points from articles from Money Control, The Hindu, and The Financial Times and merged them based on the date. The merged articles were then combined with foreign IT indices and currency exchange rates, again based on the date. However, the study dropped those records if any foreign IT indices, currency exchange rates, or economic articles were absent during this process. Therefore, this research considered only those records where all three features (foreign IT indices, currency exchange rates, and economic articles) were represented by 71 variables, and their total count was 876 records. As a result, this study was conducted on 876 records to analyze the impact of 71 variables representing foreign IT stock markets, currency exchange rates, and economic articles on the Indian IT stock market.

3.7.1.3 Data imputation

Missing data is a common issue that can arise in a machine learning setting, and imputing this missing data is an important task, especially when data is scarce, and it is essential to use all available data (Bertsimas et al., 2018; Jerez et al., 2010). While preparing the data for this research, there were a few days when some of the named entities needed to be identified in the economic articles by the Spacy library (Honnibal, 2015). The named entities with missing values were LANGUAGE Hindu, MONEY Hindu, LOC, and NORP. Despite having missing values, the other named entities, foreign IT stock market data, and currency exchange rates were still present, and this study wanted to retain this information. Therefore, rather than dropping records with a few missing values, this study imputed them with 0. Finally, this study considered all 876 records, where a value of 0 indicates the absence of a value for a particular entity. After that, this study applied statistical, machine learning, and deep-learning causality detection techniques to analyze the impact of 71 variables (representing foreign IT stock markets, currency exchange rates, and economic articles) on the Indian IT stock market.

Python Implementation

```
sentiment_all_exch_rate_all_indices_Df=sentiment_all_exch_rate_all_indices_Df.fillna(0)
```

Figure 3.2 Missing value imputation

As shown in Figure 3, this study has used fillna methods of pandas data frame to replace all the un-available values to zero, which means that where the independent variable (representation of foreign IT stock market, currency exchange rate, and new articles) has the missing value, it has been replaced by zero.

3.7.1.4 Data Standardization

Standardization is a process that involves subtracting the mean from a feature and dividing it by the standard deviation, also known as the z-score (Jamal et al., 2014). This technique is useful when the data being analyzed has varying scales and the algorithms assume that the data has a Gaussian distribution. In this research, the closing prices of foreign IT indices in dollars, the closing prices for currency exchange rates in the form of ratios, named entities in the form of summation, and article sentiments all range from 0 to 1. As a result, all the variables considered in this research are on different scales. When variables have different scales, machine learning and deep learning models may assign higher weights to variables with more minor scales and vice versa. Additionally, these models may take longer to train when variables are on different scales. Therefore, this study applied normalization to all 71 input variables (foreign IT stock markets, currency exchange rates, and economic articles) to bring the data to a uniform scale and reduce training time (L. Huang et al., 2020) for machine learning and deep learning methods.

Python Implementation

```
#### Method to scaling the dataframe
def dataframe_scaler(df,scaler_method='standard'):
  normalized_df=df.head(1)
  if 'NEWS_DATE' in df.columns:
    df.drop(columns=['NEWS_DATE'],inplace=True)
  print(scaler_method)
  if scaler_method.upper()=='STANDARD':
    print('apply standardization')
    normalized_df=(df-df.mean())/df.std()
  else:
    normalized_df=(df-df.min())/(df.max()-df.min())
  return normalized_df
```

Figure 3.3 Scaler method

Source:(Jamal et al., 2014)

As depicted in the code snippet in Figure 3.3, this study utilized the dataframe_scalar method to standardize all the independent variables (representing foreign stock markets, currency exchange rates, and economic articles) to the same scale. Machine learning and deep learning models may take longer when variables have different scales to converge on the optimal solution. Therefore, this study applied the standardization technique to transform these variables to the same scale to facilitate more efficient training.

3.7.2 Tools and Libraries

The study has fetched economic articles from Money Control, The Hindu, and the Financial Times using the BeautifulSoup library. The beautiful soup is a Python library used for pulling data from websites based on XML (Extensible Markup Language) and HTML (Hypertext Markup Language) files (Richardson, 2007). After that, Spacy and Flair's library was used to convert articles to named entities and sentiments. The transformed data were then analyzed using statistical, machine learning, and deep-learning-based causality detection methods.

The activities described above were performed in the Jupyter Lab using the Python programming. JupyterLab is an open-source, web-based interactive environment for notebooks, code, and data. It is a flexible framework that allows users to configure and arrange workflows for data science and machine learning use cases (Kluyver et al., 2016).

The research study was coded in a Jupyter notebook for collecting, processing, and applying statistical, machine learning, and deep-learning-based algorithms to confirm the causality. The code in a Jupyter notebook is organized into cells, which can be individually modified and run. The output of each cell appears just below it and is stored as part of the notebook. Although Jupyter Lab supports different programming languages, over 100 languages include Python, Java, R, Matlab, Octave, etc. Nevertheless, this study was coded in Python because it has very concise syntax, extensive availability of libraries, and ample community support (van Rossum & Python Development Team, 2018). The Jupyter notebook records the order of the steps performed while conducting research.

Moreover, tools like nbviewer and binder enable sharing of statistically rendered jupyter notebooks along with the computation environment where users can reproduce the previous executions (Kluyver et al., 2016). Thus, due to the usage of Jupyter Lab execution, the results produced in this research were shareable and reproducible (reliable).

3.8 Methods

The primary goal of this research is to examine the impact of foreign IT indices, currency exchange rates, and economic articles on the Indian IT stock using ensemble model (statistical, machine learning, and deep-learning-based) methods. The preceding sections discussed collecting, processing, and cleaning the necessary data. However, this section will also go over the various machine-learning and deep-learning methods used for research and how the voting mechanism will work.

3.8.1 NonLin Causality

NonLinCausality is a Python package that analyses causality using deep-learning methods. The package supports deep-learning methods such as Long-Term Short-Term Memory (LSTM), Gated-Recurrent Unit (GRU), and Multilayer Perceptron (MLP). These methods performed better than Granger causality at detecting non-linear causal relationships, as (Rosoł et al., 2022) reported. Hence, this research used the NonLinCausality package for analyzing the impact of foreign IT stocks, currency exchange rates, and economic articles on Indian IT stocks. Moreover, this study has only considered the LSTM and GRU variants of NonLinCausality due to their ability to capture long-term dependencies. The following section details the LSTM and GRU methods along with the MLP.

3.8.1.1 Multilayer Perceptron (MLP)

Prior to the development of multilayer perceptron, linear models were used to solve classification and regression problems. However, linear models were not able to represent all the functions, such as quadratic functions for linear regression and XOR functions for the linear classifiers. Hence, a multilayer perceptron was introduced, representing any complex non-linear function by connecting many simple processing units into a neural network. Each unit in MLP is known as a perceptron. In aggregate, these units can approximate or compute surprisingly complex functions. Hence, a multilayer perceptron is known as a universal approximator. The universal approximators can approximate any continuous function.

Figure 3 illustrates the perceptron, which receives features as input. Each of the inputs has an associated weight. The features, along with their associated weights, have been fed to the step function.

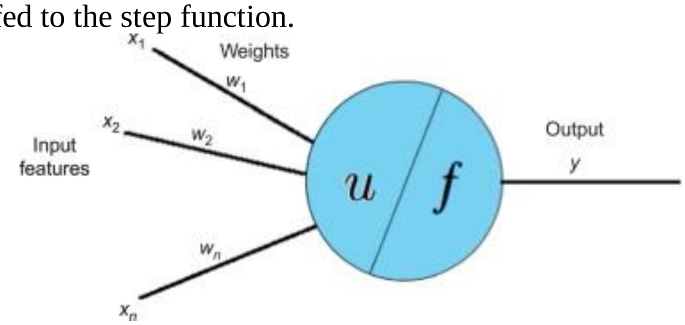

Figure 3.4 Perceptron Source:(Menzies et al., 2015)

The step function is represented by equation 1. As per the equation, if the value of a function (which can be thought of as a multiplication of input features and weights) is greater than the threshold (generally it is 0.5), then the output of perceptron is 1, else it is 0 (Menzies et al., 2015).

$$y = f(u(x)) = f(x) = \begin{cases} 1, & if u(x) > \theta \\ 0, & otherwise, \end{cases}$$

$$where\ \theta\ is\ the\ threshold$$

Equation 3.1 Step function

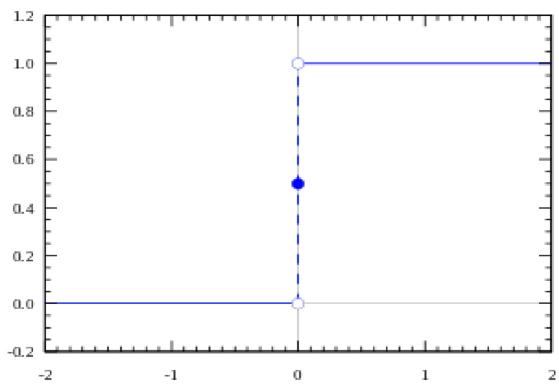

Figure 3.5. Step function

source: https://images.app.goo.gl/wv5EZR2ZY6QjG3dU7

As shown in the figure, the step function perceptron can only separate linear data. However, when these simple step functions are aggregated, MLPs can approximate any continuous function. This aggregation is achieved by combining several neurons into three layers, i.e., an input layer, an output layer, and a hidden layer.

- The input layer is fed with the data to be processed.
- The output layer performed the classification or regression task.
- A multilayer perceptron can have an arbitrary number of hidden layers in-between the input and the output layer.

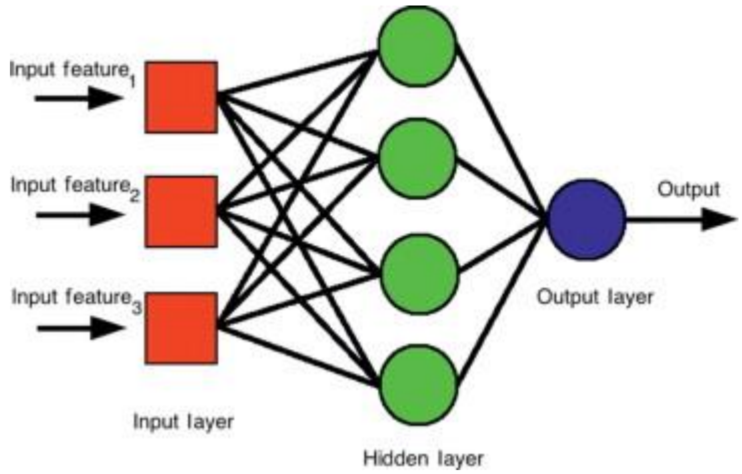

Figure 3.6 MLP architecture source:(Menzies et al., 2015)

In MLP, data flows from the input to the output layer. However, the actual weight adjustment or training occurs during back-propagation. The MLP network can approximate any continuous function and solve non-linear separable problems after the appropriate number of back-propagations (Meyer-Baese & Schmid, 2014). However, these networks are not preferred for time-series data. Because they can only forecast using a fixed number of timesteps, this research had also been implemented on the time-series data; thus, it still needed to consider the MLP implementation.

Python Implementation

```
import nonlincausality as nlc
results = nlc.nonlincausalitymeasureMLP(x=np.array(df_train), maxlag=lags,window=20,step=1,
                                        Dense_layers=2, Dense_neurons=[100, 100],
                                        x_test=np.array(df_test), run=1, add_Dropout=True,
                                        Dropout_rate=0.01, epochs_num=[50, 100],
                                        learning_rate=[0.001, 0.0001], batch_size_num=128,
                                        verbose=True, plot=True)
```

Figure 3.7 MLP implementation using nonlincausality framework.

Figure 3.7 illustrates the multilayer perceptron implementation of non-linear causality. The MLP implementation has several hyperparameters (which can be controlled by the implementer), such as max lag (the maximum number of prior values in the time series to be used), Dense Layer (the number of layers of the neural network to be experimented with), Dense neurons (the number of neurons in each dense layer of the MLP), add Dropout (whether to drop out a portion of the neurons in each layer), Dropout rate (in case add Dropout is True, the fraction of neurons to be dropped out in each layer), epochs_num (the number of iterations the neural network should perform in order to converge), and learning rate (the size of the step the network will take in order to converge).

3.8.1.2 Long-Term Short-Term Memory (LSTM)

LSTM networks are recurrent neural network variants (RNN). The LSTM was first introduced in 1997 (Hochreiter & Schmidhuber, 1997). The LSTM network addresses RNN's limitation of being unable to learn long-term dependencies. The LSTM network, on the other hand, has the unique capability of remembering long-term dependencies. As a result, the LSTM network is an obvious choice for processing, prediction, and classification of time series data. In addition, unlike traditional feed-forward networks, it has feedback connections. As a result, it can process the entire sequence of data and individual data points.

LSTM networks consist of LSTM cells rather than standard neural network layers. These LSTM cells are primarily composed of the three components listed below.

- Input gate
- Forget gate.
- Output gate

The graphical representation of LSTM is shown below.

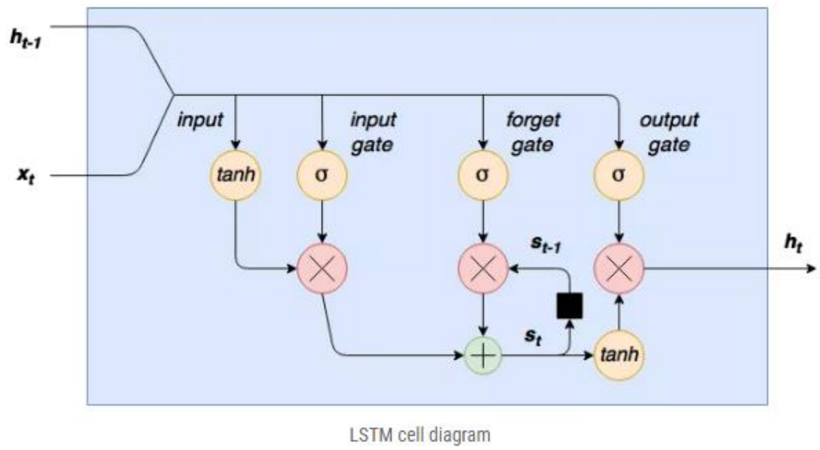

Figure 3.8 LSTM architecture

source: https://adventuresinmachinelearning.com/keras-lstm-tutorial/

As illustrated in Figure 3.8, LSTM has used majorly two activation functions: the sigmoid and the tanh activation. The details on these activation functions are given in the later part of the chapter. Now let us look at the architecture of LSTM. On the left side, the LSTM cell is fed with a new word/sequence value $\boldsymbol{X_t}$ and concatenated with the previous output from the cell $\boldsymbol{h_{t-1}}$.

The combined input is then squeezed between -1 and 1 via the tanH layer. After that, the input is fed to the input gate. The input gate is a layer of sigmoid-activated nodes. These nodes' output is multiplied by the tanH layer's output. Because the sigmoid activation function can have values ranging from 0 to 1, an input gate can turn on or off specific input values.

The forget gate is the next step in the information flow through the cell. The LSTM cell has two internal states: current and previous. Both states are combined with the input data to form the effective layer of recurrence. Using addition rather than multiplication reduces the risk of a vanishing gradient. This layer assists the network in determining which state variables should be remembered or forgotten.

The output layer of an LSTM cell is the penultimate layer. It is made up of the tanH activation function. The output of the tanH function is controlled by the output gate. This gate controls the values that can be output from the cell. The LSTM cell's math can be expressed as follows:

Input Gate

The tanH activation function limits input to the LSTM cell between -1 and 1. which can be expressed as follows:

$$g = tanh(b^g + x_t U^g + h_{t-1} V^g)$$

Equation 3.2 LSTM's input gate

U^g: $weight\ for\ the\ input$

V^g: $weight\ for\ the\ previous\ cell\ output$

b^g: $Input\ bias$

The output of the input gate, which is based on the sigmoid function, is expressed as follows:

$$i = \sigma\left(b^i + x_t U^i + h_{t-1} V^i\right)$$

Equation 3.3 LSTM's output of the input gate

U^i: $weight\ for\ the\ input$

V^i: $weight\ for\ the\ previous\ cell\ output$

b^i: $Input\ bias$

Finally, the squashed input is elementwise multiplied by the output of the input gate. In this operation, the value of I determines the importance given to the current time step.

$$g * i$$

Equation 3.4 LSTM's input gate element-wise multiplication

$$* \ is\ the\ operator\ for\ element - wise\ multiplication$$

Forget Gate

The output of the forget gate is expressed as follows:

$$f = \sigma(b^f + x_t U^f + h_{t-1} V^f)$$

Equation 3.5 LSTMs need to remember gate output.

$U^f: weight\ for\ the\ input$

$V^f: weight\ for\ the\ previous\ cell\ output$

$b^f: Input\ bias$

The output of the forget gate is multiplied by the previous state value and added to the output of the input gate. This calculation determines how much weight is assigned to previous time steps, effectively determining how much information from previous time steps the network wants to forget and how much information about current time steps it wants to retain. This arrangement of the LSTM network helped the study learn important patterns and forget unimportant ones of foreign IT stock markets, currency exchange rates, and economic articles to identify their impact on the Indian IT stock market.

$$s_t = \ s_{t-1} * f + g * i$$

Equation 3.6. LSTMs forget gate and input gate multiplication.

Output Gate

The output gate is expressed as follows:

$$o = \sigma(b^o + x_t U^o + h_{t-1} V^o)$$

Equation 3.7 LSTM’s output gate

$U^o: weight\ for\ the\ input$

$V^o: weight\ for\ the\ previous\ cell\ output$

b^{o}: *Input bias*

The forget gate output is squeezed between -1 and 1 using the tanH activation function. The resultant is then multiplied by the sigmoid function output o at the output gate.

$$h_{t=}tanh\,(s_t)\ * o$$

Equation 3.8 LSTM's final output

In summary, the LSTM network learns how much previous information should be forgotten and how much emphasis should be given to the current information. Thus, In the context of this research, LSTM implementation of non-causality had learned how much information to be forgotten from the earlier instance and how much importance to be given to the current instances of foreign IT indices, currency exchange rate, and economic articles to analysis the impact on the Indian IT stocks.

Python Implementation

```
import nonlincausality as nlc
results = nlc.nonlincausalityLSTM(x=np.array(df_train),LSTM_layers=2, LSTM_neurons=[25, 25],
                                  maxlag=lags,window=20,step=1, Dense_layers=2,
                                  Dense_neurons=[100, 100], x_test=np.array(df_test),
                                  run=3, add_Dropout=True, Dropout_rate=0.01,
                                  epochs_num=[50, 100], learning_rate=[0.001, 0.0001],
                                  batch_size_num=128, verbose=True, plot=True)
```

Figure 3.9 LSTM implementation using nonlincausality framework.

Figure 3.9 illustrates the implementation of non-linear causality using a long-short-term memory (LSTM) network. The LSTM implementation has several hyperparameters that the implementer can control, including maxlag (the maximum number of prior values in the time series to be used), LSTM_layers (the number of LSTM layers to experiment with), LSTM_neurons (the number of LSTM cells in each LSTM layer), Dense_Layer (the number of neural network layers after the LSTM layers), Dense_neurons (the number of neurons in each dense layer), run (the number of repetitions of the neural network training process to obtain the network with the lowest residual sum of squares (RSS) value), add_Dropout (whether to drop out a portion of the neurons in each layer for regularization), Dropout_rate (if add_Dropout is True, the fraction of neurons to be dropped out in each layer), epochs_num (the number of iterations the neural network should perform in order to converge), and learning_rate (the learning rate to be used in the training process).

3.8.1.3 Gated Recurrent Unit (GRU)

A gated recurrent unit (GRU) is another variant of a recurrent neural network (RNN) (Cho et al., 2014). Like LSTM, it also intends to adjust neural network weights to solve the vanishing gradient problem. Like LSTM, the GRU is different from RNN in terms of supporting the gating mechanism for the hidden state. This gated mechanism controls the update and resets operations on the hidden state. For instance, if the first token is of great importance, the network will learn not to update the hidden state after the first observation.

Similarly, the network will learn to skip irrelevant temporary observations. Finally, the network will learn to reset the latent state whenever required. The basis of these update and reset operations are the following two gates:

- Update gate
- Reset gate

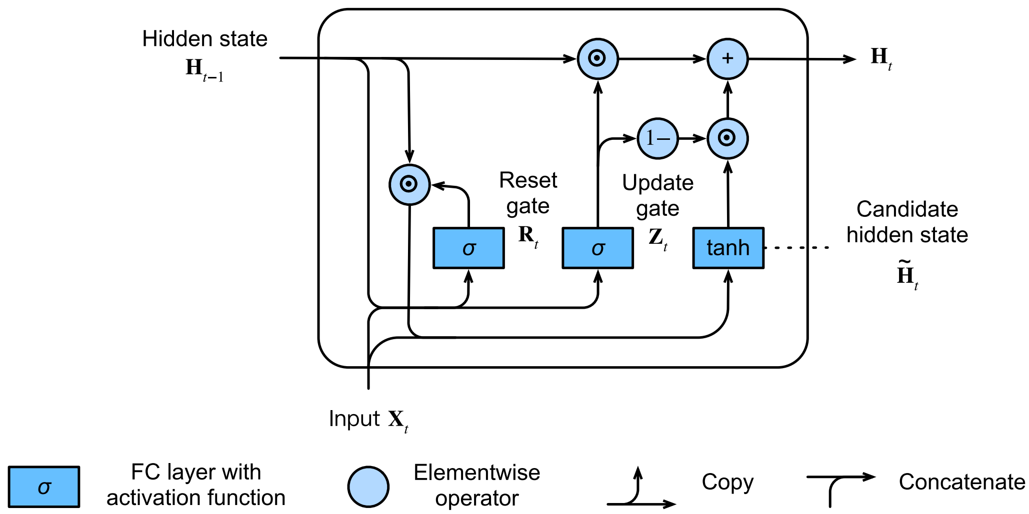

Figure 3.10 GRU architecture

source: https://d2l.ai/chapter_recurrent-modern/gru.html

Update gate

The update gate in the GRU network is equivalent to the forget gate and input gate in the LSTM network. The update gate determines how much information should be forgotten and which new information from a specific point in time should be considered for further calculation. The calculation of the update gate is based on the sigmoid function.

$$u^{(t)} = \sigma(b^{\mathrm{u}} + x_t U^{\mathrm{u}} + h_{t-1}\mathrm{W}^{\mathrm{u}})$$

Equation 3.9 GRU's update gate.

$U^u: weight\ for\ the\ input$

$\mathrm{W}^u: weight\ for\ the\ previous\ cell\ output$

$b^{\mathrm{u}}: Input\ bias$

Reset Gate.

On the other hand, the reset gate controls the amount of information from the previous state used to compute the candidate state. The following expression can express the reset gate operation:

$$r^{(t)} = \sigma(b^r + x_t U^r + h_{t-1} W^r)$$

Equation 3.10 GRU's reset gate.

$U^{\mathrm{r}}: weight\ for\ the\ input$

$\mathrm{W}^{\mathrm{r}}: weight\ for\ the\ previous\ cell\ output$

$b^{\mathrm{r}}: Input\ bias$

The output of the reset gate is used to calculate the candidate state. The candidate state is then used to compute the GRU cell state at a given time. The candidate state is calculated by using the hyperbolic tangent function.

$$h^t = tanh(b^{\mathrm{h}} + x_t U^h + r^t h_{t-1} \mathrm{W}^{\mathrm{h}})$$

Equation 3.11 GRU's candidate state

$U^{\mathrm{h}}: weight\ for\ the\ input$

$\mathrm{W}^{\mathrm{h}}: weight\ for\ the\ previous\ cell\ output$

$b^{\mathrm{h}}: Input\ bias$

The value of the state h^{t} is computed using the update gate value. The following is the formulation of the same.

$$h^t = (1 - u^t)h^{t-1} + u^t h^t$$

Equation 3.12 GRU's final output

With the help of the reset and update gates, the GRU networks can forget and remember the information for successive time steps. Although GRU and LSTM are unaffected by the vanishing and exploding gradients, the GRU network has an advantage over the LSTM network in terms of fewer parameters and less computational complexity.

Python Implementation

```
import nonlincausality as nlc
results = nlc.nonlincausalityGRU(x=np.array(df_train),GRU_layers=2, GRU_neurons=[25, 25],
                                 maxlag=lags,window=20,step=1, Dense_layers=2,
                                 Dense_neurons=[100, 100], x_test=np.array(df_test),
                                 run=3, add_Dropout=True, Dropout_rate=0.01,
                                 epochs_num=[50, 100], learning_rate=[0.001, 0.0001],
                                 batch_size_num=128, verbose=True, plot=True)
```

Figure 3.11 GRU implementation using nonlincausality framework.

Figure 3.11 illustrates the implementation of non-linear causality using a Gated Recurrent Unit (GRU) network. The GRU implementation has several hyperparameters that the implementer can control, including maxlag (the maximum number of prior values in the time series to be used), GRU_layers (the number of GRU layers to experiment with), GRU_neurons (the number of GRU cells in each GRU layer), Dense_Layer (the number of neural network layers after the GRU layers), Dense_neurons (the number of neurons in each dense layer), run (the number of repetitions of the neural network training process to obtain the network with the lowest residual sum of squares (RSS) value), add_Dropout (whether to drop out a portion of the neurons in each layer for regularization), Dropout_rate (if add_Dropout is True, the fraction of neurons to be dropped out in each layer), epochs_num (the number of iterations the neural network should perform in order to converge), and learning_rate (the learning rate to be used in the training process).

Like LSTM networks, GRU networks also learn how much previous information should be forgotten and how much emphasis should be given to the current information but with less complexity. Thus, this research also considered GRU implementation of non-causality, which had learned how much information to be forgotten from the earlier instance and how much importance to be given to the current instances of foreign IT indices, currency exchange rate, and economic articles to analyze the impact on the Indian IT stocks.

3.8.1.4 Activation Functions

Activation functions are the most important parts of neural network design. The choice of the activation function in the hidden layers determines how well the network learns the training dataset. The choice of activation function in the output layer determines the model's predictions (Sharma et al., 2020).

Activation functions are the mathematical equations determining whether a neuron will get activated. Generally, the output range of the activation function lies between 0 and 1, -1 and 1, or between 0 and x. At a very high level, activation functions are of 3 types.

- Binary Step Function
- Linear Activation Function
- Non-Linear Activation Function

The understanding of the activation mentioned above function is important for the context of this research. These functions are the major constituent of any deep-learning architecture, and LSTM and GRU implementation of nonLinCausality are not different. Hence, I recommend going through the details of activation functions in the following sections.

3.8.1.4.1 Binary Step Function

The binary step function is the most basic function for constraining the output values. When the input value exceeds the threshold, the threshold-based classifier activates the neuron; otherwise, the neuron will not get activated (Sharma et al., 2020).

$$f(x) = \begin{cases} 0, & x < th \\ 1, & x \geq th \end{cases}$$

Equation 3.13 Binary step function

$f(x)$: $Binary\ step\ function\ for\ input\ value\ x$

x: $Input\ to\ the\ binary\ step\ function$

th: $Threshold\ for\ the\ binary\ step\ function$

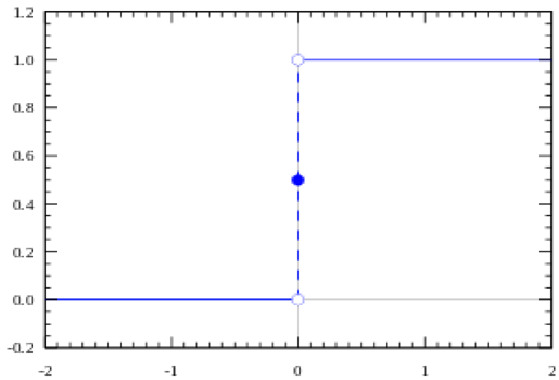

Figure 3.12 Binary step function

source: https://images.app.goo.gl/wv5EZR2ZY6QjG3dU7

Binary activation functions are limited to binary classification only, i.e., they can only distinguish between two different classes. However, they are unsuitable for use cases with more than two classes.

3.8.1.4.2 Linear activation function

The linear activation functions are directly proportional to the supplied input. The binary step function is limited to producing an output of 0 or 1. whereas the linear activation function considers the input values while producing the output. Hence, they are more expressive (Sharma et al., 2020).

$$f(x) = cx$$

Equation 3.14 Linear activation function

$f(x)$: *Linear activation function*

c: *constant value*

x: *Input to the linear activation function*

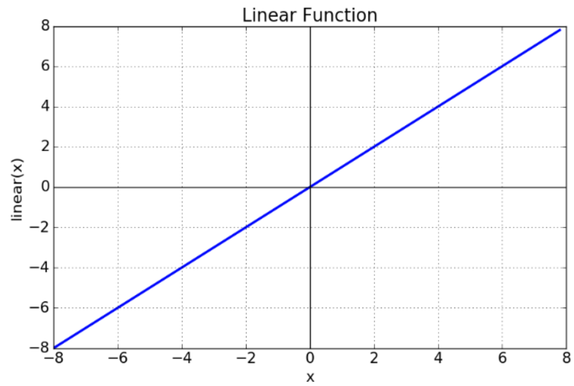

Figure 3.13 Linear activation function

source: https://images.app.goo.gl/rbiTXQCZdeToHrbp7

The following are the drawbacks associated with the linear activation functions:

Since the derivative of the linear activation function is constant, it is not dependent on the value of the input. It gets updated with the same factor during back-propagation.

Regardless of how many linear activations hidden layers are added to a neural network, The output of the last layer would not be more expressive than the first layer. In other words, all the layers in the neural network would collapse into one layer.

3.8.1.4.3 Non-Linear Activation Function

Non-linear activation functions allow the network to learn the complex mapping between input and output. These functions, in theory, can learn any composite function to match the network's output (Sharma et al., 2020). There are various kinds of non-linear activation functions.

- Sigmoid
- TanH (Hyperbolic Tangent)
- ReLu (Rectified Linear Unit)
- Leaky ReLu

- Parametric ReLu
- SoftMax

3.8.1.4.3.1 Sigmoid activation function

The most common non-linear activation function is the sigmoid activation function. It has an S-shaped curve ranging from 0 to 1 (Nwankpa et al., 2018).

$$f(x) = \frac{1}{1 + exp^{(-x)}}$$

Equation 3.15 Sigmoid activation function

$$f(x): Sigmoid\ activation\ function\ for\ given\ input\ x$$

$$x: Input\ to\ the\ activation\ function$$

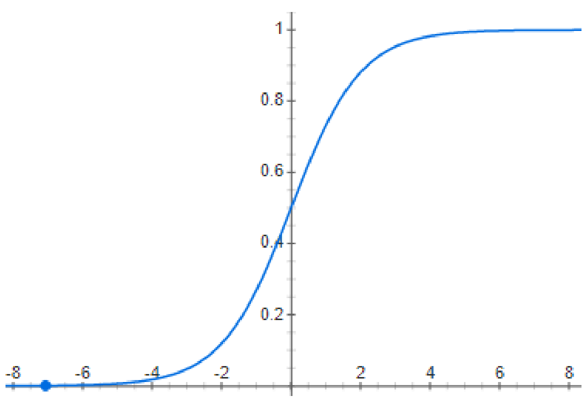

Figure 3.14 Sigmoid activation function

The following factors limit sigmoid activation functions:

Since sigmoid functions vary between 0 and 1. Hence, sigmoid functions are not symmetric about zero. The sign of all output values is the same.

Sigmoid functions also suffer from the gradient descent problem, which occurs when large input values are scaled between 0 and 1. Thus their derivative during back-propagation becomes too small to produce a satisfactory output (Hu et al., 2021).

3.8.1.4.3.2 TanH Activation Function

TanH activation functions are like sigmoid activation functions, but they are symmetric around the origin. As a result, the tanH activation function can easily model input with strong negative, positive, and neutral values. Furthermore, the tanH activation function's output ranges between -1 and 1 (Datta, 2020). The tanH activation function is expressed as follows.

$$tanH(x) = \frac{exp^{x} - exp^{-x}}{exp^{x} + exp^{-x}}$$

Equation 3.16 tanH activation function

$$tanH(x): Hyperbolic\ tangent\ activation\ function$$

$$x: Input\ to\ hyperbolic\ tangent\ activation\ function$$

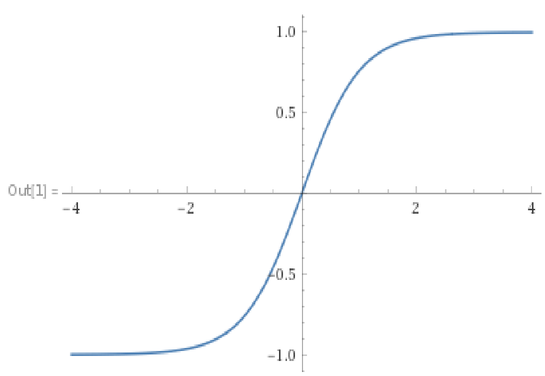

Figure 3.15 TanH activation function

source: https://www.wolframalpha.com/input/?i=tanh+function&x=0&y=0

Despite being preferred over the sigmoid activation function, the tanH activation function suffers from the vanishing gradient problem. Since it squeezes the values between $(-\infty, +\infty)$ to between (-1 to +1), the derivatives for too large and too small values become very small to give a satisfactory prediction. Moreover, they can only be used in the hidden layers because the tanH function ranges between -1 to 1. The TanH activation function is also widely used in LSTM and GRU implementation. This research has also used the LSTM and GRU implementation of nonLinCausality. Hence, the understanding of tanH activation functions is important for this research.

3.8.1.4.3.3 Rectified linear unit activation function.

The ReLu activation function is the most widely used in the neural network. The output of the ReLu function varies between 0 and x, where x is the input to the activation function. This function has much better convergence than tanH and sigmoid functions. Moreover, the ReLu does not suffer from the problem of vanishing gradients (Agarap, 2018). The mathematical equation for the ReLu function is as follows:

$$ReLu(x) = max\ (0, x)$$

Equation 3.17. ReLu activation function

$$ReLu(x): Rectified\ Linear\ activation\ function\ for\ given\ input\ x$$

$$x: Input\ to\ the\ rectified\ linear\ activation\ function$$

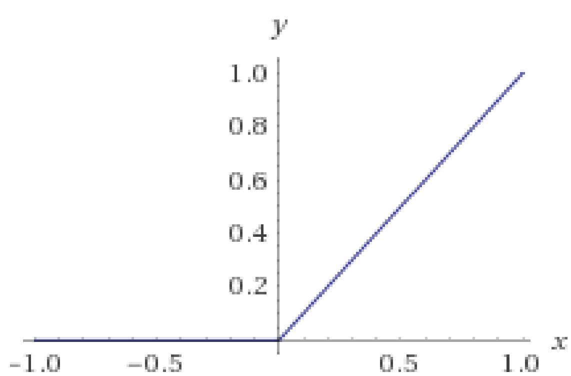

Figure 3.16 ReLu activation function

source: https://www.wolframalpha.com/input/?i=relu+function

Despite being the most popular activation function, Relu has the following drawbacks:

The output of ReLu is only sometimes centered on zero because it ranges between 0 and ∞. The non-zero-centered function's gradients are either all positive or all negative. Hence, the weight vector needs more updates to get adequately trained, which eventually increases the model training time. However, the zero-centered activation function ensures that the mean activation value is around zero, which causes activations to become normalized. Eventually, the machine learning models converge faster on normalized data (Datta, 2020).

Since Relu's activation function ranges between 0 to ∞. Hence, classification problems can only be used in the hidden layers. The ReLu activation function is only available in the hidden layers.

A large gradient flowing through the ReLu neuron may cause weight and bias to get updated in such a way that the neuron will never reactivate again on any data point. If this situation occurs, the gradient flow through that unit will be zero for the training from that point on. This type of phenomenon is also known as the dying neural problem.

3.8.1.4.3.4 Leaky Rectified linear unit activation function

The Leaky ReLu has been introduced to eliminate the dying neuron problem of the ReLu activation function. The output of Leaky ReLu lies between 0.01 and x, where x is the input to the activation function. The Leaky ReLu has a small positive value to keep the neuron alive and learning even when the value is negative (Favorskaya & Andreev, 2019). The following is the expression for the Leaky ReLu activation function.

$$LeakyReLu(x) = max\ (0.01x, x)$$

Equation 3.18 Leaky ReLu activation function

$$LeakyReLu(x): Leaky\ Relu\ activation\ function\ on\ input\ x$$

$$x: Input\ to\ the\ leaky\ ReLu\ activation\ function$$

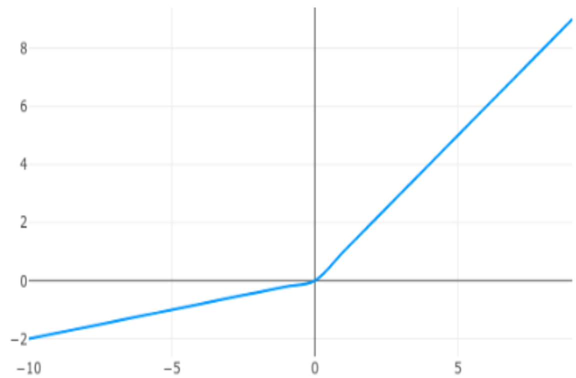

Figure 3.17 LeakyReLu activation function

source: https://images.app.goo.gl/59S4PwBAq5DNKCjN6

Even after solving the dying neuron problem, The Leaky ReLu activation function does not provide a consistent prediction for the negative values.

3.8.1.4.3.5 Parametric Rectified linear unit activation function

The parametric ReLu activation function used alpha (∝) as the minimum range, unlike the constant value of 0.01, as in the case of the leaky ReLu activation function. Thus, the range of the parametric ReLu activation function is between alpha (∝) and x, where x is the input to the activation function (Favorskaya & Andreev, 2019). The following is the expression for the parametric ReLu.

$$ParametricReLu(x) = max\ (\propto x, x)$$

Equation 3.19 Parametric ReLu activation function

$$ParametricReLu(x): Parametric\ ReLu\ activation\ function\ on\ input\ x$$

$$x: Input\ to\ the\ parametric\ ReLu\ activation\ function$$

$\propto: lower\ range\ value\ to\ be\ learned\ during\ back - propagation$

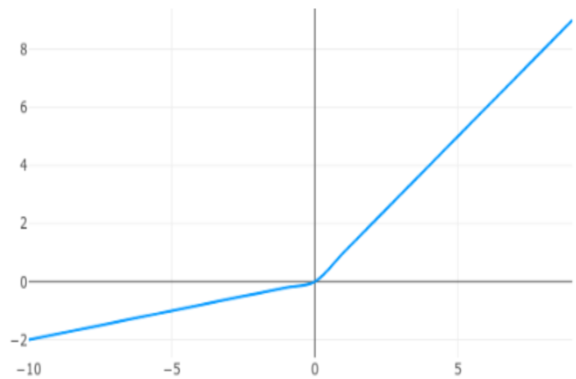

Figure 3.18 Parametric ReLu activation function

source: https://images.app.goo.gl/59S4PwBAq5DNKCjN6

Additionally, the temporal causal discovery framework (Nauta et al. (2019)) was also one of the frameworks used for causality detection. Moreover, the TCDF framework has used a parametric ReLu activation function as part of its framework. The details of TCDF have been provided in the later part of the chapter. Hence, the details of PReLu help us understand the temporal causal discovery framework.

3.8.1.4.3.6 SoftMax activation function

The SoftMax activation function is used to solve the multiclass classification problem. The output of the SoftMax activation function ranged between 0 and 1. The SoftMax activation function is only used for the output layer (B. Chen et al., 2017). The SoftMax function's mathematical equation is given below.

$$Softmax(x_i) = \frac{e^{x_i}}{\sum_{j=0}^{k} \mathrm{e}^{\mathrm{x_j}}}$$

Equation 3.20 Softmax activation function

$Softmax(x_i): Softmax\ function\ for\ given\ input\ x_i$

$x_i: Input\ on\ which\ softmax\ function\ to\ be\ applied$

The SoftMax function is also used to implement the temporal causal discovery framework (Nauta et al. (2019)) to find the potential causes in the time series. Hence, it is worthwhile to understand the SoftMax function for this study.

3.8.1.5 Statistically significant assessment

To assess the statistical significance of models, the authors of NonLinCausality calculated the model's prediction error when using the X and Y time series and when using the X time series. They then used the Wilcoxon signed-rank test to see if the error given by X and Y time series was smaller than the error given by the X time series.

The Wilcoxon signed-rank test is a non-parametric statistical hypothesis test used to compare data. It should be used when the difference between two data sets is not normally distributed (Wilcoxon, 1945).

In nonLinCausality, the null hypothesis is that the median absolute prediction error for a model based on the past value of X is equal to or smaller than for a model based on the past values of X and Y. In contrast, the alternative hypothesis is that the model based on the past value of X and Y has a smaller median absolute prediction error. The differences between the absolute errors of both models are calculated to determine the hypothesis. After that, the ranks are assigned to obtain differences. Each rank is allocated with either a positive or negative sign, and then the sum of these positive and negative ranks is calculated. The Wilcoxon test uses smaller summed-up positive or negative ranks as test statistics. Following that, the test statistics are compared to a critical value. If the test statistics are smaller than the critical value, the study will reject the established null hypothesis and vice versa.

As previously stated, the nonLinCausality package supports multi-layered perceptron, long-term, short-term memory, gate recurrent units, and traditional statistical techniques. The GRU and LSTM implementations of these techniques perform well with time-series data. Furthermore, the data used in this study is time-series in nature. As a result, the GRU and LSTM implementations of nonLinCausality were used in this study.

3.8.2 DoWhy

The DoWhy library is an open-source Python framework that emphasizes causal assumptions (A. Sharma & Kiciman, 2020). It is based on the causal graph to state and test causal assumptions. Most of the prior causal inference libraries consist primarily of statistical estimators. However, the successful application of causal inference requires specifying assumptions about the underlying observed data and validating those assumptions with statistical estimators. As a result, the "DoWhy" library offers an application programming interface (API) that includes four common steps for causal analysis: 1) In the form of a causal graph, the model encapsulates prior knowledge. 2) Using graph-based methods to identify causal effects, 3) estimating the identified estimand using statistical methods, and 4) refuting the obtained estimate using robustness checks on the initial model's assumptions.

Many existing Python libraries for causality detection only focus on estimation. However, the DoWhy library differs from other libraries in robustness checking. Furthermore, the DoWhy library is built on top of two powerful frameworks: graphical models (Pearl, 2010) and potential outcomes (Imbens & Rubin, 2015). The graphical models were used to model assumptions and identify the non-parametric causal effects.

However, potential outcomes were used to make estimates. Furthermore, the DoWhy library supports estimator libraries such as EconML (Oprescu et al., 2019) and CausalML (Chen et al., 2020).

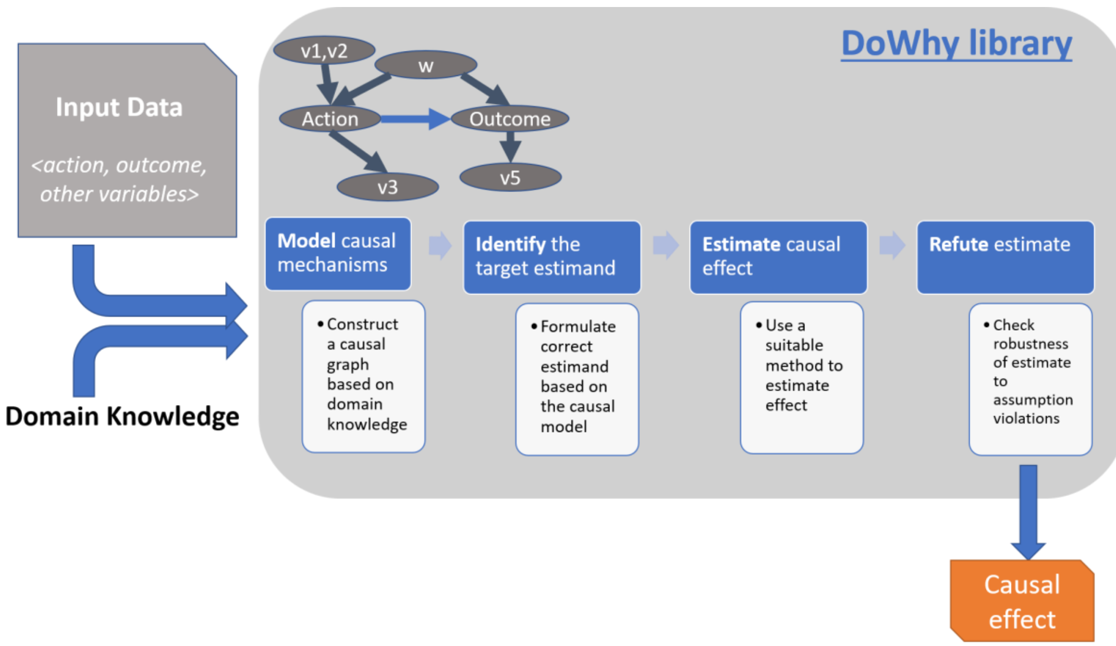

Figure 3.19 DoWhy Library framework

3.8.2.1 The Four step of causal inference

As aforementioned, DoWhy is based upon four common steps for causal analysis. The details for these four steps are as follows:

Model the Causal Question:

DoWhy is used to make causal assumptions using a causal graphical model. The analyst can use a graph to represent partial or complete prior knowledge about variables. In the case of a partial graph, DoWhy automatically considers the remaining variables as potential cofounders. Thus, this research assumed that all the variables would impact the

Indian IT stock prices and let the framework find out the variables causing the change in the Indian IT stock market.

Identify the causal estimand:

The DoWhy explores all the possible ways of identifying the desired causal effect based on the graphical model. The supported identification criteria are:

Back door criteria: The back door criterion focuses on only identifying the direct effect of X, or treatment variable, on Y, or the outcome variable. Through intelligent conditioning on the set of covariates, which block all the indirect paths from X to Y. Pearl (2010) explains that the back door criterion is the most common approach for identifying the causal effects in observational research to condition the possible cofounders. Thus, backdoor criteria were used in this study to estimate the causal effect of the foreign stock market, currency exchange, and economic events on Indian IT stocks.

Front door criteria: Conversely to the back door criteria, the front door criteria find the set of variables M, which mediate the causal influence of X or the treatment variable on the Y or the outcome variable. In other words, all the direct paths from X to Y pass through M. So, if we can identify the effects of X on M and the effects of M on Y, then we can combine their effects to get the final effect of treatment variable X on the outcome variable Y (Rohilla Shalizi, n.d.).

Instrumental variables: If there exists a variable I, which affects X, the treatment variable, and affects the outcome variable Y, but only by influencing the variable X. If somehow, we can find out the effect of I on Y and the effect of I on X, then eventually, we can also get the effect of treatment variable X on the outcome variable Y. Thus, in this case, I is an instrumental variable for the effect of X on Y (Rohilla Shalizi, n.d.).

Mediation (Direct and Indirect effect identification): Mediation is the process through which a treatment X causes an outcome Y. The total effect of a mediator can be broken down into two parts: the direct and indirect effects. The direct effect is the effect of treatment variable X on the outcome variable Y in the absence of the mediator. The indirect effect is the effect of treatment variable X on the outcome variable Y through the mediator (Gunzler et al., 2013).

Estimate the causal effect:

The DoWhy library includes methods that use both the backdoor criterion and instrumental variables. The following are the various suitable methods for estimating the effects:

Several methods for assigning treatment include propensity-based stratification, propensity score matching, and inverse propensity weighting. Propensity-based stratification divides observations into strata with similar propensity scores to balance the observed variables between treated and control units within each stratum.

Propensity score matching is a statistical matching technique that creates an artificial control group by matching each treated unit with a non-treated unit of similar characteristics, allowing researchers to estimate the intervention's effect. In contrast, inverse propensity treatment weighting (IPTW) is a statistical method that creates groups that are otherwise similar when examining the effect of a treatment or exposure. Instead of matching treatment and control groups on a selected set of confounders, the IPTW approach uses the entire cohort and addresses many confounders. Everyone in the cohort is assigned a weight based on the likelihood that they would be exposed to the treatment effect under investigation. Applying these weights when conducting statistical tests or regression models reduces or removes the impact of confounders (Haukoos & Lewis, 2015).

Several methods for estimating the outcome model include linear regression and generalized linear models. Linear regression is a linear approach for modeling the relationship between a scalar response and one or more explanatory variables. When there is only one explanatory variable, it is called simple linear regression. When there are multiple explanatory variables, it is called multiple linear regression. It is important to note that the error distribution in a linear regression model should be normally distributed (Weisberg, 2005).

On the other hand, generalized linear models (GLMs) also model the relationship between a scalar response and explanatory variables, but the relationship is not necessarily linear. As a result, the error distribution in a GLM is not required to be normally distributed and can instead assume an exponential family of distribution (Nimon, 2012).

Several methods for identifying instrumental variables include the binary instrument or Wald estimator, two-staged least squares, and regression discontinuity. Instrumental variables are variables that do not directly affect the outcome but rather influence the selection of treatment conditions. The Wald estimator assesses constraints on statistical parameters based on the weighted distance between the unrestricted estimate and its hypothesized value under the null hypothesis, with the weight being the precision of the estimates (Andrews et al., 2018). Two-staged least squares regression uses instrumental variables that are uncorrelated with the error terms to compute the estimated values of the problematic predictors in the first stage and then uses those computed values to estimate a linear regression model of the dependent variable in the second stage (Young, 2022). Regression discontinuity design (RDD) requires that all potentially relevant variables, besides the treatment and outcome variables, be continuous when the treatment and outcome discontinuities occur.

Methods for backdoor criteria and general mediation, such as two-stage linear regression, the two-stage linear regression has already been defined above.

DoWhy supports integration with EconML and CausalML packages for estimating the conditional average treatment effect (CATE). EconML is a python library used for estimating the heterogenous treatment effects from observational data through machine learning algorithms (Oprescu et al., 2019). Whereas CausalML is also a Python package that provides a suite of uplift modeling techniques and causal inference methods using machine learning algorithms. It provides an interface to estimate the Conditional Average Treatment Effect (CATE) or Individual Treatment Effect (ITE) from the experimental data (Chen et al., 2020). However, this study used the double machine learning (DML) estimation method of EconML's library due to its ability to estimate the heterogeneous treatment effect.

Refute the obtained estimate:

To validate the effect of the causal estimators, the DoWhy library supports a variety of refutation methods. The following refutation methods are supported. The first refutation method is the add a random common cause. It does check whether the estimation method's estimate changes when the independent random variable is added to the dataset. The causal estimate should stay the same if our assumption were initially correct. As the initial assumption of this study, all the variables of the foreign stock market, currency exchange rate, and economic articles would impact the Indian IT stock market. Hence, if the new causal estimates do not change much from the original estimates, then our assumption is correct.

The following estimation method is placebo treatment, which observes the effect of replacing a valid treatment variable with an independent random variable on the estimated causal effect. In the process, placebo treatment randomly assigns any covariate as a treatment and re-runs the analysis. If our assumptions were correct, then this newly found estimate should be zero. To be very sure of the causal effect, this study has also considered the result of placebo treatment. Then, this study has also considered the results of data subset refutation that creates a subset of data like the cross-validation and validates whether the causal estimates vary across subsets. If the original assumptions were correct, there should be a slight variation in the estimations.

The final test included in this study was to add an unobserved common cause that examines the estimated effect of including a common cause (cofounder) in the dataset that is related to the treatment and outcome. If the initial assumptions were correct, then the original and new estimates should be similar. Besides the above refutation methods, simulated outcomes also examine the effect on the estimated causal effect of replacing the outcome with a simulated outcome based on a known data generation process. It is expected that it will match the effect parameter from the data generation process. However, this method was not included in our study.

In summary, the DoWhy library focuses on developing the correct causal model and testing its assumptions, resulting in a more robust and accurate implementation for causality detection. Therefore, the DoWhy library was used in the study to determine the causal impact of the foreign stock market, currency exchange rate, and economic articles on the Indian IT stock.

Python Implementation

```
all_results_list=[]
for col in treatment_col_list:
  dict_dowhy={}
  ## Step 1. Form causal graph based upon prior knowledge.
  model=CausalModel(
        data = sentiment_all_exch_rate_all_indices_normal_Df,
        treatment=col,
        outcome=target_column,
        common_causes=[col1 for col1 in treatment_col_list if col1!=col]
        )
  dict_dowhy['treatment']=col
  ## Step 2. Using Causal graph to identify the causal effect.
  identified_estimand = model.identify_effect(proceed_when_unidentifiable=True)
  ## Step 3. Estimating identified estimand using statistical and machine learning methods.
  estimate = model.estimate_effect(identified_estimand,
                                   method_name="backdoor.linear_regression",
                                   test_significance=True)
  ## Step 4. Refuting the obtained estimate using robustness check.
  refutel = model.refute_estimate(identified_estimand,estimate, "random_common_cause")
  dict_dowhy['RCC_estimated_effect']=refutel.estimated_effect
  dict_dowhy['RCC_new_effect']=refutel.new_effect
  dict_dowhy['RCC_p_value']=refutel.refutation_result['p_value']
  refute2_results=model.refute_estimate(identified_estimand, estimate,
                                        method_name="placebo_treatment_refuter")
  dict_dowhy['PTR_estimated_effect']=refute2_results.estimated_effect
  dict_dowhy['PTR_new_effect']=refute2_results.new_effect
  dict_dowhy['PTR_p_value']=refute2_results.refutation_result['p_value']
  refute3_results=model.refute_estimate(identified_estimand, estimate,
                                        method_name="data_subset_refuter")
  dict_dowhy['DSR_estimated_effect']=refute3_results.estimated_effect
  dict_dowhy['DSR_new_effect']=refute3_results.new_effect
  dict_dowhy['DSR_p_value']=refute3_results.refutation_result['p_value']
  all_results_list.append(dict_dowhy)
```

Figure 3.20 DoWhy implementation

In this study, the DoWhy library was used in Python to implement the steps for identifying causal effects. The code shown in Figure 3.20 demonstrates the implementation of the DoWhy library. It begins by forming a causal graph based on prior knowledge using the CausalModel object. It was assumed that the independent variables representing the foreign IT stock market, currency exchange rate and economic articles impact the Indian IT stock market. The causal graph was then used to identify the causal effect using the identify_effect function of the CausalModel. Next, the estimand was estimated using the backdoor—linear_regression model.

Finally, random_common_cause, placebo_treatment_refuter, and data_subset_refuter were used to verify the robustness of the estimated effect obtained in the previous step.

3.8.3 Temporal Causal Discovery Framework

It consists of an attention-based convolution neural network to discover causal relationships in the time-series data. This framework employs multiple convolution neural networks, with each network receiving all the observed time series as input. However, the attention mechanism has been used to attend to a specific time series for forecasting a future time series.

As previously stated, the Temporal Causal Discovery Framework (TCDF) employs a Convolution Neural Network (CNN) and an attention mechanism, which are discussed in the following sections.

3.8.3.1 Convolution Neural Network (CNN)

A CNN is a feed-forward neural network made up of convolution layers. Convolution is a mathematical linear operation performed between two matrices. The CNN architecture is made up of several layers. The convolution layer, non-linear layer, pooling layer, and fully connected layer are among these layers. Convolutional and fully connected layers learn the parameters, whereas pooling and non-linear layers do not. The CNN architecture demonstrated high performance on the machine learning problem with fewer parameters (Albawi et al., 2017). The elements of convolutional neural networks are as follows:

- Convolution layer
- ReLu layer
- Pooling layer
- Fully connected layer

Convolution Layer

At each time step, the CNN convolution layer slides a kernel over the input and computes the dot product between the input and the kernel. Consider the 5x5 image with a 3x3 kernel. The kernel is slid over the image, and the dot product is computed to obtain the convolved feature matrix.

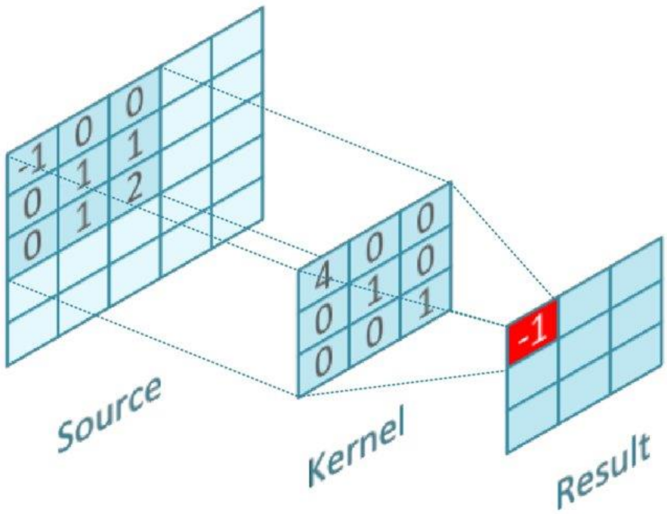

Figure 3.21 Convolution operation

source: (Wicht, 2017)

Relu Layer

The extracted feature maps from convolution operations are then passed to the ReLu layer. ReLu performs the element-by-element operation, setting all negative pixels to 0. It causes the network to become non-linear and generates the rectified feature map. The significant advantage of using the ReLu activation layer is that it saves the network from the vanishing gradient problem.

Pooling Layer

A pooling layer is used to reduce the dimensionality of the feature map. This layer will not learn any parameters during the training of CNN. The popular methods for implementing pooling are average and maximum pooling. The average pooling layer samples the input by dividing it into rectangular pooling regions and computing the average values for each.

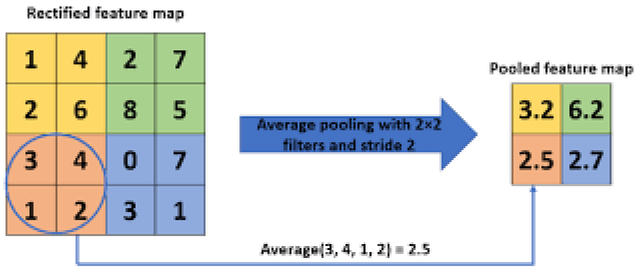

Figure 3.22 Average pooling

source:(Gholamalinezhad & Khosravi, n.d.)

The max pooling layer performs the down-sampling by dividing the input into rectangular pooling regions and taking the maximum values for each region.

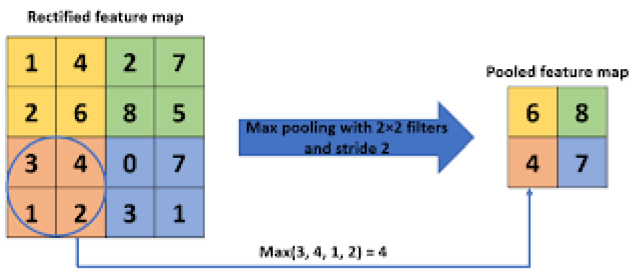

Figure 3.23 Max pooling

source:(Gholamalinezhad & Khosravi, n.d.)

Fully connected layer

The fully connected layer is also called the flattened layer. This layer converts convolution and pooled feature maps to two-dimensional arrays.

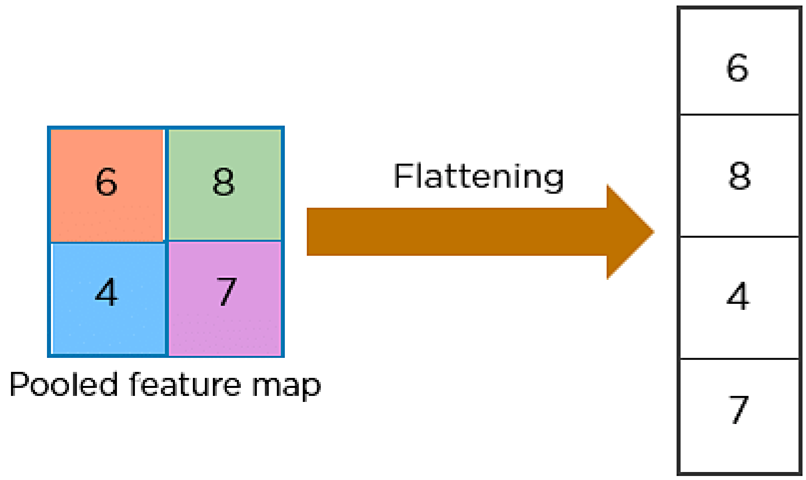

Figure 3.24 Flatten Layer

source:(Biswal, 2022)

3.8.3.2 Attention mechanism in neural network

An attention mechanism allows a neural network to focus on a subset of the input. The concept of attention has always been used in convolution neural networks, where the attention mechanism only selects the relevant parts of the image for object recognition. Attention is now a common feature of sequence networks as well. The attention mechanism allows the model to focus on the vital information in the input data. The TCDF has also used an attention mechanism to determine which input variables are causally associated with the predicted variables.

The TCDF's unique attention-based CNN architecture claims to achieve cutting-edge performance in discovering causal relationships in continuous time series data from the financial and neuroscientific domains. As a result, the outcome of the TCDF framework was also considered in this study to determine the causal relationship between the foreign stock market, currency exchange rate, economics articles, and Indian IT stocks.

The following graphical representation explains how the attention model works. Based on the source input, the model attempts to generate the t-th target. The attention mechanism can be found here (Bahdanau et al., 2015).

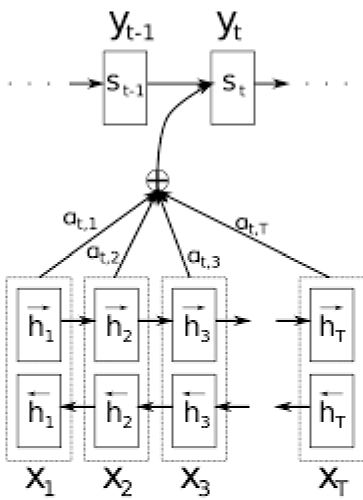

Figure 3.25 Attention mechanism

source:(Bahdanau et al., 2015)

Python Implementation

```
"""Run TCDF"""
%matplotlib inline
%run -i "runTCDF.py" --plot --epochs 1000 --log_interval 250
 --data /content/gdrive/MyDrive/DBA/data/full_data.csv
```

Figure 3.26 TCDF implementation

Figure 3.26 illustrates how the temporal causal discovery framework was used in this study to identify the representation of the foreign IT stock market, currency exchange rate, and economic articles on the Indian IT stock market. As shown in the code snippet, the runTCDF.py script uses four parameters: plot (which will plot the causal graph), epochs (the number of iterations used for convergence), log_interval (the interval at which intermediate results should be logged), and data (which specifies all the variables for which the causal impact is to be identified) to identify the causal impact between the variables.

3.8.4 Granger causality test

Granger causality is widely used for determining the causal relationship between two-time series variables. It is based on the idea that one variable, X, is said to "Granger cause" another variable, Y if the inclusion of the past values of X in a predictive model for Y results in a significantly improved forecasting ability compared to a model that only includes past values of Y.

The methodology for Granger causality involves building two linear regression models: one model in which the dependent variable Y is regressed on its own lagged values and the lagged values of X, and another model in which Y is regressed only on its own lagged values. The models are compared using statistical tests such as the F-test or the likelihood ratio test. If the first model provides a significantly better fit to the data than the second model, it can be concluded that X Granger causes Y.

It is important to note that Granger causality only indicates a direction of causality, and not the strength or the mechanism of the causal relationship. In addition, the results of a Granger causality analysis should be interpreted in the context of other evidence and knowledge about the underlying system.

In conclusion, the use of Granger causality provides a powerful tool for analyzing the causal relationships between time series variables in various fields such as economics, finance, and neuroscience. However, it should be applied with caution and in conjunction with other methods to provide a comprehensive understanding of causal relationships.

Hence, the study employed Granger Causality as a statistical method to establish the causal relationship between various economic variables and the Indian IT stock market. As already discussed, Granger Causality is based on the concept that if one time-series variable can be used to effectively forecast another time-series variable, then it is considered to "Granger-cause" that variable. This technique has become the most widely used method for verifying the usefulness of one variable for forecasting another, despite its limitation in implying a true causal relationship.

The application of Granger Causality in this study helped determine the direction of causality between the foreign stock markets, currency exchange rates, economic articles, and the Indian IT stock market. This information was critical in understanding the impact of these variables on the Indian IT stock market and will contribute to the robustness of our results.

Python Implementation

```
from statsmodels.tsa.stattools import grangercausalitytests
import numpy as np
# maxlag=10
test = 'ssr_chi2test'

def grangers_causation_matrix(data, variables,target, test='ssr_chi2test', verbose=False):

    df = pd.DataFrame(np.zeros((len(target), len(variables))), columns=variables, index=target)
    for c in df.columns:
        for r in df.index:
            test_result = grangercausalitytests(data[[r.split('-')[0], c]], maxlag=int(r.split('-')[1]), verbose=False)
            p_values = [round(test_result[i+1][0][test][1],4) for i in range(int(r.split('-')[1]))]
            if verbose: print(f'Y = {r}, X = {c}, P Values = {p_values}')
            min_p_value = np.min(p_values)
            df.loc[r, c] = min_p_value
    df.columns = [var + '_x' for var in variables]
    df.index = [var for var in target]
    return df

granger_causality_results_10=grangers_causation_matrix(df_train_transformed, variables = df_train_transformed.columns,target=['Nifty_Price-10','Nifty_Price-20','Nifty_Price-30'])
```

Figure 3.27 Granger Causality implementation

3.9 Justification for methodology

Granger causality and the Johansen co-integration test are two traditional methods for detecting causality. These methods, however, need to work better with non-linear data or small sample sizes. On the other hand, methods based on machine learning and deep learning are expected to perform well on small and non-linear data. Thus, for robust and convincing results for causality detection, this study considered five different statistical, machine-learning, and deep-learning algorithms, namely DoWhy (Sharma and Kiciman (2020)), nonLinCausality GRU (Rosoł et al. (2022)), nonLinCausality LSTM (Rosoł et al. (2022)), and Granger Causality test (Granger, 1969).

3.10 Validity

The validity of the study defines the extent to which the concept is measured in the quantitative study (SÜRÜCÜ & MASLAKÇI, 2020). When conducting research, it is critical to consider the validity and reliability of data collection tools and methodology. Furthermore, there are three types of validity: content, construct, and criterion. Content validity examines whether the instrument measures all the variables it was designed to measure. Thus, as mentioned in the data source section, this research gathered foreign stock market data, currency exchange rate data, and economic articles from trusted sources.

The second type of validity examines how well the research has translated the construct into concrete and measurable characteristics (Taherdoost, 2018). Thus, this study used a spacy library to convert economic articles into concise and measurable name entities. The study considered the 18 named entities, which cover the majority of the aspects of text data.

Criterion validity is the final measure of validity. The criterion for validity assesses how well different instruments measure the same variable (Taherdoost, 2018). As a result, the study considered the majority output voting of the five different statistical, machine-learning, and deep-learning algorithms mentioned above. The inclusion of ensemble models' joint statements strengthens the research's validity.

3.11 Reliability

Reliability refers to the consistency of a measure. It ensures that the study will achieve the same results consistently by using the same method under the same circumstances (Heale & Twycross, 2015). This study was based on historical data from secondary sources such as Investing.com, The Hindu, The Financial Times, and Money Control. These data sources are well-known and trustworthy. Furthermore, the research has been conducted using those mentioned statistical, machine learning, and deep-learning methods using rerunnable python codes on a Jupyter notebook, which produced consistent results for the same data with the same settings. As a result, the deterministic nature of this study makes it more reliable.

3.12 Research design limitation

While conducting research, a researcher is expected to think critically not only about the benefits of the research but also about its limitations. The research's limitations are potential flaws beyond the researcher's control (Ross & Bibler Zaidi, 2019). As a result, the study identified some potential flaws. The first uncertainty is the researcher's use of named entities as an operationalized measure of economic articles. However, while the research ensured that 18 named entities were used to represent an article, a few named entities were overlooked. Which could have given the economic articles a better representation.

The other potential area for improvement of that study might be the duration of data collection. Since the study has collected data from 1-Jan-2018 to 31-Dec 2021, it also includes the era when COVID was observed. Hence, this research might produce relevant results for the COVID period and may not be generalizable.

3.13 Summary

The research was an exploratory study that followed the quantitative research principle. The research relied on secondary data sources. The data was collected from January 01, 2018, to December 31, 2021. Following data collection, it was passed to the causality framework, which included the DoWhy (Sharma and Kiciman (2020)), nonLinCausality GRU (Rosoł et al. (2022)), nonLinCausality LSTM (Rosoł et al. (2022)), the temporal causal discovery framework (Nauta et al. (2019)), and Granger Causality (Granger, 1969) algorithms. Following that, the study obtained the output of the previously stated model and used the majority voting mechanism to confirm the most influential variable on the Indian IT stock market.

Furthermore, machine learning and deep-learning-based methods are expected to outperform Granger causality on small and non-linear data. However, Granger causality detection methods are well-suited for linear data Granger causality. Thus employing an ensemble model approach (including the statistical, machine learning, and deep-learning methods) is justified for this research. Moreover, the research's reliability and validity were considered while conducting research. Finally, the potential research design flaws are defined in terms of operationalizing the economic articles and non-generality due to the data collection timeframe chosen.

CHAPTER IV:

RESULTS

4.1 Introduction

This chapter presents the study's findings on the causal relationship between foreign market indices, currency exchange rates, economic articles, and the Indian IT stock market. The chapter begins by discussing the results of basic analysis methods such as bivariate analysis, Pearson correlation, and Spearman correlation. Although these techniques shed light on the interrelationship of the variables, they did not prove causation. Therefore, the chapter also includes a detailed analysis of the results of machine-learning and deep-learning-based causality detection methods. Towards the end, the chapter also includes the traditional Granger causality test results and summarizes all the findings.

4.2 Sample for analysis

This study has obtained foreign stock market data, currency exchange rates, and economic articles from January 1, 2018, to December 31, 2022. The sources for these data are listed in Table 1, Table 2, and Table 3, respectively. Additionally, this study has analyzed the impact of five foreign stock market indices (France, Germany, the UK, the USA, and Japan), six currency exchange rates (EUR/INR, JPY/INR, GBP/INR, SGD/INR, MUR/INR, and USD/INR), 18 named entities, and the positive and negative sentiment of Money Control, The Hindu, and The Financial Times on the Indian IT stock market. Seventy-one variables were analyzed with a sample size of 876 records to examine the impact on the Indian IT stock market.

4.3 Bivariate analysis

The primary goal of this study is to investigate the impact of foreign stock markets, currency exchange rates, and economic articles on Indian IT stocks. However, before analyzing the impact of these variables on the Indian IT stock market, this study conducted a bivariate analysis to visualize the relationship between the two variables. The bivariate analysis examines how the independent or explanatory variable explains the dependent or outcome variable, and it only explores the association between two variables without implying a causal relationship (Bertani et al., 2018). Therefore, this study examined the visual relationships between each depiction of a foreign stock market, each representation of currency exchange rates, and each representation of economic articles and Indian IT stocks.

4.3.1 Foreign stock market and Indian IT stock

As previously mentioned in Chapter 3, this study focused on the IT stock markets of foreign countries, including France, Germany, the UK, the USA, and Japan. These countries were selected because they were most featured in prior studies. Therefore, this study used the FRANCE_INDEX_Price, GERMAN_INDEX_Price, UK_INDEX_Price, USA_INDEX_Price, and JAPAN_INDEX_Price as representations of major foreign IT stock markets, while the Nifty_price represented the Indian IT stock market. The relationship between each foreign stock market and the Indian IT stock market is depicted in the following figure:

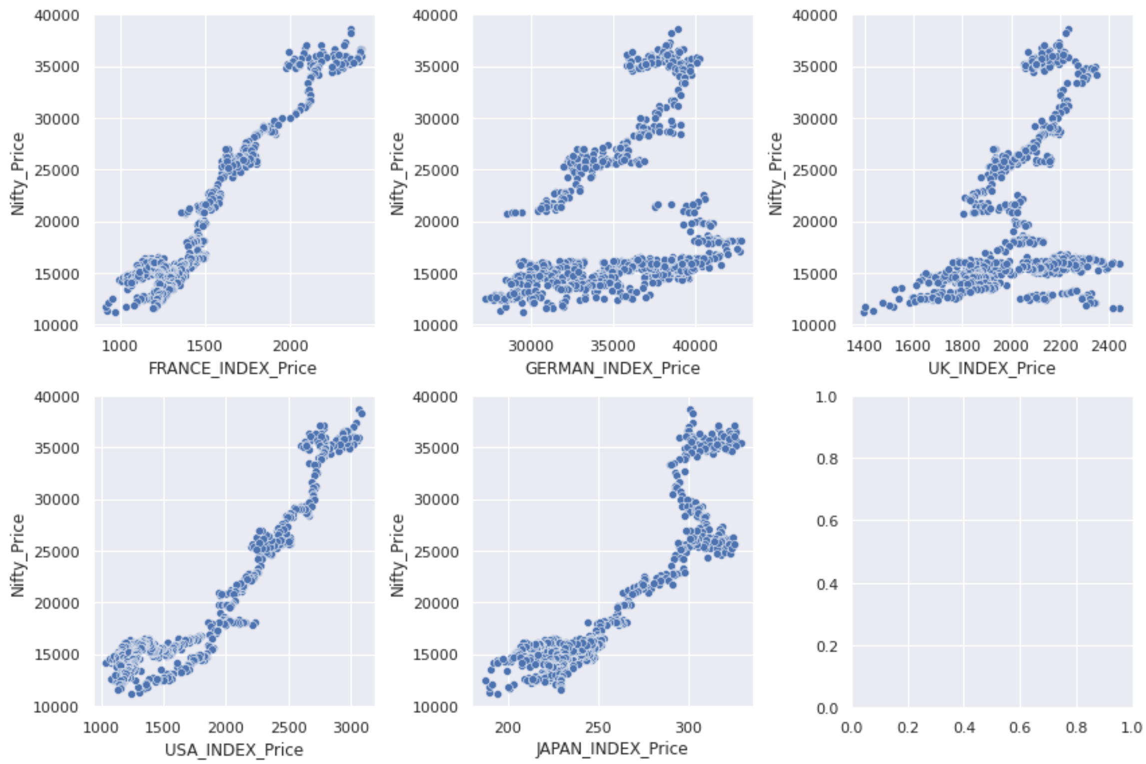

Figure 4.1 Foreign stock market vs. Indian IT stock

As depicted in Figure 4.1, the Nifty IT Price increased along with an increase in the France IT stock market, indicating that the France stock market influenced the Indian IT stock market. However, it should be noted that this observed effect is only a correlation and cannot be definitively concluded as causation. This study applied machine learning, deep learning-based causality detection techniques, and statistical techniques to establish causation. Additionally, the graph in Figure 4.1 reveals a weak positive relationship between the German IT stock market and Nifty IT, suggesting that the German IT stock market had a minimal impact on the Indian IT stock market. The graph in Figure 4.1 also showed a positive relationship between the UK IT stock market and Nifty IT, but this correlation could have been stronger. Therefore, the UK IT stock market seems to have a limited effect on the Indian IT market.

On the other hand, the graph in Figure 4.1 demonstrated a strong positive relationship between the USA IT stock market and Nifty IT, indicating that the USA IT stock market impacted the Indian IT stock market. However, as previously mentioned, this correlation does not prove causation. Therefore, the study applied statistical, machine learning, and deep learning-based causality detection techniques to establish a causal relationship between the US IT stock market and the Indian IT stock market. Similarly, the graph in Figure 4.1 revealed a positive relationship between the Japanese IT stock market and Nifty IT, implying that the Japanese IT stock market also influenced the Indian IT stock market. However, statistical, machine learning, and deep learning-based techniques must also establish the causal relationship between these variables.

4.3.2 Currency exchange rate and Indian IT stock.

One of the main focal points of this study was to investigate the influence of various currency exchange rates on the Indian IT stock market. The currency exchange rates that have been analyzed in this study include the EUR/INR, JPY/INR, GBP/INR, SGD/INR, MUR/INR, and USD/INR rates, which are determined by the financial market and reflect the buying, selling, and exchanging of currencies at current or predetermined prices. The impact of these exchange rates on the Indian IT stock market has been illustrated in the following figures:

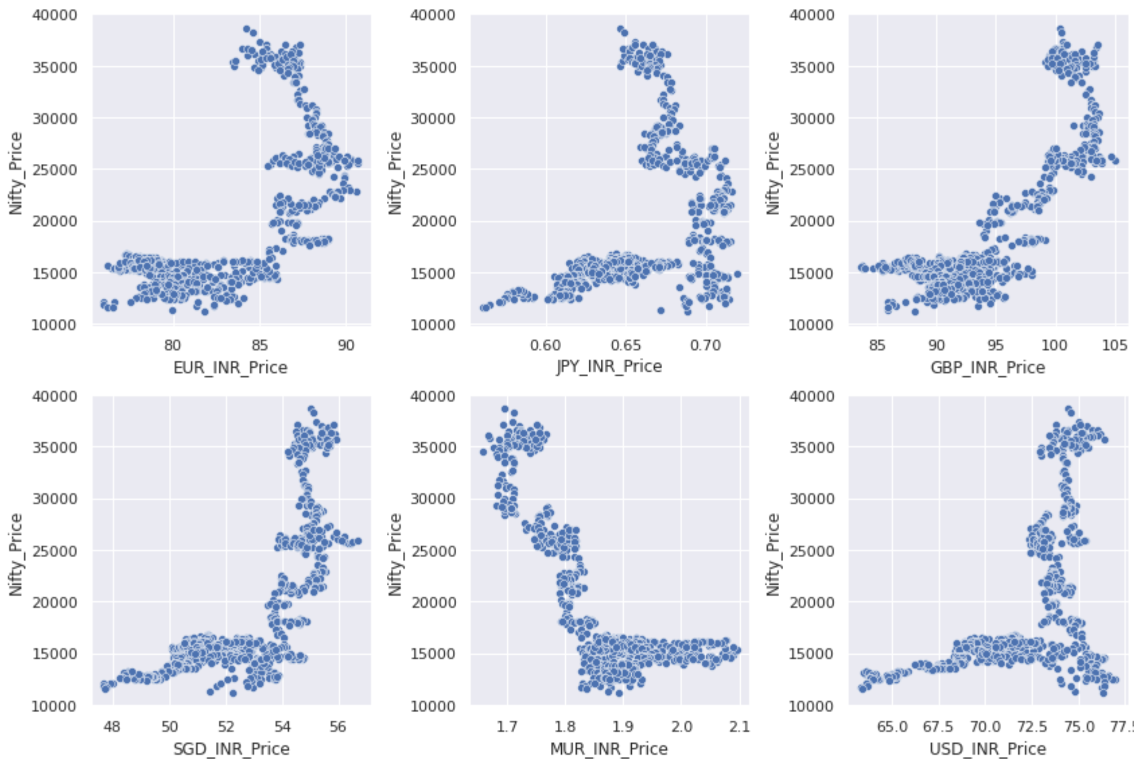

Figure 4.2 Currency exchange rate vs. Indian IT stock

The results of this study, as illustrated in Figure 4.2, suggest that the EUR/INR and the JPY/INR exchange rates have a weak positive relationship with the Nifty IT index until a certain point, after which no trend is observed. These findings suggest that neither the EUR/INR nor the JPY/INR exchange rates significantly impact the Indian IT stock market. In contrast, the GBP/INR exchange rate has a strong positive trend with the Nifty IT index. However, this relationship was further analyzed using statistical, machine learning, and deep-learning-based causality detection techniques to confirm causality confidently. The SGD/INR exchange rate did not exhibit a strong trend with the Nifty IT index, indicating that it did not significantly impact the Indian IT stock market. However, the MUR/INR exchange rate has an inverse relationship with the Nifty IT index, meaning that as the MUR/INR ratio increases, the Nifty IT price decreases. The USD/INR exchange rate initially had a weak positive trend with the Nifty IT index. However, this trend dissipated later, suggesting that it did not significantly impact Indian IT stocks. These findings were also confirmed using statistical, machine learning, and deep-learning-based causality detection techniques, the results of which will be presented later in this chapter.

4.3.3 Economic articles and Indian IT stock

As discussed in Chapter 3, this study collected economic articles from three different sources: Money Control, The Hindu, and The Financial Times. All these articles are written in English, so to make the textual data compatible with machine learning algorithms, the articles were represented using 18 different named entities. These named entities are described in detail in Table 6. To gain a deeper understanding of the content of these articles, the positive and negative sentiments expressed in them were also analyzed. This study section has investigated the relationship between the sentiment expressed in the articles from Money Control, The Hindu, and The Financial Times and the Nifty IT index.

4.3.3.1 Money control and Indian IT stock

MoneyControl.com was launched in late 1999, just as the dot-com bubble began to burst and disrupt the financial market. Despite facing various ups and downs, the platform has grown to receive over 17 million visitors per month across all its platforms (web, mobile, and tablet), making it the largest online financial platform in India. This financial portal has become a significant financial news source, helping investors make informed investment decisions. Therefore, this study has analyzed financial news articles from MoneyControl.com. As previously mentioned in the methodology section, to make the news article data compatible with machine-learning models, the articles were converted into 18 named entities, and their sentiments were analyzed. Figure 23 illustrates the relationship between the 18 named entities extracted from MoneyControl.com and the Indian IT stock market. Figure 24 shows the relationship between the sentiments expressed in the articles (positive or negative) and the Indian IT stock market.

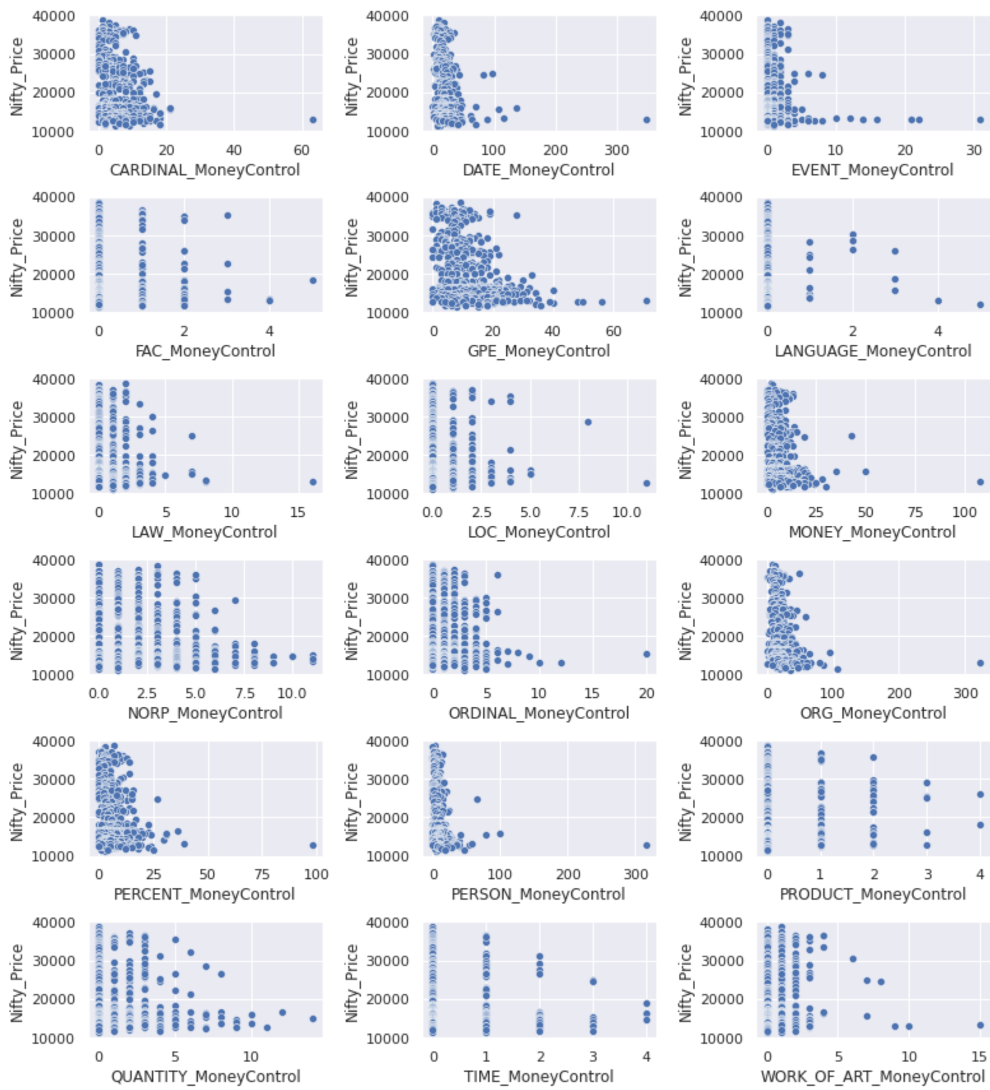

Figure 4.3 Money control named entities vs. Indian IT stock.

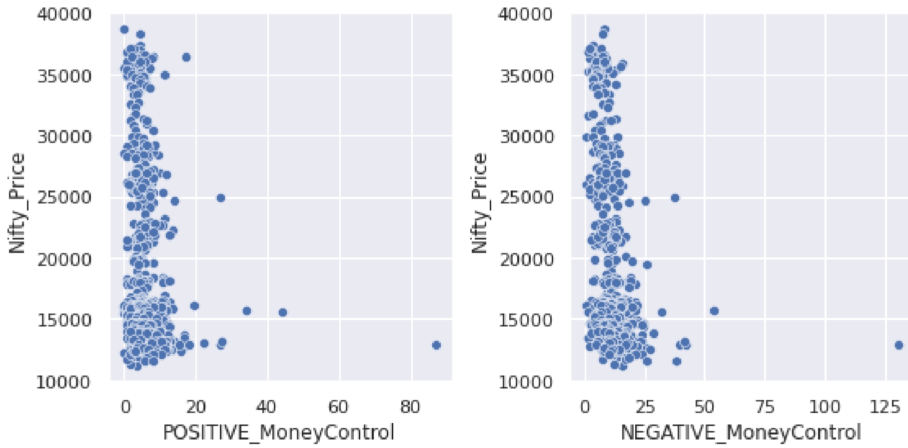

Figure 4.4 Money control sentiments vs. Indian IT stock

As illustrated in Figure 4.3, the CARDINAL named entity and the Nifty IT price did not exhibit any discernible trend. Thus, the CARDINAL-named entity does not significantly impact the Indian IT stock market. In contrast, the Nifty price decreases as the count of DATE and GPE-named entities increases, although this trend is relatively weak. As a result, the DATE-named entity does not significantly influence the Indian IT stock market. Similarly, the EVENT-named entity did not exhibit any noticeable trend with the Nifty price, indicating that it does not significantly impact the Indian IT stock market. However, this study has employed statistical, machine learning, and deep-learning-based causality detection techniques to confirm any potential impact on the Indian IT stock market.

Furthermore, the named entity FAC appears to have a weak negative linear relationship with the Nifty price, indicating that it may not significantly influence the Indian IT stock market. Similarly, the named entity LANGUAGE exhibited a weak inverted relationship with the Nifty IT price, indicating a potential lack of impact on the

Indian IT stock market. The named entities LAW and LOC also showed weak inverted relationships with the Nifty IT price, suggesting a minimal impact on the Indian IT stock market. Furthermore, the named entities MONEY, NORP, ORDINAL, and ORG do not appear to have any discernible impact on the market, as the graph does not indicate any clear trend between these entities and the Nifty IT price. However, the entities PERCENT, PERSON, PRODUCT, QUANTITY, TIME, and WORK_OF_ART exhibited weak inverse relationships with the Nifty IT indices. This study has utilized statistical, machine learning, and deep-learning-based causality detection techniques, to ascertain the impact of these entities on the Indian IT stock market.

As depicted in Figure 4.4, the correlation between sentiments extracted from economic articles and the Nifty IT price, which represents the Indian IT stock market, is not discernible. Both graphs in Figure 4.4 suggested that positive and negative sentiments did not exhibit any consistent trend for the Nifty IT price, leading us to conclude that these sentiments may not impact the Indian IT stock market. However, the study has utilized statistical, machine learning, and deep-learning-based causality detection techniques to confirm the influence of various variables on the Indian IT stock market. The results of these algorithms will be presented later in this chapter.

4.2.3.2 The Hindu and Indian IT stock

The Hindu is a renowned English-language daily newspaper owned by The Hindu Group, a company headquartered in Chennai, Tamil Nadu. Initially, it began as a weekly publication in 1878 and later started daily publication in 1889. It is one of India's major and second-most circulated English-language newspapers. The Hindu publishes various sections, including an economy section. This study has focused on the economic articles published in the economy section of The Hindu. The study has converted the text articles into named entities and sentiments to make these articles compatible with machine-learning models, as discussed in the methodology section. The relationship between Nifty IT and the named entities extracted from the economic articles of The Hindu is depicted in Figure 33. Similarly, Figure 34 illustrates the relationship between the POSITIVE and NEGATIVE sentiments of the news articles and the Nifty IT (which represents Indian IT stocks).

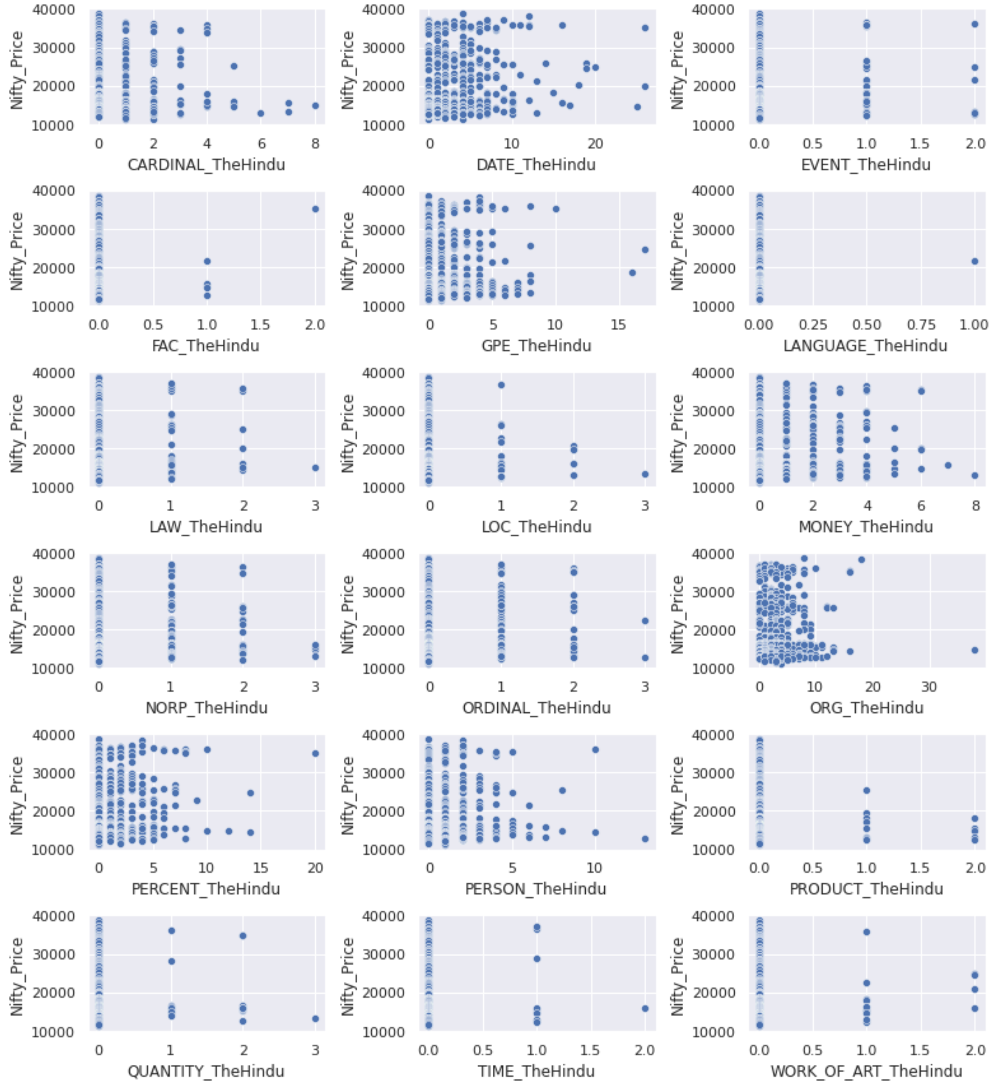

Figure 4.5 The Hindu named entities vs. Indian IT stock.

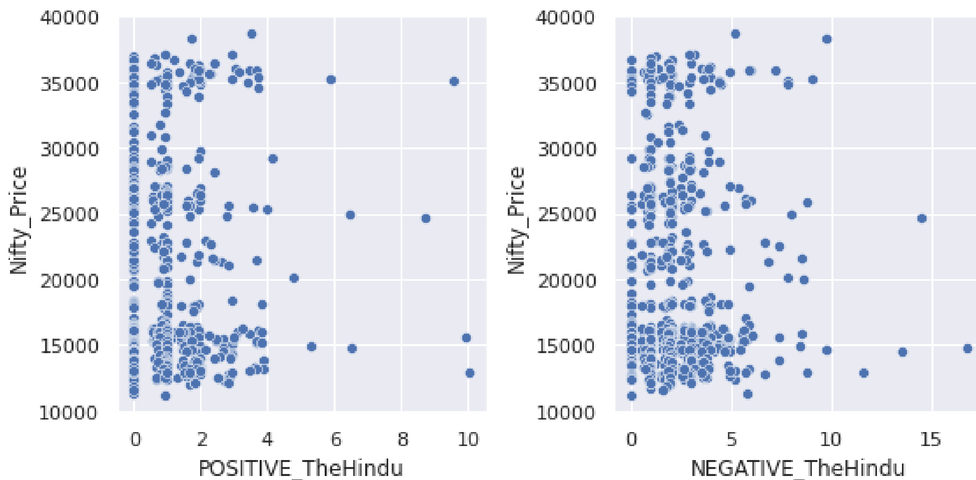

Figure 4.6 The Hindu sentiments vs. Indian IT stock

As depicted in Figure 4.5, the CARDINAL named entity, which represents numerals, appeared to have a trend with the Nifty IT price. The Nifty IT price seems to decrease when the count of CARDINAL-named entities increases, indicating that the CARDINAL-named entity may potentially impact the Indian IT stock market. This study, however, used statistical, machine learning, and deep-learning-based causality detection techniques to confidently confirm the impact of the CARDINAL-named entity on the Indian IT stock market. Furthermore, the DATE-named entity did not exhibit a visible trend with the Nifty IT price, implying that it may not impact the Indian IT stock market. Statistical, machine learning, and deep-learning-based causality detection techniques have further validated this hypothesis.

Similarly, the FAC (which represents buildings, airports, highways, and bridges), GPE (including countries, cities, and states), ORG (representing companies, agencies, and institutions), and PERCENT (including percentages) named entities do not exhibit a visible linear trend with the Nifty IT price, potentially indicating that they may not impact the Indian IT stock market. On the contrary, the Hindu's named entities such as EVENT (including named battles, wars, sports, and events), LANGUAGE (representing any named language), LAW (named documents made into laws), LOC (such as non-GPE locations, mountains, and water bodies), MONEY (containing monetary values including units), and NORP (nationalities or religious or political groups) have a weak negative relationship with the Indian IT stock market, suggesting that these entities may hurt the Indian IT stock market. To confidently confirm the same, this study relied on the results of statistical, machine learning, and deep-learning-based causality detection techniques. Similarly, the following named entities such as ORDINAL (regarding a particular order such as first, second, etc.), PERSON (including people and fiction), product (representing objects, vehicles, foods, etc. (not services)), QUANTITY (representing measurement, as of weight or distance), TIME, and WORK_OF_ART (such as the titles of books, songs, etc.) have a weak negative relationship with the Nifty IT price, potentially indicating that they may hurt the Indian IT stock market, resulting in a decrease in the Nifty IT price with an increase in the values of these entities. To confirm this hypothesis, the results of ensemble model (statistical, machine learning, and deep-learning causality detection) based techniques should be considered.

Figure 4.6 depicts the relationship between the sentiments of economic articles from The Hindu newspaper and the performance of Indian IT stocks. The charts show that POSITIVE sentiments have a very weak positive trend with the Nifty IT price, while NEGATIVE sentiments have a very weak negative trend with the Nifty IT price, which suggests that both POSITIVE and NEGATIVE sentiments may impact the Indian IT stock market. However, the confirmation of this relationship must be determined through the results of statistical, machine learning, and deep-learning-based causality detection techniques, which are presented later in this chapter.

4.3.3.3 Financial Times and Indian IT stock

The Financial Times (FT) is a highly respected British daily newspaper focused on business and current economic affairs. It was first launched as the London Financial Guide on January 10, 1888, and was subsequently renamed to the Financial Times on February 13 of the same year. According to the Global Capital Market Survey, the Financial Times is considered the most important business read by more than 36% of the sample population. In addition, it is widely regarded as the most credible publication for reporting financial and economic concerns among the worldwide professional investment community. As previously described in the methodology section, this study has extracted named entities and news sentiments from the economic articles of the FT. To transform the textual data into a form that is understandable by machine learning algorithms. Therefore, Figure 4.7 visually analyzed the relationship between the sentiments of FT economic articles and the Indian IT stock market, while Figure 36 analyzed the relationship between the named entities extracted from the Financial Times' economic articles and the Indian IT stock market.

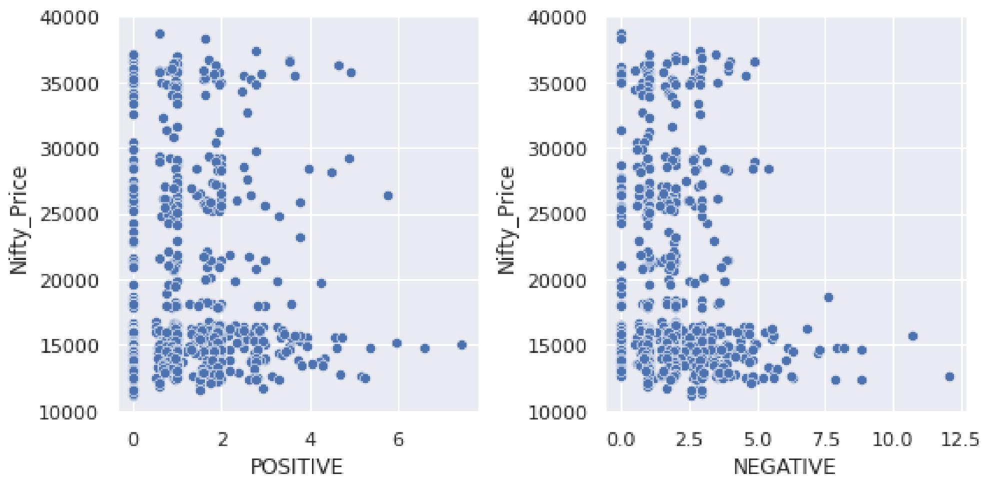

Figure 4.7 Financial news sentiments vs. Indian IT stock

As shown in Figure 4.7, POSITIVE news does not exhibit a clear trend against the Nifty IT price, indicating that it may not significantly impact the Indian IT stock market. However, NEGATIVE articles have a weak inverse relationship with the Nifty IT price, suggesting that an increase in NEGATIVE sentiments may decrease the price of the Indian IT stock market. This study has utilized an ensemble model (including statistical, machine learning, and deep learning) based on causality detection algorithms to confirm any potential impact. The results of these causality detection algorithms will be discussed later in this chapter.

Figure 4.8 Financial named entities vs. Indian IT stock.

As depicted in Figure 4.8, the named entities such as CARDINAL (representing numerals), LANGUAGE (any named language), LAW (named documents made into laws), LOC (including non-GPE locations, mountains, and water bodies), MONEY (monetary values including units), NORP (including nationalities or religious or political groups), ORDINAL (any order such as first, second, etc.), ORG (including companies, agencies, institutions, etc.), PRODUCT (objects, vehicles, foods, etc. (not services)), QUANTITY (measurement, as of weight or distance), TIME, and WORK_OF_ART (such as the title of books, songs, etc.) have a weak negative linear relationship with the Nifty IT stock, which suggests that these named entities may hurt the Indian IT stock market. However, these are only visual trends, and the causality of these named entities on the Indian stock market will be discussed later in this chapter.

On the other hand, named entities such as DATE, EVENT (including named battles, wars, sports, events, etc.), FAC (including buildings, airports, highways, bridges, etc.), GPE (countries, cities, and states), PERCENT (including %), and PERSON (people, including fictional) did not exhibit a clear visual trend with the Nifty IT price. Therefore, these named entities may not impact the Indian IT stock market. However, the impact of these entities will also be discussed in a later part of this chapter.

4.4 Pearson correlation

After visually analyzing the relationship between foreign stock markets, currency exchange rates, and economic articles with the Indian IT stock market, a Pearson correlation test was conducted to assess the strength of these relationships. The Pearson coefficient values range between -1 and 1. When the value of the Pearson coefficient is greater than 0.5, the correlation is considered positive. In contrast, when the value of the Pearson coefficient is less than -0.5, the correlation is considered negative. Otherwise, the correlation is considered none (Isaac & Chikweru, 2018).

Therefore, Table 3.8 represents the strength of the relationship between the dependent variables of foreign stock markets, currency exchange rates, economic news articles, and the Indian IT stock market.

Table 4.1 Data description for the currency exchange rate.

Independent Variable	Dependent Variable	Pearson Coefficient	Correlation Type
FRANCE_INDEX_Price	Nifty_Price	0.959051216	Positive
GERMAN_INDEX_Price	Nifty_Price	0.338037685	None
UK_INDEX_Price	Nifty_Price	0.371909347	None
USA_INDEX_Price	Nifty_Price	0.94057715	Positive
JAPAN_INDEX_Price	Nifty_Price	0.904975796	Positive
EUR_INR_Price	Nifty_Price	0.677600086	Positive
JPY_INR_Price	Nifty_Price	0.3533216	None
GBP_INR_Price	Nifty_Price	0.816059377	Positive
SGD_INR_Price	Nifty_Price	0.722188309	Positive
MUR_INR_Price	Nifty_Price	-0.1866521	None
CARDINAL_MoneyControl	Nifty_Price	-0.241338514	None
DATE_MoneyControl	Nifty_Price	-0.16315781	None
EVENT_MoneyControl	Nifty_Price	-0.093278734	None
FAC_MoneyControl	Nifty_Price	-0.072959892	None
GPE_MoneyControl	Nifty_Price	-0.326525184	None
LANGUAGE_MoneyControl	Nifty_Price	-0.001875018	None
LAW_MoneyControl	Nifty_Price	-0.127759609	None
LOC_MoneyControl	Nifty_Price	-0.088660654	None
MONEY_MoneyControl	Nifty_Price	-0.229516838	None
NORP_MoneyControl	Nifty_Price	-0.224111464	None
ORDINAL_MoneyControl	Nifty_Price	-0.100188021	None
ORG_MoneyControl	Nifty_Price	-0.258656123	None
PERCENT_MoneyControl	Nifty_Price	-0.185120748	None
PERSON_MoneyControl	Nifty_Price	-0.207139006	None
PRODUCT_MoneyControl	Nifty_Price	0.089928362	None
QUANTITY_MoneyControl	Nifty_Price	-0.197884122	None
TIME_MoneyControl	Nifty_Price	-0.139285669	None
WORK_OF_ART_MoneyControl	Nifty_Price	0.056210563	None
CARDINAL_TheHindu	Nifty_Price	-0.023129342	None
DATE_TheHindu	Nifty_Price	0.150796157	None

EVENT_TheHindu	Nifty_Price	0.031769739	None
FAC_TheHindu	Nifty_Price	0.023001294	None
GPE_TheHindu	Nifty_Price	0.023126432	None
LANGUAGE_TheHindu	Nifty_Price	0.010353878	None
LAW_TheHindu	Nifty_Price	0.052245881	None
LOC_TheHindu	Nifty_Price	-0.049817505	None
MONEY_TheHindu	Nifty_Price	0.074948259	None
NORP_TheHindu	Nifty_Price	-0.040934538	None
ORDINAL_TheHindu	Nifty_Price	0.088573808	None
ORG_TheHindu	Nifty_Price	0.039942688	None
PERCENT_TheHindu	Nifty_Price	0.162734588	None
PERSON_TheHindu	Nifty_Price	-0.012102322	None
PRODUCT_TheHindu	Nifty_Price	-0.059143098	None
QUANTITY_TheHindu	Nifty_Price	-0.018282565	None
TIME_TheHindu	Nifty_Price	-0.007209617	None
WORK_OF_ART_TheHindu	Nifty_Price	-0.01223774	None
CARDINAL	Nifty_Price	-0.052851192	None
DATE	Nifty_Price	-0.146825004	None
EVENT	Nifty_Price	-0.01276896	None
FAC	Nifty_Price	-0.013781695	None
GPE	Nifty_Price	-0.082300297	None
LANGUAGE	Nifty_Price	0.022882903	None
LAW	Nifty_Price	-0.04764884	None
LOC	Nifty_Price	-0.061480957	None
MONEY	Nifty_Price	-0.059573914	None
NORP	Nifty_Price	-2.34E-05	None
ORDINAL	Nifty_Price	-0.016859372	None
ORG	Nifty_Price	-0.101162239	None
PERCENT	Nifty_Price	-0.086118728	None
PERSON	Nifty_Price	0.002883129	None
PRODUCT	Nifty_Price	-0.037304052	None
QUANTITY	Nifty_Price	-0.01320673	None
TIME	Nifty_Price	-0.050904584	None
WORK_OF_ART	Nifty_Price	0.052694352	None
POSITIVE_MoneyControl	Nifty_Price	-0.149591271	None
NEGATIVE_MoneyControl	Nifty_Price	-0.30855929	None
POSITIVE_TheHindu	Nifty_Price	0.086438383	None
NEGATIVE_TheHindu	Nifty_Price	0.041436437	None
POSITIVE	Nifty_Price	-0.046783214	None
NEGATIVE	Nifty_Price	-0.125822824	None

USD_INR_Price	Nifty_Price	0.529832478	Positive

As shown in Table 4.1, the FRANCE_INDEX_Price, USA_INDEX_Price, and JAPAN_INDEX_Price positively correlates with the Nifty IT stock. Furthermore, the strength of these correlations is greater than 90%, indicating a high likelihood that FRANCE, USA, and Japanese stock markets will impact the Indian IT stock market. However, to confirm this impact, this study has utilized an ensemble model (including statistical, machine learning, and deep learning) based on causality detection techniques. The details of these causality detection algorithms will be discussed later in this chapter.

Furthermore, the currency exchange rates such as EUR_INR_Price, GBP_INR_Price, and SGD_INR_Price positively correlates with the Nifty IT price. The strength of these correlations is greater than 65% but less than 90%, indicating that these currency exchange rates may have a moderate impact on the Indian IT stock market. The details of whether these currency exchange rates truly cause the Indian IT stock market will be discussed later in this chapter. However, the USD-INR_Price has a positive relationship with Nifty IT, but the strength of this relationship is only around 52%. Therefore, the USD_INR may negatively impact the Indian IT stock market.

On the other hand, the MUR__INR_Price has a negative correlation with the Nifty IT, and the strength of this relationship is quite strong (nearly 80%), indicating that an increase in the MUR_INR price may result in a decrease in the Nifty IT price. Thus, the MUR_INR may hurt the Indian IT stock market. Other than these, the remaining independent variables did not correlate with the Nifty IT. However, to decisively confirm this, this study has utilized statistical, machine learning, and deep-learning-based causality detection techniques, which will be discussed in a later part of this chapter.

4.5 Spearman correlation

The Spearman correlation measure is used to assess the strength of monotonic relationships. A monotonic relationship is one in which: 1) as the value of one variable increases, the value of the other variable also increases, or 2) as the value of one variable increases, the value of the other variable decreases. However, the increase or decrease in the other variable does not need to be at a constant rate. In contrast, in a linear relationship, the rate of increase or decrease in values is constant (Ali & Al-Hameed, 2022). The result of the Spearman correlation also varies between -1 and 1. Suppose the value of the Spearman correlation is greater than 0.5. In that case, the correlation is considered positive, while if the value of the Spearman correlation is less than -0.5, the correlation is considered negative. Otherwise, the correlation is none. For more information on the Spearman correlation, please refer to Chapter 3.

Table 4.2 Data description for the currency exchange rate.

Independent Variable	Dependent Variable	Spearman Coefficient	Correlation Type
FRANCE_INDEX_Price	Nifty_Price	0.774724818	Positive
GERMAN_INDEX_Price	Nifty_Price	0.48181311	None
UK_INDEX_Price	Nifty_Price	0.455122845	None
USA_INDEX_Price	Nifty_Price	0.858622268	Positive
JAPAN_INDEX_Price	Nifty_Price	0.809044124	Positive
EUR_INR_Price	Nifty_Price	0.595852152	Positive
JPY_INR_Price	Nifty_Price	0.479276378	None
GBP_INR_Price	Nifty_Price	0.719538968	Positive
SGD_INR_Price	Nifty_Price	0.75595535	Positive
MUR_INR_Price	Nifty_Price	-0.140944604	None

CARDINAL_MoneyControl	Nifty_Price	-0.260166628	None
DATE_MoneyControl	Nifty_Price	-0.241972123	None
EVENT_MoneyControl	Nifty_Price	-0.069870477	None
FAC_MoneyControl	Nifty_Price	-0.126037835	None
GPE_MoneyControl	Nifty_Price	-0.329537936	None
LANGUAGE_MoneyControl	Nifty_Price	0.01335185	None
LAW_MoneyControl	Nifty_Price	-0.166670531	None
LOC_MoneyControl	Nifty_Price	-0.103336104	None
MONEY_MoneyControl	Nifty_Price	-0.313594934	None
NORP_MoneyControl	Nifty_Price	-0.212064299	None
ORDINAL_MoneyControl	Nifty_Price	-0.077669623	None
ORG_MoneyControl	Nifty_Price	-0.382193599	None
PERCENT_MoneyControl	Nifty_Price	-0.219586587	None
PERSON_MoneyControl	Nifty_Price	-0.387414283	None
PRODUCT_MoneyControl	Nifty_Price	0.102055278	None
QUANTITY_MoneyControl	Nifty_Price	-0.214628961	None
TIME_MoneyControl	Nifty_Price	-0.152048209	None
WORK_OF_ART_MoneyControl	Nifty_Price	0.054196381	None
CARDINAL_TheHindu	Nifty_Price	-0.040747581	None
DATE_TheHindu	Nifty_Price	0.085057514	None
EVENT_TheHindu	Nifty_Price	0.027794034	None
FAC_TheHindu	Nifty_Price	-0.016637116	None
GPE_TheHindu	Nifty_Price	-0.008892267	None
LANGUAGE_TheHindu	Nifty_Price	0.024397518	None
LAW_TheHindu	Nifty_Price	0.047735381	None
LOC_TheHindu	Nifty_Price	-0.043630944	None
MONEY_TheHindu	Nifty_Price	0.035781138	None
NORP_TheHindu	Nifty_Price	-0.030415166	None
ORDINAL_TheHindu	Nifty_Price	0.097647109	None
ORG_TheHindu	Nifty_Price	0.021915853	None
PERCENT_TheHindu	Nifty_Price	0.108225223	None
PERSON_TheHindu	Nifty_Price	-0.044811292	None
PRODUCT_TheHindu	Nifty_Price	-0.053437265	None
QUANTITY_TheHindu	Nifty_Price	-0.020219156	None
TIME_TheHindu	Nifty_Price	-0.034472358	None
WORK_OF_ART_TheHindu	Nifty_Price	-0.006069277	None
CARDINAL	Nifty_Price	-0.068058325	None
DATE	Nifty_Price	-0.160129302	None
EVENT	Nifty_Price	-0.024531813	None
FAC	Nifty_Price	0.003902795	None

GPE	Nifty_Price	-0.104649307	None
LANGUAGE	Nifty_Price	0.020068057	None
LAW	Nifty_Price	-0.045381771	None
LOC	Nifty_Price	-0.064200864	None
MONEY	Nifty_Price	-0.081293429	None
NORP	Nifty_Price	-0.025392249	None
ORDINAL	Nifty_Price	-0.076798083	None
ORG	Nifty_Price	-0.070072231	None
PERCENT	Nifty_Price	-0.096707651	None
PERSON	Nifty_Price	-0.043634371	None
PRODUCT	Nifty_Price	-0.042629503	None
QUANTITY	Nifty_Price	-0.044579484	None
TIME	Nifty_Price	-0.074067178	None
WORK_OF_ART	Nifty_Price	0.020561907	None
POSITIVE_MoneyControl	Nifty_Price	-0.191650824	None
NEGATIVE_MoneyControl	Nifty_Price	-0.383381408	None
POSITIVE_TheHindu	Nifty_Price	0.023257549	None
NEGATIVE_TheHindu	Nifty_Price	0.002666837	None
POSITIVE	Nifty_Price	-0.0583115	None
NEGATIVE	Nifty_Price	-0.151781891	None
USD_INR_Price	Nifty_Price	0.524350189	Positive

As illustrated in Table 4.2, the JAPAN_INDEX_Price and USA_INDEX_Price has a strong positive Spearman correlation with Nifty IT. The strength of this correlation is greater than 80%, indicating that changes in the stock markets of JAPAN and the USA may have a strong positive impact on the Indian IT stock market. However, it should be noted that this is only a correlation. This study used statistical, machine learning, and deep-learning-based causality detection techniques to convert this correlation into causation. The results of these causality detection techniques will be discussed later in this chapter.

Similarly, the FRANCE_INDEX_Price, GBP_INR_Price, and SGD_INR_Price has a positive correlation (with a strength of more than 70%) with the Nifty IT, indicating that an increase in the stock market for FRANCE, the GBP, and INR ratio, and the SGD and INR ratio may increase the price of the Indian stock market. However, the variables such as EUR_INR_Price and USD_INR_Price also exhibited a small positive correlation (just above 50%), suggesting an increase in the EUR and INR and USD and INR currency exchange rates may lead to an increase in the Indian IT stock market. This study has utilized ensemble model (statistical-based, machine learning, and deep-learning-based) causality detection algorithms to confirm this, which will be discussed later in this chapter. Other than the abovementioned variable, no other variables strongly correlated with the Nifty IT. Specifically, all the named entities and the sentiment variables extracted from the economic articles' sources such as Money Control, The Hindu, and The Financial Times did not show a strong correlation with the Indian IT stock market. However, this does not necessarily mean that these variables do not impact the Indian IT stock market. As mentioned earlier, the impact of these variables can only be confidently confirmed by examining the results of statistical, machine learning, and deep-learning-based causality detection methods.

4.6 NonLinCausality

After conducting bivariate and correlation analysis, this study has identified potential variables that may impact the Indian IT stock market. However, to positively confirm whether these features impacted the Indian stock market, this study applied statistical, machine learning, and deep-learning-based causality detection algorithms. One of the algorithms used in this study was nonLinCausality, which utilized GRU and LSTM implementations for causality detection. The results of the GRU and LSTM implementations of nonLinCausality will be discussed in the following sections.

4.6.1 NonLinCausality GRU

The GRU, or Gated Recurrent Unit, is a recurrent neural network that aims to solve the vanishing gradient problem. It is a leaner (less complex) variation of the LSTM (long-term, short-term memory) network and has been shown to produce equally good results. In this study, the GRU implementation of nonLinCausality was used with a lag of 10, 20, and 30 timesteps. The algorithm considered the data from the last 10 days, 20 days, and 30 days to detect causality between foreign stock markets, currency exchange rates, economic articles from different sources, and the Indian IT stock market. The results of considering the last 10 days, 20 days, and 30 days of data for causality detection on the Indian IT stock market will be discussed in detail in the following sections.

4.6.1.1 NonLinCausality GRU with ten days lag

The following tables show the representations of foreign stock markets, currency exchange rates, and economic articles (from Money Control, The Hindu, and The Financial Times) that impact the Indian IT stock market based on the last ten days of data.

Table 4.3 Most impacted variables considering ten days lag.

Independent Variable	Lag	P_value
FRANCE_INDEX_Price	10	0.031475128
LANGUAGE_MoneyControl	10	5.6E-09
LAW_MoneyControl	10	0.019470166
PERSON_MoneyControl	10	0.049036487
TIME_MoneyControl	10	0.0000305
FAC_TheHindu	10	0.0000414
LANGUAGE_TheHindu	10	5.6E-09
LOC_TheHindu	10	0.000000021
PRODUCT_TheHindu	10	5.6E-09
WORK_OF_ART_TheHindu	10	0.0000113
EVENT	10	2.10283E-07
FAC	10	1.05472E-08
LANGUAGE	10	1.03895E-07

LOC	10	0.018905773
MONEY	10	9.0711E-06
NORP	10	0.039984316
ORDINAL	10	0.001374078
PERCENT	10	0.016288158
PRODUCT	10	9.73622E-08
QUANTITY	10	6.42513E-07

As shown in Table 4.3, this study used the p-value to identify the variables that impacted the Indian IT stock market. If the p-value for a particular variable was less than 0.05, it meant that the variable impacted the Indian IT stock market. Based on the last 10 days of data, 20 variables had the most impact on the Indian IT stock market. Specifically, out of these 20 variables, the FRANCE_INDEX_price, or the French stock market, was the only foreign stock that impacted the Indian IT stock market. However, there were four named entities extracted from the articles of Money Control, such as LANGUAGE_MoneyControl, LAW_MoneyControl, PERSON_MoneyControl, and TIME_MoneyControl, that had an impact on the Indian IT stock market based on the last ten days of data.

Similarly, the five named entities fetched from the articles of The Hindu that had an impact on the Indian IT stock market within the last ten days of data were FAC_TheHindu, LANGUAGE_TheHindu, LOC_TheHindu, PRODUCT_TheHindu, and WORK_OF_ART_TheHindu. Ten other entities extracted from The Financial Times articles also impacted the Indian IT stock market. These ten named entities were EVENT, FAC, LANGUAGE, LOC, MONEY, NORP, ORDINAL, PERCENT, PRODUCT, and QUANTITY. From Table 4.3, it was found that the currency exchange rate and economical article sentiment variables did not have any impact on the Indian IT stock market based on the last ten days of data. Among foreign stock markets, only the French market had an impact on the Indian IT stock market.

4.6.1.2 NonLinCausality GRU with 20 days lag

The tables below depict the representation of the global stock market, currency exchange rate, and economic articles (such as Money Control, Hindu, and Financial Times) that have had the greatest impact on Indian IT stocks during the last 20 days, based on data from the last 20 days.

Table 4.4 Most impacted variables considering 20 days lag.

Independent Variable	Lag	P_value
FRANCE_INDEX_Price	20	4.68E-07
USA_INDEX_Price	20	2.72E-06
LANGUAGE_MoneyControl	20	2.69E-07
LOC_MoneyControl	20	0.028525244
PERSON_MoneyControl	20	0.0392055
TIME_MoneyControl	20	0.000194997
EVENT_TheHindu	20	0.01246736
LANGUAGE_TheHindu	20	3.90E-07
LOC_TheHindu	20	2.72E-06
PRODUCT_TheHindu	20	2.69E-07
EVENT	20	3.26E-05
FAC	20	1.76E-05
LAW	20	2.69E-07
LOC	20	0.002730968
MONEY	20	0.000255148
ORG	20	0.032205787
PERCENT	20	0.003582159
QUANTITY	20	0.000194997

As shown in Table 4.4, there are 18 variables with a p-value less than 0.05 based on 20 days of lag data, indicating that these variables impact the Indian IT stock market.

Among these 18 variables, two are related to foreign stock markets (FRANCE_INDEX_Price and USA_INDEX_Price) and impact the Indian IT stock market. Additionally, four named entity variables extracted from Money Control articles (LANGUAGE_MoneyControl, LOC_MoneyControl, PERSON_MoneyControl, and TIME_MoneyControl) have had an impact on the Indian IT stock market. Four named entity variables from articles in The Hindu (EVENT_TheHindu, LANGUAGE_TheHindu, LOC_TheHindu, and PRODUCT_TheHindu) have also had an impact on the Indian IT stock market. Finally, eight named entity variables from the Financial Times articles (EVENT, FAC, LAW, LOC, MONEY, ORG, PERCENT, and QUANTITY) have impacted the Indian IT stock market. However, this study did not find any currency exchange rate or economic article sentiment variables that impacted the Indian IT stock market, even with 20 days of lag data. It is worth noting that the French and US stock markets have impacted the Indian IT stock market.

4.6.1.3 NonLinCausality GRU with 30 days lag

The following tables show the variables from the foreign stock market, currency exchange rate, and different economic article sources such as Money Control, The Hindu, and The Financial Times that have the most impact on the Indian IT stocks considering the last 30 days of data.

Table 4.5 Most impacted variables considering 30 days lag.

Independent Variable	Lag	P_value
FRANCE_INDEX_Price	30	1.19E-07
MUR_INR_Price	30	0.009778619
LANGUAGE_MoneyControl	30	1.19E-07
MONEY_MoneyControl	30	0.006147265

PERSON_MoneyControl	30	0.001207829
TIME_MoneyControl	30	0.009778619
EVENT_TheHindu	30	0.001067996
LANGUAGE_TheHindu	30	1.19E-07
LAW_TheHindu	30	0.00216949
PRODUCT_TheHindu	30	1.19E-07
CARDINAL	30	0.019190073
DATE	30	0.002425671
EVENT	30	3.02E-05
FAC	30	5.13E-06
LANGUAGE	30	7.63E-05
LAW	30	1.19E-07
LOC	30	0.011662364
MONEY	30	2.38E-07
NORP	30	0.00743258
ORDINAL	30	0.000149012
ORG	30	0.000729203
PERCENT	30	0.006764293
WORK_OF_ART	30	0.00743258

According to the data presented in Table 4.5, there are 23 variables with a p-value less than 0.05 based on 30 days of lag data, indicating that these variables have significantly impacted Indian IT stocks. Among these 23 variables, one is related to the foreign stock market (FRANCE_INDEX_Price) and has impacted the Indian IT stock market. Interestingly, one currency exchange rate variable (MUR_INR_Price) has also impacted the Indian IT stock market with a lag of 30 days. In addition, four named entity variables extracted from Money Control articles (LANGUAGE_MoneyControl, MONEY_MoneyControl, PERSON_MoneyControl, and TIME_MoneyControl) have had an impact on the Indian IT stock market. Four named entity variables from articles in The Hindu (EVENT_TheHindu, LANGUAGE_TheHindu, LAW_TheHindu, and PRODUCT_TheHindu) have also had an impact on the Indian IT stock market with a lag of 30 days. 13 named entity variables from Financial Times articles (CARDINAL, DATE,

EVENT, FAC, LANGUAGE, LAW, LOC, MONEY, NORP, ORDINAL, ORG, PERCENT, and WORK_OF_ART) have also had an impact on the Indian IT stock market. However, Table 4.6 does not show any foreign stock market variables or economic article sentiments that have impacted the Indian IT stock market.

According to the data presented in Table 4.6, the GRU's implementation of nonLinCausality with various time lags has identified 31 unique cumulative representations of foreign stock markets, currency exchange rates, and economic articles from various sources that have had an impact on the Indian IT stock market. These representations can be broken down as follows: Two foreign stock market variables (FRANCE_INDEX_Price and USA_INDEX_Price) have impacted the Indian IT stock market. At the same time, the MUR_INR_Price (MUR to INR) currency exchange rate has also had an impact. In addition, six named entities (LANGUAGE_MoneyControl, LAW_MoneyControl, PERSON_MoneyControl,

TIME_MoneyControl, LOC_MoneyControl, and MONEY_MoneyControl) extracted from Money Control articles have had an impact on the Indian IT stock market. Seven named entities (FAC_TheHindu, LANGUAGE_TheHindu, LOC_TheHindu, PRODUCT_TheHindu, WORK_OF_ART_TheHindu, EVENT_TheHindu, and LAW_TheHindu) from The Hindu have had an impact on the Indian IT stock market. Fifteen named entities (EVENT, FAC, LANGUAGE, LOC, MONEY, NORP, ORDINAL, PERCENT, PRODUCT, QUANTITY, LAW, ORG, CARDINAL, DATE, and WORK_OF_ART) from Financial Times articles have also had an impact on the Indian IT stock market.

As shown in Figure 4.9, the variables on the Y-axis are plotted against the number of times they occurred at the different time lags (10 days, 20 days, and 30 days) on the X-axis. Variables with a count greater than one can be considered more reliable than others. Therefore, the following 19 variables can be considered more impactful:

Table 4.6 Most impacted variables summary for GRU nonLinCausality

Independent Variables	GRU implementation of NonLinCausality		
	10 days	20 days	30 days
FRANCE_INDEX_Price	Yes	Yes	Yes
LANGUAGE_MoneyControl	Yes	Yes	Yes
LAW_MoneyControl	Yes	No	No
PERSON_MoneyControl	Yes	Yes	Yes
TIME_MoneyControl	Yes	Yes	Yes
FAC_TheHindu	Yes	No	No
LANGUAGE_TheHindu	Yes	Yes	Yes
LOC_TheHindu	Yes	Yes	No
PRODUCT_TheHindu	Yes	Yes	Yes
WORK_OF_ART_TheHindu	Yes	No	No
EVENT	Yes	Yes	Yes
FAC	Yes	Yes	Yes
LANGUAGE	Yes	No	Yes
LOC	Yes	Yes	Yes
MONEY	Yes	Yes	Yes
NORP	Yes	No	Yes
ORDINAL	Yes	No	Yes
PERCENT	Yes	Yes	Yes
PRODUCT	Yes	No	No
QUANTITY	Yes	Yes	No
USA_INDEX_Price	No	Yes	No
LOC_MoneyControl	No	Yes	No
EVENT_TheHindu	No	Yes	Yes
LAW	No	Yes	Yes
ORG	No	Yes	Yes
MUR_INR_Price	No	No	Yes
MONEY_MoneyControl	No	No	Yes
LAW_TheHindu	No	No	Yes

CARDINAL	No	No	Yes
DATE	No	No	Yes
WORK_OF_ART	No	No	Yes

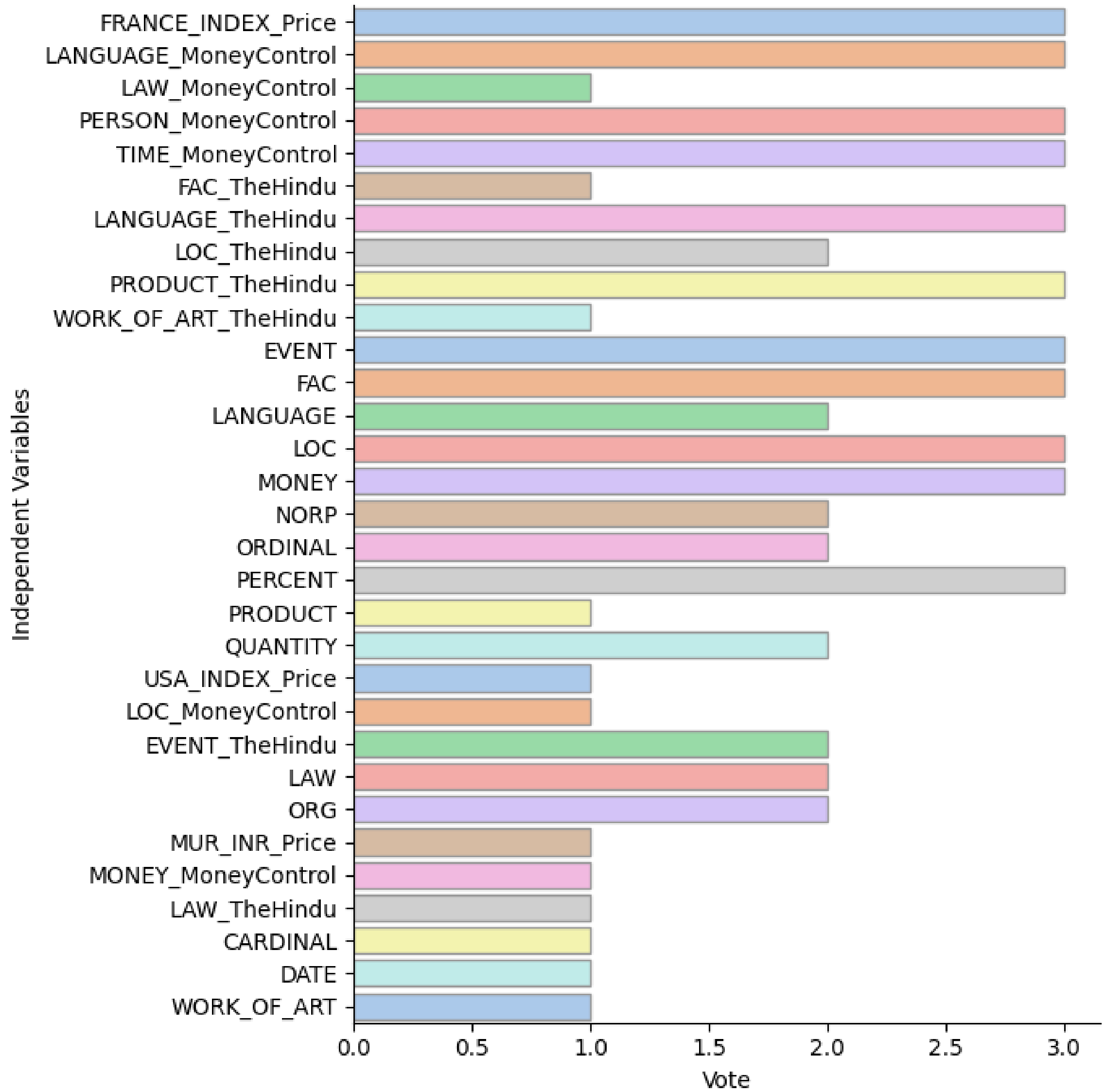

Figure 4.9 Independent Variables vs. Vote count for nonLinCausality GRU

4.6.2 NonLinCausality LSTM

LSTM stands for Long Short-Term Memory networks. Like GRU networks, LSTM networks can also solve the vanishing gradients problem. LSTM networks were the first to solve this problem, and the GRU architecture was later developed based on the LSTM architecture but with fewer complexities. While LSTM networks are more complex than GRU networks, they are still more accurate than GRUs on longer sequence datasets. Therefore, this study has used the LSTM implementation of nonLinCausality with lags of 10, 20, and 30 timesteps. The lags of 10, 20, and 30 timesteps mean that the algorithm has considered data from the last ten days, 20 days, and 30 days, respectively, to detect the impact of foreign stock markets, currency exchange rates, and economic articles from various sources (including Money Control, The Hindu, and The Financial Times) on the Indian IT stock market. The details of the nonLinCausality LSTM implementation have already been discussed in Chapter 3. However, the results of using lags of 10 days, 20 days, and 30 days for causality detection on the Indian IT stock market will be discussed in more detail in the following sections.

4.6.2.1 NonLinCausality LSTM with ten days lag

The following tables show the foreign stock market, currency exchange rate, and economic articles (from Money Control, the Hindu, and the Financial Times) that have impacted the Indian IT stock market, considering the last ten days of data.

Table 4.7 Most impacted variables considering ten days lag for LSTM nonLinCausality.

Independent Variables	Lag	P_value
USA_INDEX_Price	10	1.49174E-08
DATE_MoneyControl	10	0.029803285
EVENT_MoneyControl	10	0.016785313
FAC_MoneyControl	10	0.000918672
LANGUAGE_MoneyControl	10	5.6008E-09
LAW_MoneyControl	10	0.00034558
ORG_MoneyControl	10	0.006654999

PERSON_MoneyControl	10	0.009890772
TIME_MoneyControl	10	1.49174E-08
CARDINAL_TheHindu	10	0.037947457
EVENT_TheHindu	10	0.011601029
FAC_TheHindu	10	0.002188101
LANGUAGE_TheHindu	10	5.6008E-09
LOC_TheHindu	10	1.29914E-08
NORP_TheHindu	10	0.000109971
PRODUCT_TheHindu	10	5.6008E-09
WORK_OF_ART_TheHindu	10	6.5746E-08
FAC	10	2.75866E-08
LANGUAGE	10	1.03895E-07
LAW	10	5.6008E-09
LOC	10	0.000377328
MONEY	10	3.73552E-06
ORDINAL	10	0.000192589
ORG	10	0.026680223
PERCENT	10	1.63511E-05
PRODUCT	10	1.03895E-07
QUANTITY	10	2.75306E-05
POSITIVE_MoneyControl	10	0.023165394
NEGATIVE_MoneyControl	10	0.00034558

According to the data presented in Table 4.7, there are 29 variables representing foreign stock markets, currency exchange rates, and economic articles from various sources that have impacted the Indian IT stock market. Of these 29 variables, only USA_INDEX_Price, representing the USA market index price, has had an impact on the Indian IT stock market. However, eight named entities from Money Control (DATE_MoneyControl, EVENT_MoneyControl, FAC_MoneyControl, LANGUAGE_MoneyControl, LAW_MoneyControl, ORG_MoneyControl, PERSON_MoneyControl, and TIME_MoneyControl) and two sentiment variables (POSITIVE and NEGATIVE) have also had an impact on the Indian IT stock market.

Similarly, eight named entities (CARDINAL_TheHindu, EVENT_TheHindu, FAC_TheHindu, LANGUAGE_TheHindu, LOC_TheHindu, NORP_TheHindu, PRODUCT_TheHindu, and WORK_OF_ART_TheHindu) from The Hindu have had an impact on the Indian IT stock market. In addition, ten more named entities (FAC, LANGUAGE, LAW, LOC, MONEY, ORDINAL, ORG, PERCENT, PRODUCT, and QUANTITY) extracted from Financial News have had an impact on the Indian IT stock market. However, the LSTM implementation of nonLinCausality with a lag of 10 days did not identify any currency exchange rates that have impacted the Indian IT stock market.

4.6.2.2 NonLinCausality LSTM with 20 days lag

The following tables show the involvement of foreign stock markets, currency exchange rates, and economic articles (from sources such as Money Control, The Hindu, and The Financial Times) that have had the most impact on Indian IT stocks based on data from the past 20 days.

Table 4.8 Most impacted variables considering 20 days lag for LSTM nonLinCausality.

Independent Variables	Lag	P_value
FRANCE_INDEX_Price	20	7.35E-07
LANGUAGE_MoneyControl	20	2.69E-07
LAW_MoneyControl	20	0.002584448
PRODUCT_MoneyControl	20	0.045656699
QUANTITY_MoneyControl	20	0.003216594
TIME_MoneyControl	20	3.56E-07
EVENT_TheHindu	20	0.001464974
LANGUAGE_TheHindu	20	2.69E-07
LOC_TheHindu	20	1.25E-06
ORDINAL_TheHindu	20	0.000806151
PRODUCT_TheHindu	20	2.69E-07
WORK_OF_ART_TheHindu	20	1.02E-05
CARDINAL	20	0.047396203
EVENT	20	9.75E-05

FAC	20	1.91E-05
LANGUAGE	20	4.49E-06
LAW	20	2.69E-07
LOC	20	0.008531319
MONEY	20	7.87E-05
PERCENT	20	0.000208648
PERSON	20	0.010837482
PRODUCT	20	3.24E-07
QUANTITY	20	2.96E-06

As shown in Table 4.8, 23 variables have a p-value less than 0.05, indicating that they are statistically essential and have impacted the Indian IT stock market. Among these 23 variables, one foreign stock market variable, FRANCE_INDEX_Price, has affected the Indian IT stock market. Additionally, five named entities (LANGUAGE_MoneyControl, LAW_MoneyControl, PRODUCT_MoneyControl, QUANTITY_MoneyControl, and TIME_MoneyControl) extracted from economic articles published on Money Control have had an impact on the Indian IT stock market. Similarly, six named entities (EVENT_TheHindu, LANGUAGE_TheHindu, LOC_TheHindu, ORDINAL_TheHindu, PRODUCT_TheHindu, and WORK_OF_ART_TheHindu) extracted from The Hindu's economic articles have had an impact on the Indian IT stock market. Furthermore, eleven named entities (CARDINAL, EVENT, FAC, LANGUAGE, LAW, LOC, MONEY, PERCENT, PERSON, PRODUCT, and QUANTITY) extracted from the Financial Times have had a significant impact on the Indian IT stock market. However, the LSTM implementation of nonLinCausality with a lag of 20 days did not identify any currency exchange rates or economic article sentiments that have impacted the Indian IT stock market.

4.6.2.3 NonLinCausality LSTM with 30 days lag

The following table shows the variables that have impacted the Indian IT stock market based on data from the past 30 days, as determined by the LSTM implementation of nonLinCausality. These variables represent foreign stock markets, currency exchange rates, and economic articles from sources such as Money Control, The Hindu, and The Financial Times.

Table 4.9 Most impacted variables considering 30 days lag for LSTM nonLinCausality.

Independent Variables	Lag	P_value
FRANCE_INDEX_Price	30	1.19E-07
DATE_MoneyControl	30	0.048994064
FAC_MoneyControl	30	1.19E-07
GPE_MoneyControl	30	0.002425671
LANGUAGE_MoneyControl	30	1.19E-07
LAW_MoneyControl	30	0.003726006
MONEY_MoneyControl	30	0.02243793
PERSON_MoneyControl	30	0.006147265
PRODUCT_MoneyControl	30	1.31E-05
TIME_MoneyControl	30	1.19E-07
EVENT_TheHindu	30	3.66E-05
LANGUAGE_TheHindu	30	1.19E-07
LOC_TheHindu	30	1.19E-07
ORDINAL_TheHindu	30	0.028126359
PRODUCT_TheHindu	30	1.19E-07
CARDINAL	30	0.012711287
DATE	30	0.003726006
GPE	30	0.008936405
LAW	30	1.19E-07
MONEY	30	0.001207829
PERCENT	30	0.000729203
PRODUCT	30	2.47E-05
QUANTITY	30	0.001207829
POSITIVE_MoneyControl	30	0.024220467

As shown in Table 4.9, 24 variables represent foreign stock markets, currency exchange rates, and economic articles from different sources that have impacted the Indian IT stock market. Out of these 24 variables, one representation from the foreign stock market, FRANCE_INDEX_Price, was statistically significant, indicating that changes in the French stock market impacted the Indian IT stock market. In addition, nine named entities (such as DATE_Moneycontrol, FAC_Moneycontrol, GPE_Moneycontrol, LANGUAGE_Moneycontrol, LAW_Moneycontrol, MONEY_Moneycontrol, PERSON_Moneycontrol, PRODUCT_Moneycontrol, And TIME_Moneycontrol) And One Sentiment (POSITIVE_Moneycontrol) Extracted from Money Control Articles also impacted the Indian IT Stock Market. Similarly, Five Named Entities (EVENT_Thehindu, LANGUAGE_Thehindu, LOC_The Hindu, ORDINAL_Thehindu, And PRODUCT_Thehindu) extracted from The Hindu articles also had an impact on the Indian IT stock market. Moreover, eight named entities (CARDINAL, DATE, GPE, LAW, MONEY, PERCENT, PRODUCT, and QUANTITY) fetched from Financial Times articles also caused the Indian IT stock market to change. However, even with a lag of 30 days, the LSTM implementation of nonLinCausality did not find a representation in the currency exchange rate that impacted the Indian IT stock market.

As illustrated in Table 4.10, the LSTM implementation of nonLinCausality has identified 40 unique variables from the foreign stock market, currency exchange rate, and economic articles from various sources that impact the Indian IT stock market. Of these 40 variables, two foreign stock market variables, USA_INDEX_Price (S&P 500 Information Technology) and FRANCE_INDEX_Price (CAC Technology) have impacted the Indian

IT stock market. Additionally, 12 named entities (such as DATE_MoneyControl, EVENT_MoneyControl, FAC_MoneyControl, LANGUAGE_MoneyControl, LAW_MoneyControl, ORG_MoneyControl, PERSON_MoneyControl, MONEY_MoneyControl, QUANTITY_MoneyControl, TIME_MoneyControl, GPE_MoneyControl, and PRODUCT_MoneyControl) and two sentiment variables (POSITIVE_MoneyControl and NEGATIVE_MoneyControl) extracted from articles in Money Control have also impacted the Indian IT stock market. Similarly, nine named entities (CARDINAL_TheHindu, FAC_TheHindu, LANGUAGE_TheHindu, LOC_TheHindu, NORP_TheHindu, ORDINAL_TheHindu, PRODUCT_TheHindu, WORK_OF_ART_TheHindu, and EVENT_TheHindu) extracted from The Hindu's articles have had an impact on the Indian IT stock market. Furthermore, 15 named entities (FAC, LANGUAGE, LAW, LOC, MONEY, ORDINAL, ORG, PERCENT, PRODUCT, QUANTITY, CARDINAL, EVENT, PERSON, DATE, and GPE) extracted from Financial Times articles have also impacted the Indian IT stock market. However, the LSTM implementation of nonCausality did not identify any representation from the currency exchange rate that has impacted the Indian IT stock market. To visualize the impact of these variables, Figure 4.10 plots the variables on the Y-axis and the number of times each variable occurred at different time lags (10 days, 20 days, and 30 days) on the X-axis. Variables with a count greater than one can be considered more confident than others. Therefore, the following 24 variables in the table below can be considered more impactful than the others.

Table 4.10 Most impacted variables summary for LSTM nonLinCausality

Independent Variables	LSTM implementation of nonLinCausality		
	10 days lag	20 days lag	30 days lag
USA_INDEX_Price	Yes	No	No
DATE_MoneyControl	Yes	No	Yes
EVENT_MoneyControl	Yes	No	No
FAC_MoneyControl	Yes	No	Yes
LANGUAGE_MoneyControl	Yes	Yes	Yes
LAW_MoneyControl	Yes	Yes	Yes
ORG_MoneyControl	Yes	No	No
PERSON_MoneyControl	Yes	No	Yes
TIME_MoneyControl	Yes	Yes	Yes
CARDINAL_TheHindu	Yes	No	No
EVENT_TheHindu	Yes	Yes	Yes
FAC_TheHindu	Yes	No	No
LANGUAGE_TheHindu	Yes	Yes	Yes
LOC_TheHindu	Yes	Yes	Yes
NORP_TheHindu	Yes	No	No
PRODUCT_TheHindu	Yes	Yes	Yes
WORK_OF_ART_TheHindu	Yes	Yes	No
FAC	Yes	Yes	No
LANGUAGE	Yes	Yes	No
LAW	Yes	Yes	Yes
LOC	Yes	Yes	No
MONEY	Yes	Yes	Yes
ORDINAL	Yes	No	No
ORG	Yes	No	No
PERCENT	Yes	Yes	Yes
PRODUCT	Yes	Yes	Yes
QUANTITY	Yes	Yes	Yes
POSITIVE_MoneyControl	Yes	No	Yes
NEGATIVE_MoneyControl	Yes	No	No
FRANCE_INDEX_Price	No	Yes	Yes
PRODUCT_MoneyControl	No	Yes	Yes
QUANTITY_MoneyControl	No	Yes	No
ORDINAL_TheHindu	No	Yes	Yes
CARDINAL	No	Yes	Yes
EVENT	No	Yes	No
PERSON	No	Yes	No
GPE_MoneyControl	No	No	Yes

MONEY_MoneyControl	No	No	Yes
DATE	No	No	Yes
GPE	No	No	Yes

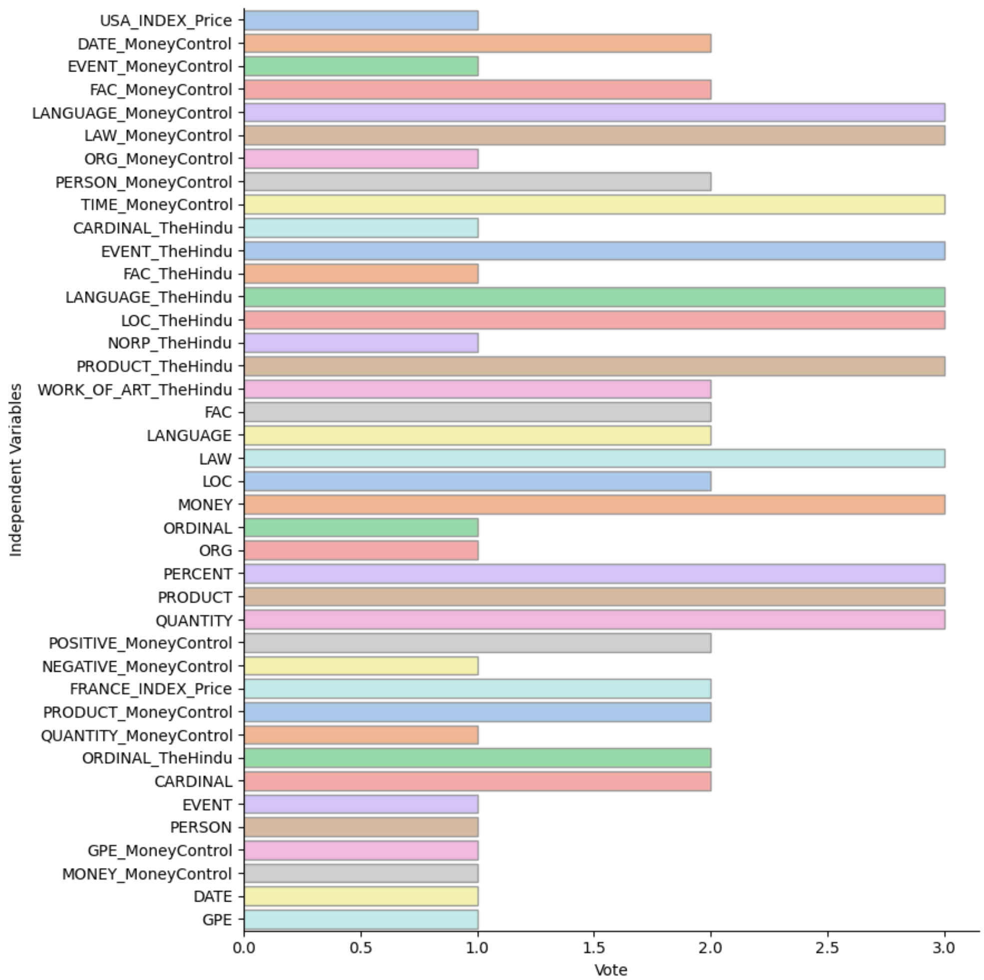

Figure 4.10 Independent Variables vs. Vote count for nonLinCausality LSTM

4.7 DoWhy

The DoWhy framework is an open-source Python framework. It is structured to facilitate causal inference in a specific domain. The DoWhy library allows for specifying some underlying assumptions about the observed data and validating those assumptions with statistical estimators. Thus, DoWhy offers a four-step framework to make causal inferences by focusing on the specified assumptions. These are the four steps: (1) forming a causal graph and verifying its clarity through domain expertise, (2) identifying all potential causes and their corresponding effects, (3) estimating causality for a given assumption using the backdoor linear regression method, and (4) refuting or validating the assumption using the Random Common Cause (RCC), Placebo Treatment Refuter (PTR), and Data Subset Refuter (DSR) methods. This study used the random common cause (RCC), placebo treatment refuter (PTR), and data subset refuter (DSR) to check whether the specified assumptions were correct or not. The details of these refutation methods have been discussed in the following sections.

4.7.1 DoWhy Results (RCC)

The Random Common Cause (RCC) method involves the addition of randomly drawn covariates to the data and re-running the analysis to assess whether the causal estimate changes. If the original assumption was correct, the causal estimate should stay the same due to adding the covariates. The results of the RCC method are presented below.

Table 4.11 RCC refutation results

Independent Variables	Estimated_effect	New_effect	Difference
GERMAN_INDEX_Price	-0.159190815	-0.159203136	1.2321E-05
UK_INDEX_Price	0.06838958	0.068393139	-3.55877E-06
GBP_INR_Price	-0.214033711	-0.214040553	6.84197E-06
MUR_INR_Price	0.077370318	0.07736461	5.7081E-06
DATE_MoneyControl	0.017212837	0.017223721	-1.0884E-05
EVENT_MoneyControl	-0.027840172	-0.027806019	-3.41526E-05
GPE_MoneyControl	-0.007174199	-0.007160389	-1.38105E-05
LANGUAGE_MoneyControl	0.004126419	0.004100723	2.56954E-05
LAW_MoneyControl	-0.005867838	-0.005900556	3.27187E-05
LOC_MoneyControl	0.001077046	0.001096568	-1.95218E-05
NORP_MoneyControl	-0.025029411	-0.025030652	1.24108E-06
PERCENT_MoneyControl	-0.020209576	-0.020205073	-4.50343E-06

PERSON_MoneyControl	0.033780842	0.033818077	-3.72353E-05
PRODUCT_MoneyControl	0.006067674	0.006094012	-2.63377E-05
QUANTITY_MoneyControl	0.002863017	0.002888729	-2.57119E-05
TIME_MoneyControl	0.007693349	0.007691448	1.90107E-06
WORK_OF_ART_MoneyControl	0.027471589	0.027461196	1.03932E-05
CARDINAL_TheHindu	0.009715857	0.00969649	1.93672E-05
DATE_TheHindu	0.010767449	0.010759055	8.39414E-06
EVENT_TheHindu	-0.00460704	-0.004599342	-7.69873E-06
FAC_TheHindu	-0.005589853	-0.005590985	1.13196E-06
GPE_TheHindu	0.002935048	0.002943837	-8.78953E-06
LANGUAGE_TheHindu	-0.00241961	-0.00241534	-4.26938E-06
LAW_TheHindu	-0.000828421	-0.000827399	-1.02213E-06
LOC_TheHindu	-0.002289362	-0.002276227	-1.31344E-05
NORP_TheHindu	-0.001729978	-0.001724704	-5.27437E-06
ORDINAL_TheHindu	-0.007017165	-0.007026745	9.58E-06
ORG_TheHindu	-0.002877715	-0.002888648	1.09328E-05
PERCENT_TheHindu	-0.015146661	-0.015133336	-1.33252E-05
PERSON_TheHindu	0.000876839	0.000881747	-4.90753E-06
QUANTITY_TheHindu	0.005600658	0.005614524	-1.38658E-05
TIME_TheHindu	0.00503788	0.005025445	1.24357E-05
EVENT	-0.009120336	-0.009143495	2.31591E-05
GPE	-0.058311088	-0.058329161	1.80733E-05
LANGUAGE	0.01885496	0.018862685	-7.72544E-06
LOC	-0.018737993	-0.018710014	-2.79791E-05
PERCENT	-0.049423202	-0.049439732	1.65292E-05
PRODUCT	-0.018710349	-0.018736155	2.58061E-05
QUANTITY	0.000435648	0.000364068	7.15796E-05
TIME	-0.040750859	-0.040836033	8.51743E-05
WORK_OF_ART	0.069726479	0.069660387	6.60918E-05
POSITIVE_MoneyControl	0.011290008	0.011281765	8.24313E-06
NEGATIVE_MoneyControl	-0.032305211	-0.032257037	-4.81735E-05
POSITIVE_TheHindu	0.012121471	0.012098221	2.32503E-05
NEGATIVE_TheHindu	0.000202642	0.000223981	-2.13391E-05
POSITIVE	-0.001254799	-0.001277382	2.25833E-05
NEGATIVE	-0.006099034	-0.006091528	-7.50641E-06
USD_INR_Price	0.102305495	0.10234274	-3.72454E-05

As seen in Table 4.11, there are 48 variables for which there was only a minimal difference between the Estimated Effect and the New Effect. These variables, which include representations of foreign stock markets, currency exchange rates, and economic articles, are believed to impact the Indian IT stock market. The breakdown of these variables is as follows: Two variables (GERMAN_INDEX_Price and UK_INDEX_Price) out of 48 represent the German (DAX Software (CXPSX)) and UK (FTSE 350 Software & Computer Services (FTNMX101010)) stock markets, respectively, and are believed to affect the Indian IT stock market. Three currency exchange rates (GBP_INR_Price, MUR_INR_Price, and USD_INR_Price) are believed to impact the Indian IT stock market. Furthermore, there are 13 named entities (such as DATE_MoneyControl, EVENT_MoneyControl, GPE_MoneyControl, LANGUAGE_MoneyControl, LAW_MoneyControl, LOC_MoneyControl, NORP_MoneyControl, PERCENT_MoneyControl, PERSON_MoneyControl, PRODUCT_MoneyControl, QUANTITY_MoneyControl, TIME_MoneyControl, and WORK_OF_ART_MoneyControl) and two sentiments (POSITIVE_MoneyControl and NEGATIVE_MoneyControl) retrieved from articles published on Money Control that are believed to have influenced the Indian IT stock market. Similarly, there are 15 named entities (CARDINAL_TheHindu, DATE_TheHindu, EVENT_TheHindu, FAC_TheHindu, GPE_TheHindu, LANGUAGE_TheHindu, LAW_TheHindu, LOC_TheHindu, NORP_TheHindu, ORDINAL_TheHindu, ORG_TheHindu, PERCENT_TheHindu, PERSON_TheHindu, QUANTITY_TheHindu, and TIME_TheHindu) and two sentiments (POSITIVE_TheHindu and NEGATIVE_TheHindu) extracted from articles in The Hindu that are believed to have impacted the Indian IT stock market. Moreover, there are nine named entities (EVENT,

GPE, LANGUAGE, LOC, PERCENT, PRODUCT, QUANTITY, TIME, and WORK_OF_ART) and two sentiments (POSITIVE and NEGATIVE) extracted from the Financial Times that are believed to have affected the Indian IT stock market. While all these variables suggest a causal impact on the Indian IT stock market, further investigation using the Placebo Treatment Refuter (PTR) and Data Subset Refuter (DSR) methods of refutation is necessary to increase confidence in these findings.

4.7.2 DoWhy results (PTR)

The Placebo Treatment Refuter (PTR) method involves randomly assigning any covariate variable as a treatment variable and re-running the analysis to determine the validity of the original assumptions. If the assumptions are correct, the newly obtained estimation should be close to zero. The results of the PTR refutation method are presented below.

Table 4.12 PTR refutation results

Independent Variables	Estimated_effect	New_effect
FRANCE_INDEX_Price	0.476616592	0
GERMAN_INDEX_Price	-0.159190815	-3.89E-16
UK_INDEX_Price	0.06838958	-1.95E-16
USA_INDEX_Price	0.545187518	1.30E-16
JAPAN_INDEX_Price	0.033250022	2.60E-16
EUR_INR_Price	0.043742784	3.89E-16
JPY_INR_Price	-0.24407207	6.49E-17
GBP_INR_Price	-0.214033711	2.60E-16
SGD_INR_Price	0.283114871	3.24E-16
MUR_INR_Price	0.077370318	6.49E-17
CARDINAL_MoneyControl	-0.012955838	0
DATE_MoneyControl	0.017212837	-1.95E-16
EVENT_MoneyControl	-0.027840172	0
FAC_MoneyControl	-0.002243125	-1.30E-16
GPE_MoneyControl	-0.007174199	0
LANGUAGE_MoneyControl	0.004126419	-2.60E-16
LAW_MoneyControl	-0.005867838	1.95E-16

LOC_MoneyControl	0.001077046	-1.30E-16
MONEY_MoneyControl	-0.017854417	1.30E-16
NORP_MoneyControl	-0.025029411	0
ORDINAL_MoneyControl	-0.008748361	1.30E-16
ORG_MoneyControl	-0.007016846	-3.89E-16
PERCENT_MoneyControl	-0.020209576	2.60E-16
PERSON_MoneyControl	0.033780842	2.60E-16
PRODUCT_MoneyControl	0.006067674	-1.95E-16
QUANTITY_MoneyControl	0.002863017	1.95E-16
TIME_MoneyControl	0.007693349	-6.49E-17
WORK_OF_ART_MoneyControl	0.027471589	-6.49E-17
CARDINAL_TheHindu	0.009715857	-6.49E-17
DATE_TheHindu	0.010767449	3.24E-16
EVENT_TheHindu	-0.00460704	-2.60E-16
FAC_TheHindu	-0.005589853	0
GPE_TheHindu	0.002935048	1.30E-16
LANGUAGE_TheHindu	-0.00241961	1.95E-16
LAW_TheHindu	-0.000828421	3.24E-16
LOC_TheHindu	-0.002289362	1.30E-16
MONEY_TheHindu	-0.002560479	-1.95E-16
NORP_TheHindu	-0.001729978	6.49E-17
ORDINAL_TheHindu	-0.007017165	0
ORG_TheHindu	-0.002877715	2.60E-16
PERCENT_TheHindu	-0.015146661	2.60E-16
PERSON_TheHindu	0.000876839	-1.30E-16
PRODUCT_TheHindu	-0.001361481	-3.24E-16
QUANTITY_TheHindu	0.005600658	0
TIME_TheHindu	0.00503788	-6.49E-17
WORK_OF_ART_TheHindu	-0.002470009	3.24E-16
CARDINAL	0.01061017	2.43E-17
DATE	-0.136596213	3.45E-17
EVENT	-0.009120336	9.13E-18
FAC	0.002654352	4.06E-18
GPE	-0.058311088	2.94E-17
LANGUAGE	0.01885496	-7.40E-17
LAW	-0.032312129	-3.75E-17
LOC	-0.018737993	1.83E-17
MONEY	-0.01038766	1.22E-17
NORP	0.056511207	-6.08E-18
ORDINAL	0.006673706	6.08E-18

ORG	-0.043415533	0
PERCENT	-0.049423202	1.32E-17
PERSON	0.079228176	1.93E-17
PRODUCT	-0.018710349	-4.87E-17
QUANTITY	0.000435648	7.10E-18
TIME	-0.040750859	-1.62E-17
WORK_OF_ART	0.069726479	1.62E-17
POSITIVE_MoneyControl	0.011290008	-1.95E-16
NEGATIVE_MoneyControl	-0.032305211	5.19E-16
POSITIVE_TheHindu	0.012121471	-1.95E-16
NEGATIVE_TheHindu	0.000202642	4.54E-16
POSITIVE	-0.001254799	-2.60E-16
NEGATIVE	-0.006099034	-2.60E-16
USD_INR_Price	0.102305495	-6.49E-17

As shown in Table 4.12, all 71 variables have a new effect close to zero, according to the PTR method of refutation. That suggests that variables representing foreign stock markets, currency exchange rates, and economic articles obtained from various sources did indeed impact the Indian IT stock market. Specifically, four out of 71 variables represent foreign stock markets (FRANCE_INDEX_Price, GERMAN_INDEX_Price, UK_INDEX_Price, USA_INDEX_Price, and JAPAN_INDEX_Price) that are believed to have influenced the Indian IT stock market. In addition, there are five currency exchange rates (EUR_INR_Price, JPY_INR_Price, GBP_INR_Price, SGD_INR_Price, and MUR_INR_Price) that may have affected the Indian IT stock market according to the PTR refutation technique. Additionally, 18 named entities (such as CARDINAL_MoneyControl, DATE_MoneyControl, EVENT_MoneyControl, FAC_MoneyControl, GPE_MoneyControl, LANGUAGE_MoneyControl, LAW_MoneyControl, LOC_MoneyControl, MONEY_MoneyControl, NORP_MoneyControl, ORDINAL_MoneyControl, ORG_MoneyControl, PERCENT_MoneyControl, PERSON_MoneyControl, PRODUCT_MoneyControl, QUANTITY_MoneyControl, TIME_MoneyControl, and WORK_OF_ART_MoneyControl) and two sentiment variables (POSITIVE_MoneyControl and NEGATIVE_MoneyControl) extracted from articles published on Money Control are believed to have had an impact on the Indian IT stock market. Similarly, there are 18 named entities (CARDINAL_TheHindu, DATE_TheHindu, EVENT_TheHindu, FAC_TheHindu, GPE_TheHindu, LANGUAGE_TheHindu, LAW_TheHindu, LOC_TheHindu, MONEY_TheHindu, NORP_TheHindu, ORDINAL_TheHindu, ORG_TheHindu, PERCENT_TheHindu, PERSON_TheHindu, PRODUCT_TheHindu, QUANTITY_TheHindu, TIME_TheHindu, and WORK_OF_ART_TheHindu) and two sentiment variables (POSITIVE_TheHindu and NEGATIVE_TheHindu) retrieved from articles in The Hindu that are believed to have impacted the Indian IT stock market. Moreover, there are 18 named entities (CARDINAL, DATE, EVENT, FAC, GPE, LANGUAGE, LAW, LOC, MONEY, NORP, ORDINAL, ORG, PERCENT, PERSON, PRODUCT, QUANTITY, TIME, and WORK_OF_ART) and two sentiment variables (POSITIVE and NEGATIVE) extracted from the Financial Times that are believed to have influenced the Indian IT stock market. Analysis of the PTR refutation technique results shows that it supports all the assumptions made in the framing phase of the DoWhy library.

4.7.3 DoWhy Results (DSR)

The Data Subset Refuter (DSR) method involves creating a random subset of the data, like cross-validation, and examining whether the causal estimates vary across the subsets. If the original assumptions were correct, there should not be a significant deviation from the initially estimated effect. The following table presents the results of the DSR method.

Table 4.13 DSR refutation results

Independent Variables	Estimated_effect	New_effect
USA_INDEX_Price	0.545187518	0.5454652
CARDINAL_MoneyControl	-0.012955838	-0.0134291
FAC_MoneyControl	-0.002243125	-0.0018031
LOC_MoneyControl	0.001077046	0.000857
ORDINAL_MoneyControl	-0.008748361	-0.0085177
ORG_MoneyControl	-0.007016846	-0.0070873
PRODUCT_MoneyControl	0.006067674	0.0062318
QUANTITY_MoneyControl	0.002863017	0.0027802
TIME_MoneyControl	0.007693349	0.0078907
CARDINAL_TheHindu	0.009715857	0.0098657
GPE_TheHindu	0.002935048	0.0028301
ORDINAL_TheHindu	-0.007017165	-0.0071536
ORG_TheHindu	-0.002877715	-0.0028756
PERCENT_TheHindu	-0.015146661	-0.0149654
PERSON_TheHindu	0.000876839	0.0008526
LOC	-0.018737993	-0.0194715
ORDINAL	0.006673706	0.0067291
POSITIVE_MoneyControl	0.011290008	0.0108725
POSITIVE_TheHindu	0.012121471	0.0118272
POSITIVE	-0.001254799	-0.0009638
NEGATIVE	-0.006099034	-0.0058304

As shown in Table 4.13, according to the DSR method, 21 variables representing foreign stock markets, currency exchange rates, and economic articles from various sources have impacted the Indian IT stock market. Of these 21 variables, one represents a foreign stock market (USA_INDEX_Price) and has influenced the Indian IT stock market. Additionally, eight named entities (CARDINAL_MoneyControl, FAC_MoneyControl, LOC_MoneyControl, ORDINAL_MoneyControl, ORG_MoneyControl, PRODUCT_MoneyControl, QUANTITY_MoneyControl, and

TIME_MoneyControl) and one sentiment variable (POSITIVE_MoneyControl) extracted from articles published on Money Control are believed to have impacted the Indian IT stock market. Similarly, six named entities (CARDINAL_TheHindu, GPE_TheHindu, ORDINAL_TheHindu, ORG_TheHindu, PERCENT_TheHindu, and PERSON_TheHindu) and one sentiment variable (POSITIVE_TheHindu) retrieved from articles in The Hindu are believed to have had an impact on the Indian IT stock market. Moreover, two named entities (LOC and ORDINAL) and two sentiment variables (POSITIVE and NEGATIVE) are believed to have influenced the Indian IT stock market. However, according to the results of the DSR method, none of the currency exchange rate representations have impacted the Indian IT stock market.

Table 4.14 summarizes the results of the DoWhy framework using different refutation methods. As previously mentioned, the PTR refutation method found that all variables representing foreign stock markets, currency exchange rates, and economic articles impacted the Indian IT stock market. However, the RCC and DSR refutation methods found that 54 and 21 variables impacted the Indian IT stock market. As shown in Table 4.14, each refutation method has a value of "Yes" or "No" for each independent variable, with "Yes" indicating that the independent variable does impact the Indian IT stock market and "No" indicating that it does not. In Figure 4.11, the y-axis represents the independent variables that may impact the Indian IT stock market, and the x-axis shows the votes of the RCC, PTR, and DSR refutation methods of the DoWhy library. The independent variables that receive at least two votes from the different refutation methods can be considered the more confident variables that may affect the Indian IT stock market. These most voted variables are:

GERMAN_INDEX_Price, *UK_INDEX_Price*, *GBP_INR_Price*, *MUR_INR_Price*, *DATE_MoneyControl*, *EVENT_MoneyControl*, *GPE_MoneyControl*, *LANGUAGE_MoneyControl, LAW_MoneyControl, LOC_MoneyControl, NORP_MoneyControl, PERCENT_MoneyControl, PERSON_MoneyControl, PRODUCT_MoneyControl, QUANTITY_MoneyControl, TIME_MoneyControl, WORK_OF_ART_MoneyControl, CARDINAL_TheHindu, DATE_TheHindu, EVENT_TheHindu, FAC_TheHindu, GPE_TheHindu, LANGUAGE_TheHindu, LAW_TheHindu, LOC_TheHindu, NORP_TheHindu, ORDINAL_TheHindu, ORG_TheHindu, PERCENT_TheHindu, PERSON_TheHindu, QUANTITY_TheHindu, TIME_TheHindu, EVENT, GPE, LANGUAGE, LOC, PERCENT, PRODUCT, QUANTITY, TIME, WORK_OF_ART, POSITIVE_MoneyControl, NEGATIVE_MoneyControl, POSITIVE_TheHindu, NEGATIVE_TheHindu, POSITIVE, NEGATIVE, USD_INR_Price, USA_INDEX_Price, CARDINAL_MoneyControl, FAC_MoneyControl, ORDINAL_MoneyControl, ORG_MoneyControl, and ORDINAL.*

Table 4.14 Summary of all refutation methods

Independent Variables	DoWhy Results Refuter		
	RCC	PTR	DSR
GERMAN_INDEX_Price	Yes	Yes	No
UK_INDEX_Price	Yes	Yes	No
GBP_INR_Price	Yes	Yes	No
MUR_INR_Price	Yes	Yes	No
DATE_MoneyControl	Yes	Yes	No
EVENT_MoneyControl	Yes	Yes	No
GPE_MoneyControl	Yes	Yes	No
LANGUAGE_MoneyControl	Yes	Yes	No
LAW_MoneyControl	Yes	Yes	No
LOC_MoneyControl	Yes	Yes	Yes
NORP_MoneyControl	Yes	Yes	No
PERCENT_MoneyControl	Yes	Yes	No
PERSON_MoneyControl	Yes	Yes	No
PRODUCT_MoneyControl	Yes	Yes	Yes
QUANTITY_MoneyControl	Yes	Yes	Yes
TIME_MoneyControl	Yes	Yes	Yes
WORK_OF_ART_MoneyControl	Yes	Yes	No

CARDINAL_TheHindu	Yes	Yes	Yes
DATE_TheHindu	Yes	Yes	No
EVENT_TheHindu	Yes	Yes	No
FAC_TheHindu	Yes	Yes	No
GPE_TheHindu	Yes	Yes	Yes
LANGUAGE_TheHindu	Yes	Yes	No
LAW_TheHindu	Yes	Yes	No
LOC_TheHindu	Yes	Yes	No
NORP_TheHindu	Yes	Yes	No
ORDINAL_TheHindu	Yes	Yes	Yes
ORG_TheHindu	Yes	Yes	Yes
PERCENT_TheHindu	Yes	Yes	Yes
PERSON_TheHindu	Yes	Yes	Yes
QUANTITY_TheHindu	Yes	Yes	No
TIME_TheHindu	Yes	Yes	No
EVENT	Yes	Yes	No
GPE	Yes	Yes	No
LANGUAGE	Yes	Yes	No
LOC	Yes	Yes	Yes
PERCENT	Yes	Yes	No
PRODUCT	Yes	Yes	No
QUANTITY	Yes	Yes	No
TIME	Yes	Yes	No
WORK_OF_ART	Yes	Yes	No
POSITIVE_MoneyControl	Yes	Yes	Yes
NEGATIVE_MoneyControl	Yes	Yes	No
POSITIVE_TheHindu	Yes	Yes	Yes
NEGATIVE_TheHindu	Yes	Yes	No
POSITIVE	Yes	Yes	Yes
NEGATIVE	Yes	Yes	Yes
USD_INR_Price	Yes	Yes	No
FRANCE_INDEX_Price	No	Yes	No
USA_INDEX_Price	No	Yes	Yes
JAPAN_INDEX_Price	No	Yes	No
EUR_INR_Price	No	Yes	No
JPY_INR_Price	No	Yes	No
SGD_INR_Price	No	Yes	No
CARDINAL_MoneyControl	No	Yes	Yes
FAC_MoneyControl	No	Yes	Yes
MONEY_MoneyControl	No	Yes	No

ORDINAL_MoneyControl	No	Yes	Yes
ORG_MoneyControl	No	Yes	Yes
MONEY_TheHindu	No	Yes	No
PRODUCT_TheHindu	No	Yes	No
WORK_OF_ART_TheHindu	No	Yes	No
CARDINAL	No	Yes	No
DATE	No	Yes	No
FAC	No	Yes	No
LAW	No	Yes	No
MONEY	No	Yes	No
NORP	No	Yes	No
ORDINAL	No	Yes	Yes
ORG	No	Yes	No
PERSON	No	Yes	No

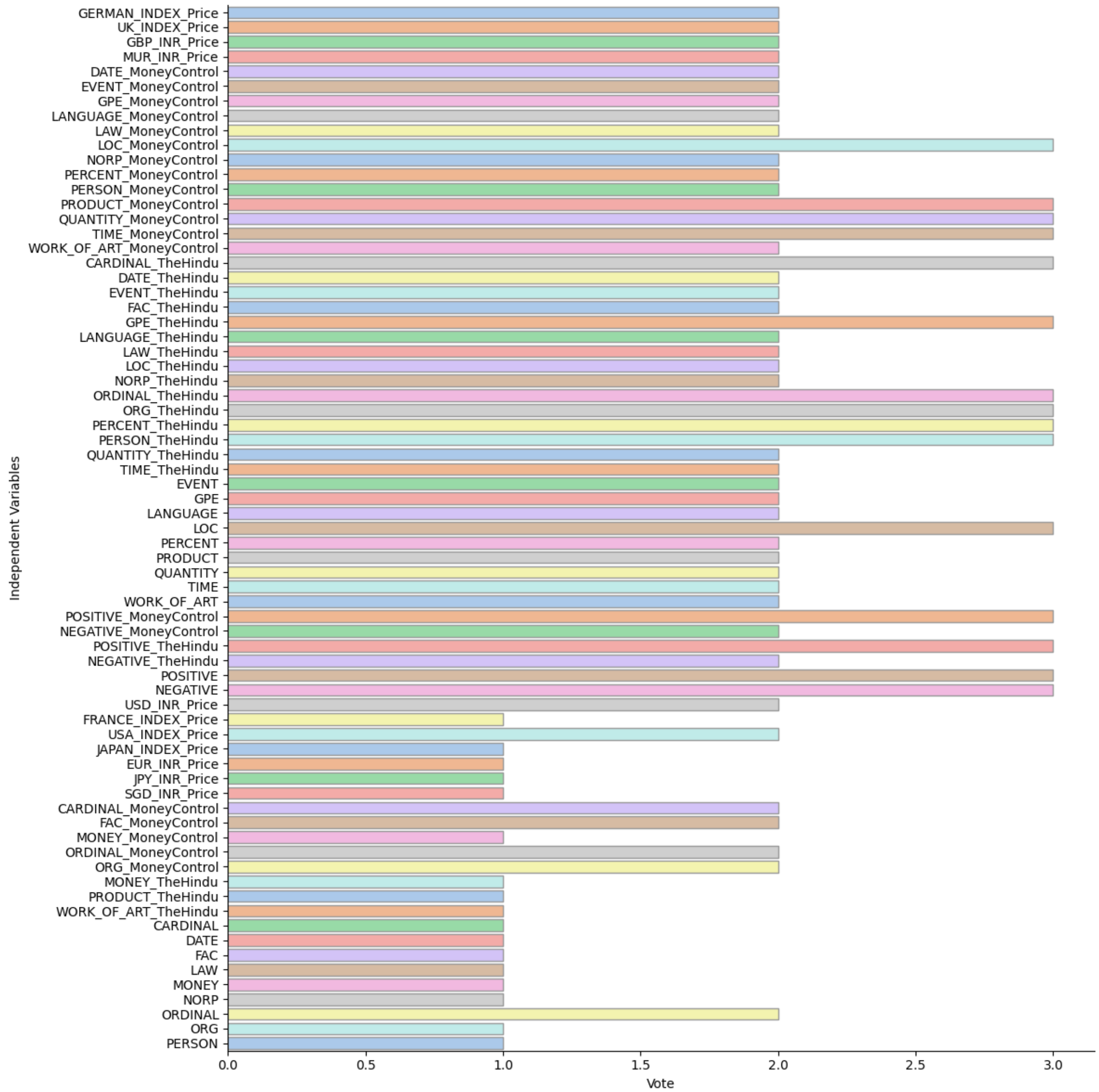

Figure 4.11 Independent Variables vs. Vote count for DoWhy's refutation methods

4.8 Temporal Causal Discovery Framework

The Temporal Causal Discovery Framework (TCDF) is a deep learning-based approach implemented in Python for identifying the causal relationships between different time series. Its inputs consist of multiple time series data, such as variables representing foreign IT indices, currency exchange rates, named entities, sentiments extracted from economic articles, and the Indian IT index. TCDF employs an attention-based convolutional neural network with a causal validation step to produce a causal graph displaying the relationships between the various time series. It can also detect time delays between cause and effect after the parameters of the convolutional network have been tuned. This study used TCDF to assess the impact of foreign IT stock markets, currency exchange rates, and economic articles on the Indian IT stock market. The results of TCDF are presented below.

Table 4.15 Results of TCDF with time-lag

Variable 1	Variable 2	Time Lag
Nifty_Price	EVENT	2791
Nifty_Price	LOC	2713
Nifty_Price	MONEY	415
Nifty_Price	ORDINAL	2976
Nifty_Price	PERCENT	2460
Nifty_Price	EVENT_MoneyControl	3034
Nifty_Price	LOC_MoneyControl	3027
Nifty_Price	NORP_MoneyControl	235
Nifty_Price	QUANTITY_MoneyControl	2964
Nifty_Price	TIME_MoneyControl	3002
Nifty_Price	LOC_TheHindu	2687
Nifty_Price	MONEY_TheHindu	2906
Nifty_Price	NORP_TheHindu	2967
Nifty_Price	PERCENT_TheHindu	2741
Nifty_Price	PERSON_TheHindu	2923
Nifty_Price	POSITIVE_MoneyControl	550
Nifty_Price	POSITIVE_TheHindu	890
Nifty_Price	NEGATIVE_TheHindu	698
Nifty_Price	POSITIVE	948
Nifty_Price	NEGATIVE	853

Table 4.15 illustrates the relationships between three variables: the impactor (Variable 1), the impacted (Variable 2), and the time lag (the number of days prior to the

impact). The table also shows instances where independent variables impact other independent variables. As shown in Table 22, 20 variables are impacted by the Indian IT stock market. However, none of these 20 variables represent the foreign IT stock market or currency exchange rate impacted by the Indian IT stock market. Additionally, out of the 20 variables, five named entities and one sentiment from the articles of Money Control, five named entities, and two sentiments from The Hindu and The Financial Times articles have been impacted by the Indian IT stock market.

Furthermore, in all the records listed in Table 4.15, the Indian IT stock market is observed as the impactor variable rather than the impacted variable, which indicates that changes in the Indian IT stock market did impact the variables mentioned above, but the reverse was not true. Therefore, the results obtained from the TCDF algorithm did not provide any information about the impact of foreign IT stock markets, currency exchange rates, or economic articles on the Indian IT stock market. However, it is worth noting that there are 20 variables that the Indian IT stock market has impacted.

4.9 Granger Causality

Granger causality is a statistical method used to determine the causal relationship between two time-series variables. It is based on the premise that one variable, if useful for forecasting another variable, is said to "Granger-cause" that variable. On the other hand, if a variable is not useful for forecasting another variable, it is said to "fail to Granger-cause" that variable. It is important to note that the term "causality" in Granger causality is somewhat of a misnomer, as it only refers to the ability to forecast and does not necessarily imply a true causal relationship between the variables. Despite this limitation, Granger causality is the most widely used statistical technique for verifying the usefulness of one variable in forecasting another. In this study, the results of Granger causality were captured with lags of 10, 20, and 30 timesteps in addition to the results of machine-learning and deep-learning-based methods. The following sections provide a detailed description of these results.

4.9.1 Granger Causality with ten days lag

The table below illustrates the independent variables that have impacted the Indian IT stock market with a lag of 10 days. These variables include foreign stock market performance, currency exchange rates, and economic news articles.

Table 4.16 Most impacted variables considering ten days lag for Granger Causality

Independent Variable	Lag	P_value
FRANCE_INDEX_Price	10	0
GERMAN_INDEX_Price	10	0.0002
UK_INDEX_Price	10	0
USA_INDEX_Price	10	0
JAPAN_INDEX_Price	10	0.0001
JPY_INR_Price	10	0.0107
SGD_INR_Price	10	0.0025
FAC_MoneyControl	10	0.0065
GPE_MoneyControl	10	0.0236
PRODUCT_MoneyControl	10	0.015
WORK_OF_ART_MoneyControl	10	0.0102
CARDINAL_TheHindu	10	0.0473
LANGUAGE_TheHindu	10	0.0125
LOC_TheHindu	10	0.0084
LANGUAGE	10	0.0212
NORP	10	0.0237
PERCENT	10	0.027
PRODUCT	10	0.0005
POSITIVE	10	0.0166

As illustrated in Table 4.16, the statistical analysis results indicate that 19 independent variables have a statistically significant impact on the Indian IT stock market, as evidenced by a p-value less than 0.05. Among these variables, five represent foreign IT stock markets, including the FRANCE_INDEX_Price, GERMAN_INDEX_Price, UK_INDEX_Price, USA_INDEX_Price, and JAPAN_INDEX_Price. Additionally, two variables, JPY_INR_PRICE and SGD_INR_PRICE, represent currency exchange rates. Furthermore, four variables, namely FAC_MoneyControl, GPE_MoneyControl, PRODUCT_MoneyControl, and WORK_OF_ART_MoneyControl, represent named entities extracted from economic articles published on MoneyControl and were found to have an impact on the Indian IT stock market. Furthermore, out of the remaining eight independent variables, three represent named entities (CARDINAL_TheHindu, LANGUAGE_TheHindu, and LOC_TheHindu) extracted from articles published in The Hindu, which also had an impact on the Indian IT stock market. Similarly, four named entities, such as LANGUAGE, NORP, PERCENT, and PRODUCT, and a POSITIVE sentiment extracted from the Financial Times, were found to impact the Indian IT stock market. It is worth noting that sentiments extracted from Money Control and The Hindu did not appear to impact the Indian IT stock market over the 10-day lag period.

4.9.2 Granger Causality with 20 days lag

The independent variables in the following table represent the financial IT stock market, currency exchange rates, and economic articles that have been found to impact the Indian IT stock market, according to the Granger Causality Test with a 20-day lag.

Table 4.17 Most impacted variables considering 20 days lag for Granger Causality

Independent Variable	Lag	P_value
FRANCE_INDEX_Price	20	0
GERMAN_INDEX_Price	20	0.0002
UK_INDEX_Price	20	0
USA_INDEX_Price	20	0
JAPAN_INDEX_Price	20	0.0001
JPY_INR_Price	20	0.0107
SGD_INR_Price	20	0.0025
FAC_MoneyControl	20	0.0065
GPE_MoneyControl	20	0.0236
LANGUAGE_MoneyControl	20	0.0026
PRODUCT_MoneyControl	20	0.015
QUANTITY_MoneyControl	20	0.0121
WORK_OF_ART_MoneyControl	20	0.0102
CARDINAL_TheHindu	20	0.0473
GPE_TheHindu	20	0.0306
LANGUAGE_TheHindu	20	0.0125
LAW_TheHindu	20	0.0396
LOC_TheHindu	20	0.0084
LANGUAGE	20	0.0212
NORP	20	0.0237
PERCENT	20	0.027
PRODUCT	20	0.0004
POSITIVE	20	0.0166
USD_INR_Price	20	0.0429

As depicted in Table 4.17, the statistical analysis results indicate that 24 independent variables have a statistically significant impact on the Indian IT stock market, as evidenced by a p-value less than 0.05. These variables include representations of foreign IT stock markets, currency exchange rates, and economic articles. Specifically, five variables represent the foreign IT stock market (FRANCE_INDEX_Price, GERMAN_INDEX_Price, UK_INDEX_Price, USA_INDEX_Price, and JAPAN_INDEX_Price), and according to the Granger Causality Test, all these foreign IT stock markets have had an impact on the Indian IT stock market. Additionally, JPY_INR_Price, SGD_INR_Price, and USD_INR_Price represents currency exchange rates that were also found to impact the Indian IT stock market. Furthermore, six named entities (FAC_MoneyControl, GPE_MoneyControl, LANGUAGE_MoneyControl, PRODUCT_MoneyControl, QUANTITY_MoneyControl, and WORK_OF_ART_MoneyControl) extracted from economic articles published on Money Control were found to have an impact on Indian IT stock market. Additionally, five named entities (CARDINAL_TheHindu, GPE_TheHindu, LANGUAGE_TheHindu, LAW_TheHindu, and LOC_TheHindu) extracted from articles published in The Hindu were found to have an impact on the Indian IT stock market. Similarly, four named entities (LANGUAGE, NORP, PERCENT, and PRODUCT) and a sentiment (POSITIVE) extracted from the Financial Times were found to impact the Indian IT stock market. It is important to note that Table 24 did not include sentiment representations (positive or negative) for Money Control and The Hindu that have impacted the Indian IT stock market.

4.9.3 Granger Causality with 30 days lag

The independent variables in the following table represent the foreign IT stock market, currency exchange rates, and economic articles from Money Control, The Hindu, and Financial Times that have been found to have an impact on the Indian IT stock market, according to the Granger Causality test with a 30-day lag.

Table 4.18 Most impacted variables considering 30 days lag for Granger Causality

Independent Variable	Lag	P_value
FRANCE_INDEX_Price	30	0
GERMAN_INDEX_Price	30	0.0002
UK_INDEX_Price	30	0
USA_INDEX_Price	30	0

JAPAN_INDEX_Price	30	0.0001
JPY_INR_Price	30	0.0057
SGD_INR_Price	30	0.0025
FAC_MoneyControl	30	0.0065
GPE_MoneyControl	30	0.0236
LANGUAGE_MoneyControl	30	0.0026
PRODUCT_MoneyControl	30	0.015
QUANTITY_MoneyControl	30	0.0121
WORK_OF_ART_MoneyControl	30	0.0102
CARDINAL_TheHindu	30	0.0473
EVENT_TheHindu	30	0.0363
GPE_TheHindu	30	0.019
LANGUAGE_TheHindu	30	0.0021
LAW_TheHindu	30	0.0396
LOC_TheHindu	30	0.0084
LANGUAGE	30	0.0212
NORP	30	0.0237
PERCENT	30	0.027
PRODUCT	30	0.0002
POSITIVE	30	0.0166
USD_INR_Price	30	0.0429

Table 4.18 presents the Granger Causality Test results with a 30-day lag period, indicating that 25 independent variables significantly impact the Indian IT stock market. Among these, five variables representing foreign IT stock markets (FRANCE_INDEX_Price, GERMAN_INDEX_Price, UK_INDEX_Price, USA_INDEX_Price, and JAPAN_INDEX_Price) have been found to have an impact on the Indian IT stock market. Furthermore, three variables representing currency exchange rates (JPY_INR_Price, SGD_INR_Price, and USD_INR_Price) also impact the Indian IT stock market. Additionally, six named entities (FAC_MoneyControl, GPE_MoneyControl, LANGUAGE_MoneyControl, PRODUCT_MoneyControl, QUANTITY_MoneyControl, and WORK_OF_ART_MoneyControl) extracted from economic articles published on Money Control have an impact on the Indian IT stock market. Furthermore, six named

entities (CARDINAL_TheHindu, EVENT_TheHindu, GPE_TheHindu, LANGUAGE_TheHindu, LAW_TheHindu, and LOC_TheHindu) extracted from economic articles published in The Hindu have an impact on the Indian IT stock market. Similarly, four named entities (LANGUAGE, NORP, PERCENT, and PRODUCT) and a sentiment (POSITIVE) extracted from the Financial Times impact the Indian IT stock market. It is worth noting that this study found no impact of sentiments of Money Control and The Hindu on the Indian IT stock market, even over a 30-day lag period.

According to Table 4.19, the Granger Causality Test with 10-, 20-, and 30-day lag periods identified 25 unique representations of the foreign IT stock market, currency exchange rates, and economic articles that impacted the Indian IT stock market. Of these, five foreign IT stock market representations (FRANCE_INDEX_Price, GERMAN_INDEX_Price, UK_INDEX_Price, USA_INDEX_Price, and JAPAN_INDEX_Price) were identified by the Granger Causality Test in all three lag periods. They had an impact on the Indian IT stock market. Three representations of the currency exchange rate (JPY_INR_Price, SGD_INR_Price, and USD_INR_Price) impacted the Indian IT stock market. Six named entities (FAC_MoneyControl, GPE_MoneyControl, LANGUAGE_MoneyControl, PRODUCT_MoneyControl, QUANTITY_MoneyControl, and WORK_OF_ART_MoneyControl) were extracted from articles published on Money Control had an impact on the Indian IT stock market. Similarly, six named entities (GPE_TheHindu, LAW_TheHindu, MONEY_TheHindu, FAC_TheHindu, LOC_TheHindu, and ORDINAL_TheHindu) were extracted from economic articles published in The Hindu and had an impact on the Indian IT stock market.

Likewise, four named entities (LANGUAGE, NORP, PERCENT, and PRODUCT) and a sentiment (POSITIVE) extracted from the Financial Times impact the Indian IT stock market. The study did not show any impact of Money Control and Hindu sentiments on the Indian IT stock market. Figure 4.12 shows a chart of independent variables versus vote count for the Granger Causality Test. On the X-axis, the number of times a variable was identified by the Granger Causality Test with 10, 20, and 30-day lag periods. The independent variables representing the foreign IT stock market, currency exchange rates, and economic articles published on Money Control and The Hindu are shown on the Y-axis. As shown in Figure 4.12, the independent variables are:

Table 4.19 Most impacted variables summary for Granger Causality

Independent Variable	Granger Causality Test		
	10 days lag	20 days lag	30 days lag
FRANCE_INDEX_Price	Yes	Yes	Yes
GERMAN_INDEX_Price	Yes	Yes	Yes
UK_INDEX_Price	Yes	Yes	Yes
USA_INDEX_Price	Yes	Yes	Yes
JAPAN_INDEX_Price	Yes	Yes	Yes
JPY_INR_Price	Yes	Yes	Yes
SGD_INR_Price	Yes	Yes	Yes
FAC_MoneyControl	Yes	Yes	Yes
GPE_MoneyControl	Yes	Yes	Yes
PRODUCT_MoneyControl	Yes	Yes	Yes
WORK_OF_ART_MoneyControl	Yes	Yes	Yes
CARDINAL_TheHindu	Yes	Yes	Yes
LANGUAGE_TheHindu	Yes	Yes	Yes
LOC_TheHindu	Yes	Yes	Yes
LANGUAGE	Yes	Yes	Yes
NORP	Yes	Yes	Yes
PERCENT	Yes	Yes	Yes
PRODUCT	Yes	Yes	Yes
POSITIVE	Yes	Yes	Yes

LANGUAGE_MoneyControl	No	Yes	Yes
QUANTITY_MoneyControl	No	Yes	Yes
GPE_TheHindu	No	Yes	Yes
LAW_TheHindu	No	Yes	Yes
USD_INR_Price	No	Yes	Yes
EVENT_TheHindu	No	No	Yes

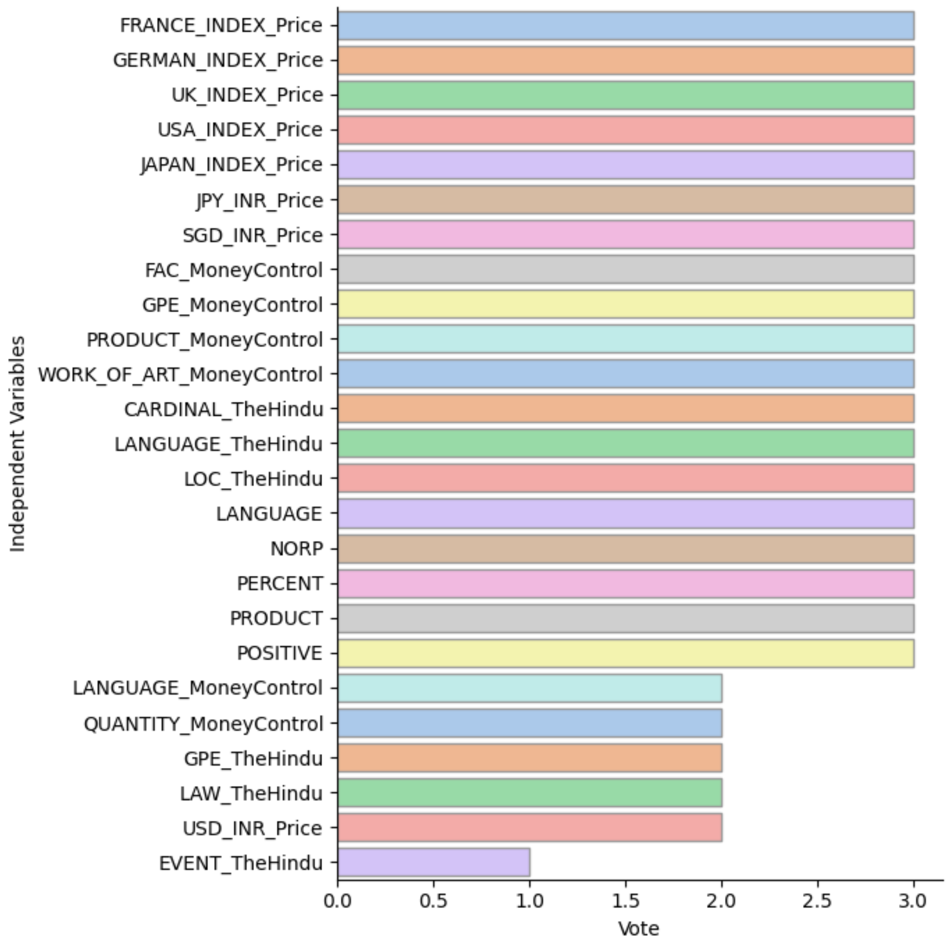

Figure 4.12 Independent Variables vs. Vote count for Granger Causality methods

4.10 Summary

This chapter began by analyzing the relationship between representations of the foreign IT stock market, currency exchange rates, economic articles, and the Indian IT stock market using bivariate analysis. The Pearson and Spearman correlation tests were then used to evaluate these relationships numerically. However, these tests only show whether the independent variables are correlated with the Indian IT stock market. To confidently determine whether the observed correlation between the independent variables and the Indian IT stock market also causes the Indian IT stock market, this study used statistical, machine learning, and deep-learning-based causality detection methods. Four causality detection algorithms were used in this study to increase confidence in the impacting independent variables: NonLinCausality, DoWhy, TCDF, and the Granger Causality Test. The results of multiple implementations of NonLinCausality (LSTM and GRU with 10-, 20-, and 30-day lags), DoWhy (different refutation methods such as RCC, PTR, and DSR), TCDF, and Granger causality (with 10-, 20-, and 30-day lags) were considered in this study. However, the results of TCDF algorithms provided an interesting insight that the named entities and sentiments of economic articles published in Money Control, The Hindu, and The Financial Times may be influenced by the Indian IT stock market. The focus of this study was to determine the impact of foreign IT stock markets, currency exchange rates, and economic articles on the Indian IT stock market rather than the other way around. Hence, TCDF implementation was not included in this study's methodology.

In this chapter, the results of the study are presented. The final chapter of this thesis will provide a detailed interpretation of the findings about the research questions, discuss the practical implications of the results, and suggest directions for future potential research.

CHAPTER V:

DISCUSSION

5.1 Introduction

The outcome of this research provides the impact of foreign stock markets, currency exchange rates, and economic articles on the Indian IT stock market. However, the findings of this research should be understood and interpreted concerning the aim and questions of the research. Thus, this chapter evaluated the results of Granger causality, machine learning, and deep learning-based causality detection methods per the research purpose, question, and hypothesis.

Moreover, this chapter reflects the research process. As a result, it discusses the research setting's limitations, potential consequences, and implications of interpreting the results. The chapter closes with recommendations for future research.

5.2 Evaluating the Research Objective.

The present research employs quantitative methods to analyze the impact of foreign IT stock markets, currency exchange rates, and economic articles on the Indian IT stock market. The methods employed in this study included machine-learning-based techniques, deep-learning-based techniques, and the Granger Test for causality detection. This ensemble of techniques, referred to as ensemble learning, as described by (Valentini, 2012), combines multiple models to solve a single problem. The primary rationale for utilizing ensemble methods is to merge the predictions of machine-learning-based, deep-learning-based, and statistical-based Granger Tests to produce more accurate and stable predictions than would be obtained from a single model.

Additionally, ensemble methods allow for both linear and non-linear data, as the literature review suggests that the Granger causality test is suitable for linear data. In contrast, machine-learning-based and deep-learning-based methods are more effective for non-linear data. Specifically, this research utilizes machine-learning-based methods (DoWhy), deep-learning-based methods (NonLinCausality (GRU) and NonLinCausality (LSTM)), and the statistical-based Granger Test for causality detection. In the following sections, the research will evaluate the objectives, research questions, and hypotheses using the ensemble method outputs.

5.3 Research objective 1 – Examine the impact of the foreign IT stock market on the Indian IT stock market using ensemble methods.

The present research endeavors to evaluate the impact of foreign IT stock markets on the Indian IT stock market. Thus, the study meticulously considered a selection of foreign IT stock markets, including Germany, the United Kingdom, the United States, France, India, and Japan, for analysis to achieve this objective. The information came from a secondary source, specifically the website www.investing.com. Subsequently, the study applied and examined the results of machine-learning-based, deep-learning-based, and statistical-based methods to gain insights into the impact of foreign IT stock markets on the Indian IT stock market. The subsequent sections will provide a detailed examination of the results obtained from each method, followed by a comprehensive analysis of the combined results to arrive at more accurate, stable, and reliable predictions.

5.3.1 Examine the impact of the foreign IT stock market using machine learning models.

The research used the “DoWhy” library to represent machine-learning-based causality detection methods. One of the assumptions of the research was that foreign IT stock markets do have an impact on the Indian IT stock market. The research has collected the result from three different refutation methods, such as refutation Random Common Cause (RCC), Placebo Treatment Refuter (PTR), and Data Subset Refuter (DSR), to check whether the specified assumption was correct. The details of the refutation methods have already been discussed in the earlier sections. The following table represents the impact of the foreign IT stock market on the Indian IT stock market using different refutation methods of the DoWhy library.

Table 5.1 Most impacted foreign IT variables using DoWhy

Independent Variables	DoWhy Results Refuter		
	RCC	PTR	DSR
GERMAN_INDEX_Price	Yes	Yes	No
UK_INDEX_Price	Yes	Yes	No
FRANCE_INDEX_Price	No	Yes	No
USA_INDEX_Price	No	Yes	Yes
JAPAN_INDEX_Price	No	Yes	No

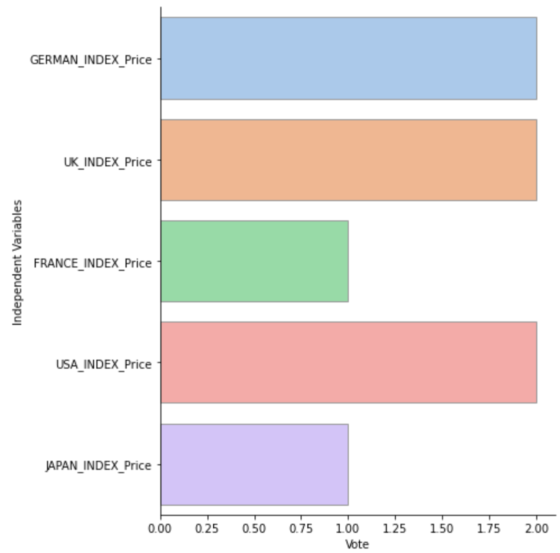

Figure 5.1 Most impacted foreign IT variables vs. Vote count for DoWhy.

The table presented displays all the independent variables that have been found to impact the Indian IT stock market, as per at least one of the refutation methods used. As depicted in the table, the German and UK IT stock markets have been determined to impact the Indian IT stock market, as per the Recursive Conditional Comparison (RCC) and Pairwise Transferability Ratio (PTR) methods of the DoWhy refutation library. Furthermore, the USA IT stock market has also been identified to impact the Indian IT stock market, per the PTR and Directed Subsampling Ratio (DSR) refutation methods of the DoWhy library. France and Japanese IT stock markets have also been found to impact the Indian IT stock market, as per only the PTR refutation method.

Consequently, to increase the confidence level, this study has considered the foreign IT representation that received the most votes: German, UK, and USA IT stock markets. More specifically, as per the results of the DoWhy library, the German, UK, and USA IT stock markets have been determined to impact the Indian IT stock market.

5.3.2 Examine the impact of the foreign IT stock market using a deep-learning-based method.

The research employs the nonLinCausality Library, which represents deep-learning-based techniques, to detect causality. To achieve this, the study implements the Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) versions of the nonLinCausality Library, as these methods have been demonstrated to be highly effective in handling time-series data. The study analyzes the impact of foreign IT stock markets on the Indian IT stock market, utilizing the LSTM and GRU implementations of non-linear causality, and focuses on lag times of 10, 20, and 30 days. The results are presented in tables representing the foreign IT stock market and its impact on the Indian IT stock market, using LSTM or GRU implementation of nonLinCausality Library for the specified lag times.

Table 5.2 Most impacted foreign IT variables using nonLinCausality (GRU)

Independent Variables	GRU implementation of nonLinCausality		
	10 days	20 days	30 days
FRANCE_INDEX_Price	Yes	Yes	Yes
USA_INDEX_Price	No	Yes	No

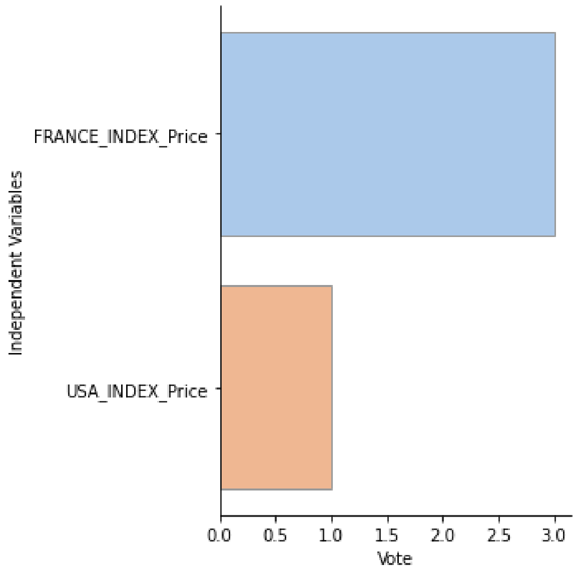

Figure 5.2 Most impacted foreign IT variables vs. Vote count for nonLinCausality (GRU).

Table 5.3 Most impacted foreign IT variables using nonLinCausality (LSTM)

Independent Variables	LSTM implementation of nonLinCausality		
	10 days lag	20 days lag	30 days lag
USA_INDEX_Price	Yes	No	No
FRANCE_INDEX_Price	No	Yes	Yes

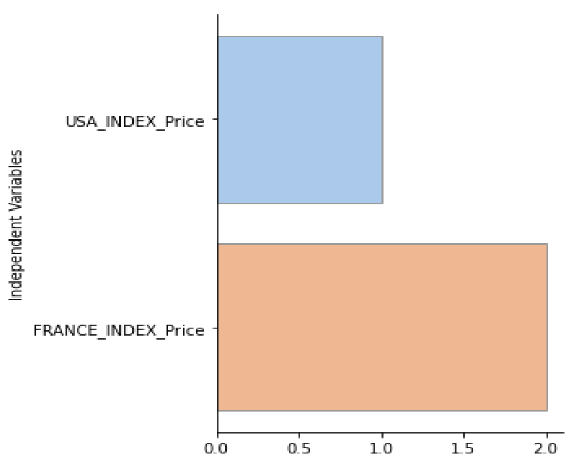

Figure 5.3 Most impacted foreign IT variables vs. Vote count for nonLinCausality (LSTM).

As demonstrated by the data presented in Tables 5.2 and 5.3, the USA IT stock market and the France IT stock market have significantly impacted the Indian IT stock market. Specifically, the results indicate that, for both the Long Short-term Memory (LSTM) and Gated Recurrent Unit (GRU) implementations of non-linear causality, the USA IT stock market has impacted the Indian IT stock market with a lag of 10 days. Additionally, the results reveal that, for both LSTM and GRU implementations, the France IT stock market has impacted the Indian IT stock market with lag times of 20 and 30 days. However, it is worth noting that, in the case of the GRU implementation, the foreign IT stock market has even impacted the Indian IT stock market with 10 days lag. Therefore, the above findings suggest that the USA IT and France IT stock markets have causatively affected the Indian IT stock market.

5.3.3 Examine the impact of the foreign IT stock market using a statistical-based method.

This research has used the Granger-Causality detection test to determine the causality relationship between the foreign IT stock market and the Indian IT stock market. The Granger-Causality test is the most widely used statistical method for verifying the usefulness of one variable in forecasting another variable. Thus, this research has analyzed the impact of the foreign IT stock market on the Indian IT stock market using lag times of 10 days, 20 days, and 30 days. The following table provides the result of the Granger Causality test with a lag of 10 days, 20 days, and 30 days.

Table 5.4 Most impacted foreign IT variables using Granger Causality Test (GRU)

Independent Variable	Granger Causality Test

	10 days lag	20 days lag	30 days lag
FRANCE_INDEX_Price	Yes	Yes	Yes
GERMAN_INDEX_Price	Yes	Yes	Yes
UK_INDEX_Price	Yes	Yes	Yes
USA_INDEX_Price	Yes	Yes	Yes
JAPAN_INDEX_Price	Yes	Yes	Yes

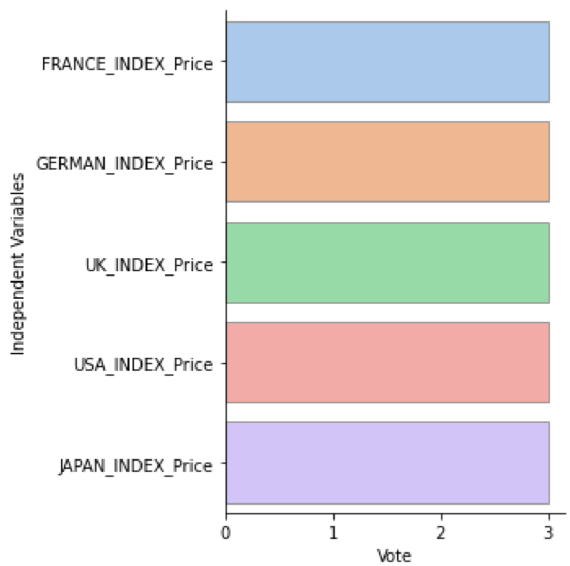

Figure 5.4 Most impacted foreign IT variables vs. Vote count for Granger Causality.

As demonstrated by table 5.4, all the foreign IT stock markets included in this research have been found to impact the Indian IT stock market, as determined by the Granger Causality Test, with a time lag of 10, 20, and 30 days. These results can be attributed to the fact that the Granger Causality Test is particularly effective in detecting causality when the dataset is linear. As previously outlined in the methodology section, the FRANCE_INDEX_Price, GERMAN_INDEX_Price, UK_INDEX_Price, USA_INDEX_Price, JAPAN_INDEX_Price, and NIFTY_INDEX_Price is represented by daily closing prices and are linear. Therefore, the above results strongly indicate that, per the Granger Causality Test, the France, German, UK, USA, and Japanese IT indices have a causal impact on the Indian IT stock market.

5.3.4 Examine the impact of the foreign IT stock market using ensemble methods.

As discussed in the preceding sections, this research has employed a variety of machine-learning, deep-learning, and statistical-based causality detection methods to analyze the impact of foreign IT stock markets on the Indian IT stock market. Each method has provided insights into how foreign IT stock markets impact the Indian IT stock market. According to the DoWhy library (a machine-learning-based method), the German, UK, and USA IT stock markets have impacted the Indian IT stock market. Meanwhile, the nonLinCausality methods (a deep-learning-based method) have determined that the USA IT and France IT stock markets have impacted the Indian IT stock market. Additionally, as per the Granger Causality (a statistical-based method), the France, German, UK, USA, and Japanese IT indices have been found to have a causal impact on the Indian IT stock market.

This study has considered the findings of all three approaches to obtain a more accurate result. The following table presents a consolidated result of all the causality detection methods used in this research.

Table 5.5 Most impacted foreign IT variables using ensemble methods (GRU)

Independent Variable	Impact on the Indian IT stocks		
	Statistical-based (Granger Test)	Machine-learning-based (DoWhy)	Deep-learning-based (nonLinCausality)
FRANCE_INDEX_Price	Yes	No	Yes
GERMAN_INDEX_Price	Yes	Yes	No
UK_INDEX_Price	Yes	Yes	No

USA_INDEX_Price	Yes	Yes	Yes
JAPAN_INDEX_Price	Yes	No	No

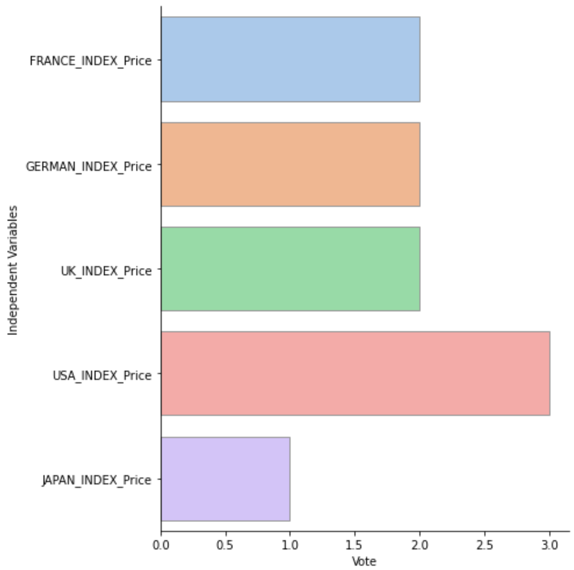

Figure 5.5 Most impacted foreign IT variables vs. Vote count for ensemble methods.

As illustrated in table 5.5, the results of the statistical-based, machine-learning-based, and deep-learning-based methods used in this research provide a strong indication that the **France, USA, UK, and German IT stock markets** have a causal impact on the Indian IT stock market. Furthermore, as depicted in Figure 4.1, there is a direct relationship between the France, USA, UK, and German IT stock markets and the Indian IT stock market. Specifically, when the closing prices of the France, USA, UK, and German IT stock markets increase, likely, the Indian IT stock market will also increase. This finding addresses one of the research questions, "*How do foreign IT indices affect the Indian IT stock market*" and supports the hypothesis that an increase in foreign IT stock markets leads to an increase in the Indian IT stock market.

5.4 Research objective 2 – Examine the impact of significant currency exchange rates on the Indian IT stock market using ensemble methods.

The present research endeavors to evaluate the impact of the currency exchange rate on the Indian IT stock market to achieve this objective, the study meticulously considered a selection of currency exchange rates, including Germany (EUR), the United Kingdom (EUR), the United States (USD), France (EUR), Japan (JPY), Singapore (SGD), and Mauritius (MUR) against India (INR) for analysis. The information came from a secondary source, specifically the website www.investing.com. Subsequently, the study applied and examined the results of machine-learning-based, deep-learning-based, and statistical-based methods to gain insights into the impact of foreign IT stock markets on the Indian IT stock market. The subsequent sections provide a detailed examination of the results obtained from each method, followed by a comprehensive analysis of the combined results to arrive at more accurate, stable, and reliable predictions.

5.4.1 Examine the impact of currency exchange rates using a machine-learning-based method.

The present study employed the DoWhy Library to represent machine-learning-based causality detection methods. One of the assumptions of the research was that the currency exchange rate of foreign currency against Indian currency would have an impact on the Indian IT stock market. To verify this assumption, the study collected results from three different refutation methods, namely the Random Common Cause (RCC), Placebo Treatment Refuter (PTR), and Data Subset Refuter (DSR). The details of these refutation methods were discussed in earlier sections of the thesis. The following table illustrates the impact of the foreign currency exchange rate against Indian currency on the Indian IT stock market, as determined by the different refutation methods of the DoWhy library.

Table 5.6 Most impacted currency exchange variables using DoWhy.

Independent Variables	DoWhy Results Refuter		
	RCC	PTR	DSR
GBP_INR_Price	Yes	Yes	No
MUR_INR_Price	Yes	Yes	No
USD_INR_Price	Yes	Yes	No
EUR_INR_Price	No	Yes	No
JPY_INR_Price	No	Yes	No
SGD_INR_Price	No	Yes	No

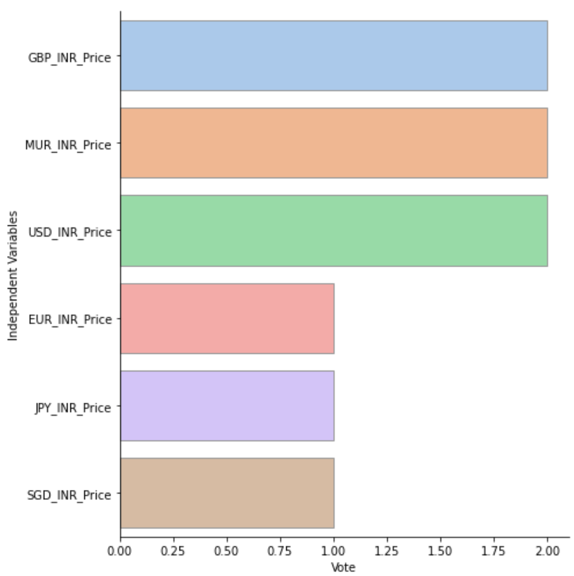

Figure 5.6 Most impacted currency exchange variables vs. Vote count for doWhy.

Table 5.6 illustrates all the independent variables that have been found to have an impact on the Indian IT stock market, as determined by at least one of the refutation methods employed in the study. As can be seen from the table, the GBP-INR, MUR-INR, and USD-INR, currency exchange rates, have been identified to have an impact on the Indian IT stock market, as per the Recursive Conditional Comparison (RCC) and Pairwise Transferability Ratio (PTR) methods of the “DoWhy” refutation library. Additionally, the EUR_INR, JPY_INR, and SGD_INR currency exchange rates have also been determined to impact the Indian IT stock market, as per the PTR refutation methods of the "DoWhy" library.

This study increased the confidence level of the results by considering the currency exchange rate representation that received the most votes: the GBP-INR, MUR-INR, and USD-INR exchange rates. More specifically, as per the results of the DoWhy library, the GBP-INR, MUR-INR, and USD-INR exchange rates have been determined to impact the Indian IT stock market significantly.

5.4.2 Examine the impact of currency exchange rates using a deep-learning-based method.

The research employs NonLinCausality Library, which represents the deep-learning-based techniques to detect causality. The study achieved the same by implementing the Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) versions of the NonLinCausality library, as these methods have been demonstrated to be highly effective in handling time-series data. The study analyzes the impact of significant currency exchange rates on the Indian IT stock market, utilizing the LSTM and GRU implementations of non-linear causality, and focuses on lag times of 10, 20, and 30 days. The results are presented in tables representing the currency exchange rates and their impact on the Indian IT stock market, using LSTM or GRU implementation of nonLinCausality Library for the specified lag times.

Table 5.7 Most impacted currency exchange variables using nonLinCausality (GRU).

Independent Variables	GRU implementation of nonLinCausality		
	10 days	20 days	30 days
MUR_INR_Price	No	No	Yes

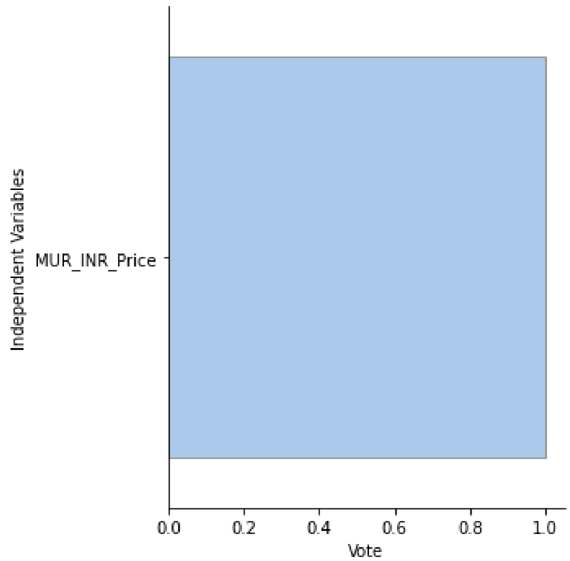

Figure 5.7 Most impacted currency exchange variables vs. Vote count for nonLinCausality (GRU).

As demonstrated by the data presented in Table 5.7, the MUR_INR exchange rate has significantly impacted the Indian IT stock market. Specifically, the results indicate that using Gated Recurrent Unit (GRU) implementations of non-linear causality, the MUR_INR exchange rate has impacted the Indian IT stock market with a lag of 30 days. Conversely, the results reveal that, for Long Short-Term Memory (LSTM) implementations, no currency exchange rate impacted the Indian IT stock market with lag times of 10, 20, and 30 days.

It is worth noting that this research has identified only one currency exchange rate with the nonLinCausality method, and that too with only a 30-day time lag. However, it is still worthwhile to consider the MUR_INR currency exchange rate as having a causative effect on the Indian IT stock market.

5.4.3 Examine the impact of currency exchange rates on the Indian IT stock market using a statistical-based method.

This research has used the Granger-Causality detection test to determine the causality relationship between the foreign IT stock market and the Indian IT stock market. The Granger-Causality test is the most widely used statistical method for verifying the usefulness of one variable in forecasting another variable. Thus, this research has analyzed the impact of significant currency exchange on the Indian IT stock market using 10 days, 20 days, and 30 days lag times. The following table provides the result of the Granger Causality test with a lag of 10 days, 20 days, and 30 days.

Table 5.8 Most impacted currency exchange variables using Granger Causality.

Independent Variable	Granger Causality Test		
	10 days lag	20 days lag	30 days lag
JPY_INR_Price	Yes	Yes	Yes
SGD_INR_Price	Yes	Yes	Yes
USD_INR_Price	No	Yes	Yes

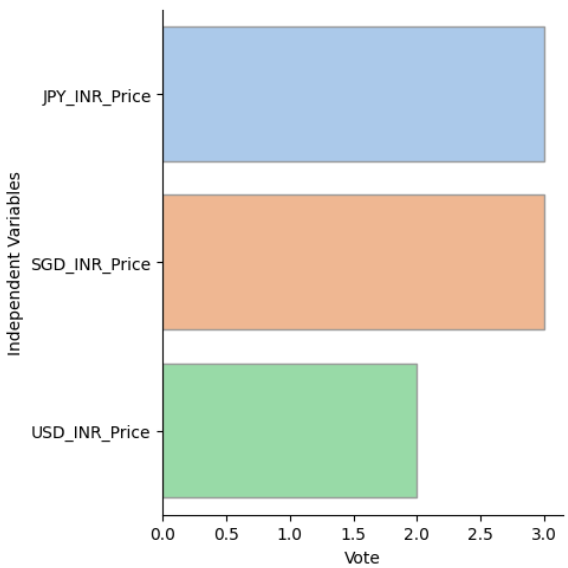

Figure 5.8 Most impacted currency exchange variables vs. Vote count for Granger causality.

The results of the Granger Causality Test, with lags of 10, 20, and 30 days, indicate that the exchange rates of JPY_INR_Price and SGD_INR_Price have a significant impact on the Indian IT stock market. Furthermore, the test results reveal that the USD_INR_Price also impacts the Indian IT stock market, with a 20 and 30-day lag. As demonstrated in the table, at least two lags of Granger causality indicate that the exchange rates of JPY_INR_Price, SGD_INR_Price, and USD_INR_Price have a causal impact on the Indian IT stock market. Therefore, it can be concluded that these exchange rates play a crucial role in influencing the Indian IT stock market, as determined by the Granger Causality Test.

5.4.4 Examine the impact of currency exchange rate on the Indian IT stock market using ensemble methods.

As discussed in the preceding sections, this research employed a variety of machine-learning, deep-learning, and statistical-based causality detection methods to analyze the impact of currency exchange rates on the Indian IT stock market. These methods provided insights into which currency exchange rates impact the Indian IT stock market. According to the DoWhy library (a machine-learning-based method), the GBP-INR, MUR-INR, USD-INR, EUR_INR, JPY_INR, and SGD_INR currency exchange rates have had an impact on the Indian IT stock market. Meanwhile, the nonLinCausality methods (a deep-learning-based method) determined that the MUR_INR currency exchange rate impacted the Indian IT stock market. Additionally, as per the Granger Causality Test (a statistical-based method), the JPY_INR, SGD_INR, and USD_INR currency exchange rate was found to have a causal impact on the Indian IT stock market.

This study considered the findings of all three approaches to obtain a more accurate result. The following table presents a consolidated result of all the causality detection methods used in this research.

Table 5.9 Most impacted currency exchange variables using ensemble methods.

Independent Variable	Impact on the Indian IT stocks		
	Statistical based (Granger Test)	Machine-learning-based (DoWhy)	Deep-learning-based (nonLinCausality)
GBP_INR_Price	No	Yes	No
MUR_INR_Price	No	Yes	Yes
USD_INR_Price	Yes	Yes	No
EUR_INR_Price	No	Yes	No
JPY_INR_Price	Yes	Yes	No
SGD_INR_Price	Yes	Yes	No

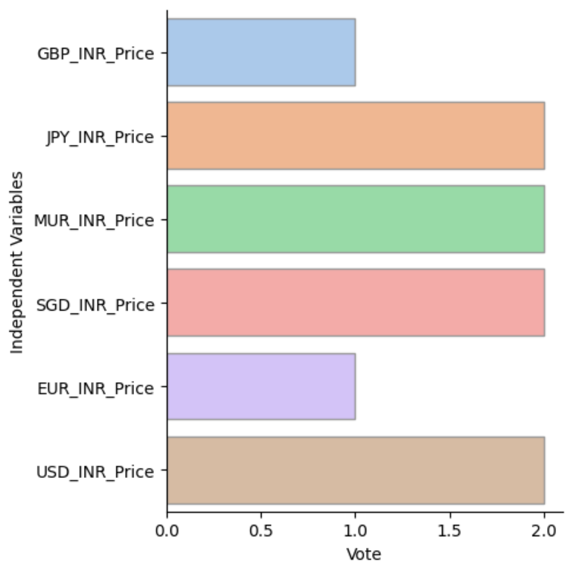

Figure 5.9 Most impacted currency exchange variables vs. Vote count for ensemble methods.

As illustrated in table 5.9, the results of the statistical-based, machine-learning-based, and deep-learning-based methods used in this research provide strong indications that the MUR_INR, USD_INR, JPY_INR, and SGD_INR currency exchange rates have a causal impact on the Indian IT stock market. Furthermore, as depicted in Figure 4.2, the

Indian IT stock market shares a positive relationship with the USD_INR, JPY_INR, and SGD_INR currency exchange rates. In contrast, it shares an inverse relationship with the MUR-INR currency exchange rate. Specifically, when the currency exchange rate of USD_INR, JPY_INR, and SGD_INR increases, likely, the Indian IT stock market will also increase. However, when the currency exchange rate of MUR-INR decreases, the Indian IT stock market increases and vice-versa.

This finding addresses one of the research questions, "*How do currency exchange rates affect the Indian IT stock market*" and supports the hypothesis that an increase in the USD_INR, JPY_INR, and SGD_INR currency exchange rate leads to an increase in the Indian IT stock market. In contrast, a decrease in the MUR-INR currency exchange rate leads to an increase in the Indian IT stock market.

5.5 Research objective 3 – Examine the impact of text features and sentiments of economic articles on the Indian IT stock market using ensemble methods.

The present research endeavors to evaluate the impact of text features and sentiments of economic articles from three different sources on the Indian IT stock market. The study achieved the same objective by meticulously considering three economic articles, Money Control, The Hindu, and The Financial Times, for analysis. Specifically, the research has named entities such as CARDINAL, DATE, EVENT, FAC, GPE, LANGUAGE, LAW, LOC, MONEY, NORP, ORDINAL, ORG, PERCENT, PERSON, PRODUCT, QUANTITY, TIME, and WORK_OF_ART and sentiments like POSITIVE, and NEGATIVE to represent the economic articles from different economic sources. After that, the study applied and examined the results of machine-learning-based, deep-learning-based, and statistical-based methods to gain insights into the impact of economic articles on the Indian IT stock market. The subsequent sections will provide a detailed examination of the results obtained from each method, followed by a comprehensive analysis of the combined results to arrive at more accurate, stable, and reliable predictions.

5.5.1 Examine the impact of text features and sentiments of economic articles using a machine-learning-based method.

The present study employed the DoWhy Library to represent machine-learning-based causality detection methods. One of the research assumptions was that economic articles' text features and sentiments would impact the Indian IT stock market. The study collected results from three different refutation methods, namely the Random Common Cause (RCC), Placebo Treatment Refuter (PTR), and Data Subset Refuter (DSR), to verify this assumption. The details of these refutation methods were discussed in earlier sections of the thesis. The following table illustrates the impact of the economic articles on the Indian IT stock market, as determined by the different refutation methods of the DoWhy library.

Table 5.10 Most impacted economic article variables using doWhy.

Independent Variables	DoWhy Results Refuter		
	RCC	PTR	DSR
DATE_MoneyControl	Yes	Yes	No
EVENT_MoneyControl	Yes	Yes	No
GPE_MoneyControl	Yes	Yes	No
LANGUAGE_MoneyControl	Yes	Yes	No
LAW_MoneyControl	Yes	Yes	No
LOC_MoneyControl	Yes	Yes	Yes
NORP_MoneyControl	Yes	Yes	No
PERCENT_MoneyControl	Yes	Yes	No
PERSON_MoneyControl	Yes	Yes	No
PRODUCT_MoneyControl	Yes	Yes	Yes

QUANTITY_MoneyControl	Yes	Yes	Yes
TIME_MoneyControl	Yes	Yes	Yes
WORK_OF_ART_MoneyControl	Yes	Yes	No
CARDINAL_TheHindu	Yes	Yes	Yes
DATE_TheHindu	Yes	Yes	No
EVENT_TheHindu	Yes	Yes	No
FAC_TheHindu	Yes	Yes	No
GPE_TheHindu	Yes	Yes	Yes
LANGUAGE_TheHindu	Yes	Yes	No
LAW_TheHindu	Yes	Yes	No
LOC_TheHindu	Yes	Yes	No
NORP_TheHindu	Yes	Yes	No
ORDINAL_TheHindu	Yes	Yes	Yes
ORG_TheHindu	Yes	Yes	Yes
PERCENT_TheHindu	Yes	Yes	Yes
PERSON_TheHindu	Yes	Yes	Yes
QUANTITY_TheHindu	Yes	Yes	No
TIME_TheHindu	Yes	Yes	No
EVENT	Yes	Yes	No
GPE	Yes	Yes	No
LANGUAGE	Yes	Yes	No
LOC	Yes	Yes	Yes
PERCENT	Yes	Yes	No
PRODUCT	Yes	Yes	No
QUANTITY	Yes	Yes	No
TIME	Yes	Yes	No
WORK_OF_ART	Yes	Yes	No
POSITIVE_MoneyControl	Yes	Yes	Yes
NEGATIVE_MoneyControl	Yes	Yes	No
POSITIVE_TheHindu	Yes	Yes	Yes
NEGATIVE_TheHindu	Yes	Yes	No
POSITIVE	Yes	Yes	Yes
NEGATIVE	Yes	Yes	Yes
CARDINAL_MoneyControl	No	Yes	Yes
FAC_MoneyControl	No	Yes	Yes
MONEY_MoneyControl	No	Yes	No
ORDINAL_MoneyControl	No	Yes	Yes
ORG_MoneyControl	No	Yes	Yes
MONEY_TheHindu	No	Yes	No
PRODUCT_TheHindu	No	Yes	No

WORK_OF_ART_TheHindu	No	Yes	No
CARDINAL	No	Yes	No
DATE	No	Yes	No
FAC	No	Yes	No
LAW	No	Yes	No
MONEY	No	Yes	No
NORP	No	Yes	No
ORDINAL	No	Yes	Yes
ORG	No	Yes	No
PERSON	No	Yes	No

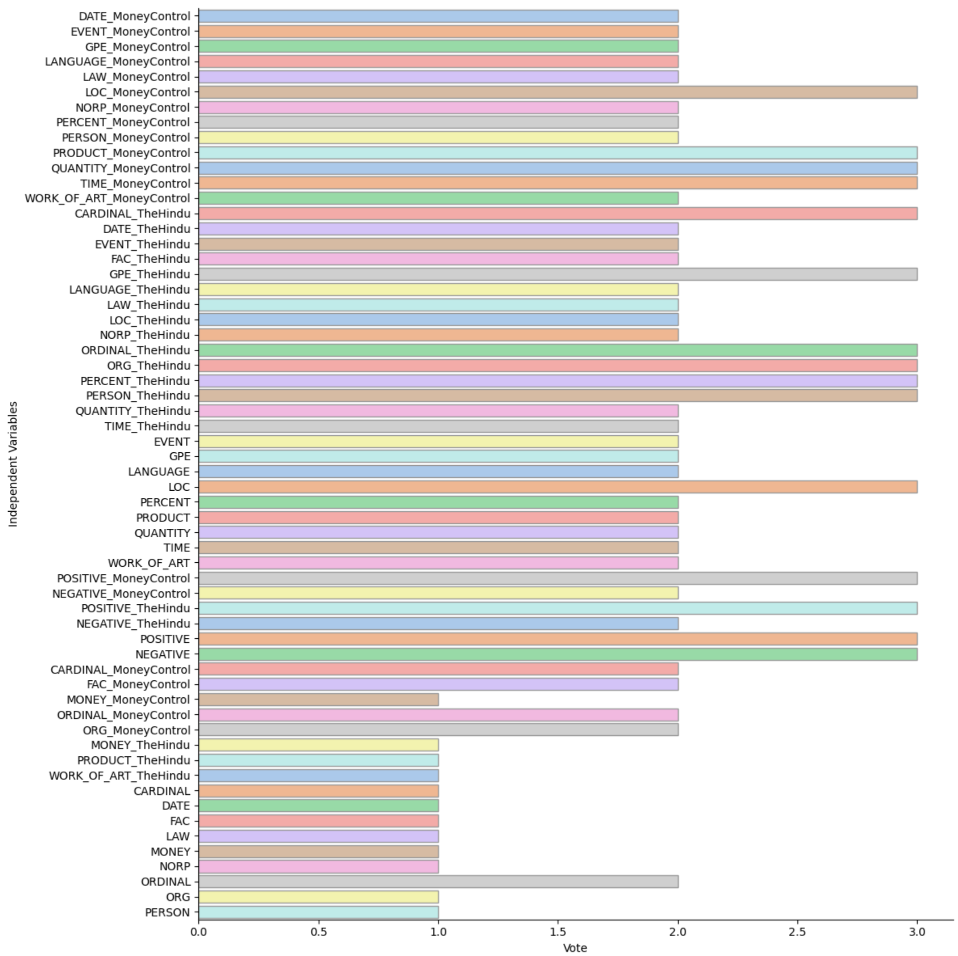

Figure 5.10 Most impacted economic articles variables vs. Vote count for doWhy.

Table 5.10 presented illustrates all the independent variables that have been found to have an impact on the Indian IT stock market, as determined by at least one of the refutation methods employed in the study. As can be seen from the table, all the named entities and sentiments that represented MoneyControl (suffix by _MoneyControl), The Hindu (suffixed by _TheHindu), and the Financial Times (not suffix by anything) have been identified to have an impact on the Indian IT stock market. However, there are 12 representations such as MONEY_MoneyControl, MONEY_TheHindu, PRODUCT_TheHindu, WORK_OF_ART_TheHindu, CARDINAL, DATE, FAC, LAW, MONEY, NORP, ORG, and PERSON impacted Indian IT stock market, as per only Pairwise Transferability Ratio (PTR) method of the DoWhy refutation library.

This study increased the confidence level of the results by considering text features and sentiments of economic articles that received at least two votes, which were DATE_MoneyControl, EVENT_MoneyControl, GPE_MoneyControl, LANGUAGE_MoneyControl, LAW_MoneyControl, LOC_MoneyControl, NORP_MoneyControl, PERCENT_MoneyControl, PERSON_MoneyControl, PRODUCT_MoneyControl, QUANTITY_MoneyControl, TIME_MoneyControl, WORK_OF_ART_MoneyControl, CARDINAL_TheHindu, DATE_TheHindu, EVENT_TheHindu, FAC_TheHindu, GPE_TheHindu, LANGUAGE_TheHindu, LAW_TheHindu, LOC_TheHindu, NORP_TheHindu, ORDINAL_TheHindu, ORG_TheHindu, PERCENT_TheHindu, PERSON_TheHindu, QUANTITY_TheHindu, TIME_TheHindu, EVENT, GPE, LANGUAGE, LOC, PERCENT, PRODUCT, QUANTITY, TIME, WORK_OF_ART, POSITIVE_MoneyControl, NEGATIVE_MoneyControl, POSITIVE_TheHindu, NEGATIVE_TheHindu, POSITIVE, NEGATIVE, CARDINAL_MoneyControl, FAC_MoneyControl, ORDINAL_MoneyControl, ORG_MoneyControl, and ORDINAL. Thus, as per the results of the DoWhy library, named entities and sentiments of economic articles from different economic sources have a significant impact on the Indian IT stock market.

5.5.2 Examine the impact of text features and sentiments of economic articles using a deep-learning-based method.

The research employs the nonLinCausality Library, which represents deep-learning-based techniques, to detect causality. The study achieved the same by implementing the Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) versions of the nonLinCausality Library, as these methods have been demonstrated to be highly effective in handling time-series data. The study analyzes the impact text features and sentiments of economic articles on the Indian IT stock market, utilizing the LSTM and GRU implementations of non-linear causality, and focuses on lag times of 10, 20, and 30 days. The results are presented in tables, which provide the representations of different economic articles and their impact on the Indian IT stock market, using LSTM or GRU implementation of nonLinCausality Library for the specified lag times.

Table 5.11 Most impacted economic article variables using nonLinCausality(GRU).

Independent Variables	GRU implementation of nonLinCausality		
	10 days	20 days	30 days
LANGUAGE_MoneyControl	Yes	Yes	Yes
LAW_MoneyControl	Yes	No	No
PERSON_MoneyControl	Yes	Yes	Yes
TIME_MoneyControl	Yes	Yes	Yes
FAC_TheHindu	Yes	No	No
LANGUAGE_TheHindu	Yes	Yes	Yes
LOC_TheHindu	Yes	Yes	No
PRODUCT_TheHindu	Yes	Yes	Yes
WORK_OF_ART_TheHindu	Yes	No	No
EVENT	Yes	Yes	Yes
FAC	Yes	Yes	Yes
LANGUAGE	Yes	No	Yes
LOC	Yes	Yes	Yes

MONEY	Yes	Yes	Yes
NORP	Yes	No	Yes
ORDINAL	Yes	No	Yes
PERCENT	Yes	Yes	Yes
PRODUCT	Yes	No	No
QUANTITY	Yes	Yes	No
LOC_MoneyControl	No	Yes	No
EVENT_TheHindu	No	Yes	Yes
LAW	No	Yes	Yes
ORG	No	Yes	Yes
MONEY_MoneyControl	No	No	Yes
LAW_TheHindu	No	No	Yes
CARDINAL	No	No	Yes
DATE	No	No	Yes
WORK_OF_ART	No	No	Yes

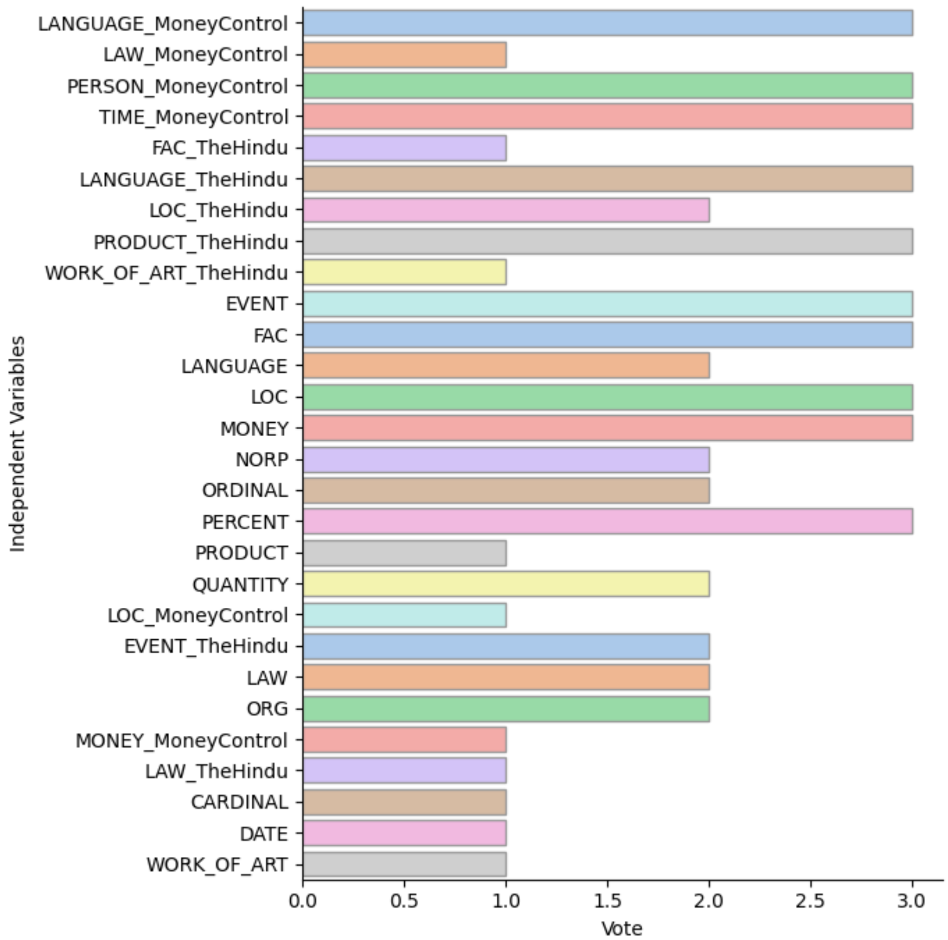

Figure 5.11 Most impacted economic articles variables vs. Vote count for nonLinCausality (GRU).

Table 5.12 Most impacted economic article variables using nonLinCausality (LSTM).

Independent Variables	LSTM implementation of nonLinCausality		
	10 days lag	20 days lag	30 days lag
DATE_MoneyControl	Yes	No	Yes
EVENT_MoneyControl	Yes	No	No
FAC_MoneyControl	Yes	No	Yes
LANGUAGE_MoneyControl	Yes	Yes	Yes
LAW_MoneyControl	Yes	Yes	Yes
ORG_MoneyControl	Yes	No	No
PERSON_MoneyControl	Yes	No	Yes
TIME_MoneyControl	Yes	Yes	Yes
CARDINAL_TheHindu	Yes	No	No
EVENT_TheHindu	Yes	Yes	Yes

FAC_TheHindu	Yes	No	No
LANGUAGE_TheHindu	Yes	Yes	Yes
LOC_TheHindu	Yes	Yes	Yes
NORP_TheHindu	Yes	No	No
PRODUCT_TheHindu	Yes	Yes	Yes
WORK_OF_ART_TheHindu	Yes	Yes	No
FAC	Yes	Yes	No
LANGUAGE	Yes	Yes	No
LAW	Yes	Yes	Yes
LOC	Yes	Yes	No
MONEY	Yes	Yes	Yes
ORDINAL	Yes	No	No
ORG	Yes	No	No
PERCENT	Yes	Yes	Yes
PRODUCT	Yes	Yes	Yes
QUANTITY	Yes	Yes	Yes
POSITIVE_MoneyControl	Yes	No	Yes
NEGATIVE_MoneyControl	Yes	No	No
PRODUCT_MoneyControl	No	Yes	Yes
QUANTITY_MoneyControl	No	Yes	No
ORDINAL_TheHindu	No	Yes	Yes
CARDINAL	No	Yes	Yes
EVENT	No	Yes	No
PERSON	No	Yes	No
GPE_MoneyControl	No	No	Yes
MONEY_MoneyControl	No	No	Yes
DATE	No	No	Yes
GPE	No	No	Yes

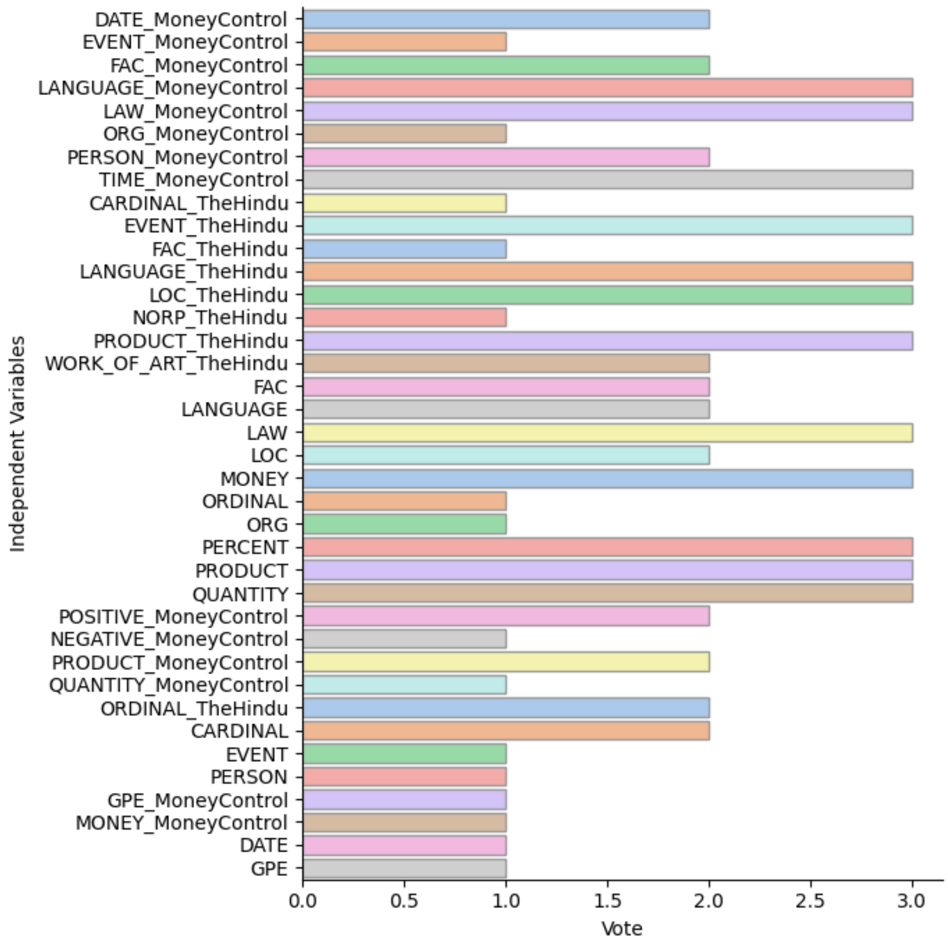

Figure 5.12 Most impacted economic articles variables vs. Vote count for nonLinCausality (LSTM).

As demonstrated by the data presented in Table 5.11 and 5.12, the economic articles from Money Control, The Hindu, and The Financial Times have significantly impacted the Indian IT stock market. Specifically, the results indicate that as per Gated Recurrent Unit (GRU) LANGUAGE_MoneyControl, PERSON_MoneyControl, TIME_MoneyControl, LANGUAGE_TheHindu, LOC_TheHindu, PRODUCT_TheHindu, EVENT, FAC, LANGUAGE, LOC, MONEY, NORP, ORDINAL, PERCENT, QUANTITY, EVENT_TheHindu, LAW, and ORG voted by at least two-time lags and impacted Indian IT stock market. However, as per Long Short-term Memory (LSTM)

DATE_MoneyControl, FAC_MoneyControl, LANGUAGE_MoneyControl, LAW_MoneyControl, PERSON_MoneyControl, TIME_MoneyControl, EVENT_TheHindu, LANGUAGE_TheHindu, LOC_TheHindu, PRODUCT_TheHindu, WORK_OF_ART_TheHindu, FAC, LANGUAGE, LAW, LOC, MONEY, PERCENT, PRODUCT, QUANTITY, POSITIVE_MoneyControl, PRODUCT_MoneyControl, ORDINAL_TheHindu, and CARDINAL voted by at least two-time lags and has impacted the Indian IT stock market. It is worth noting that this research has identified only Money Control sentiments (i.e., positive, and negative sentiments) that have impacted Indian IT, that too only with the LSTM implementation of nonLinCausality.

Thus, the above findings strongly suggest that the named entities of Money Control, The Hindu, The Financial Times, and Sentiment of Money control have had a causative effect on the Indian IT stock market.

5.5.3 Examine the impact of text features and sentiments of economic articles on the Indian IT stock market using a statistical-based method.

This research has used the Granger-Causality detection test to determine the causality relationship between the economic articles from different sources and the Indian IT stock market. The Granger-Causality test is the most widely used statistical method for verifying the usefulness of one variable in forecasting another variable. Thus, this research has analyzed the impact of major economic articles on the Indian IT stock market using lag times of 10 days, 20 days, and 30 days. The following table provides the result of the Granger Causality test with a lag of 10 days, 20 days, and 30 days.

Table 5.13 Most impacted economic article variables using Granger Causality.

Independent Variable	Granger Causality Test		
	10 days lag	20 days lag	30 days lag
FAC_MoneyControl	Yes	Yes	Yes
GPE_MoneyControl	Yes	Yes	Yes
PRODUCT_MoneyControl	Yes	Yes	Yes
WORK_OF_ART_MoneyControl	Yes	Yes	Yes
CARDINAL_TheHindu	Yes	Yes	Yes
LANGUAGE_TheHindu	Yes	Yes	Yes
LOC_TheHindu	Yes	Yes	Yes
LANGUAGE	Yes	Yes	Yes
NORP	Yes	Yes	Yes
PERCENT	Yes	Yes	Yes
PRODUCT	Yes	Yes	Yes
POSITIVE	Yes	Yes	Yes
LANGUAGE_MoneyControl	No	Yes	Yes
QUANTITY_MoneyControl	No	Yes	Yes
GPE_TheHindu	No	Yes	Yes
LAW_TheHindu	No	Yes	Yes
EVENT_TheHindu	No	No	Yes

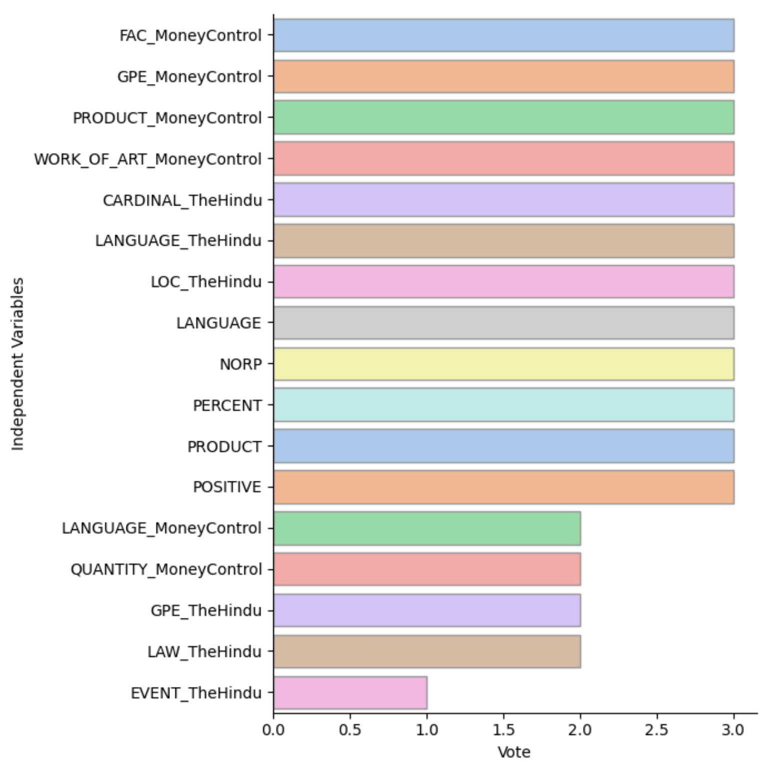

Figure 5.13 Most impacted economic articles variables vs. Vote count for Granger Causality.

As demonstrated by the table above, FAC_MoneyControl, GPE_MoneyControl, PRODUCT_MoneyControl, WORK_OF_ART_MoneyControl, CARDINAL_TheHindu, LANGUAGE_TheHindu, LOC_TheHindu, LANGUAGE, NORP, PERCENT, PRODUCT, POSITIVE, LANGUAGE_MoneyControl, QUANTITY_MoneyControl, GPE_TheHindu, and LAW_TheHindu are the named entities voted by Granger Causality Test at least with two-time lags and have an impact on the Indian IT stock market. However, as per the Granger causality test, the sentiments of economic articles published on Money Control and The Hindu did not impact the Indian IT stock market. Therefore, the above results indicated the representation of Money Control, The Hindu, and the Financial Times that has impacted the Indian IT stock market.

5.5.4 Examine the impact of text features and sentiments of economic articles on the Indian IT stock market using ensemble methods.

As discussed in the preceding sections, this research employed a variety of machine-learning, deep-learning, and statistical-based causality detection methods to analyze the impact of economic articles from Money Control, The Hindu, and the Financial Times on the Indian IT stock market. Each of these methods provided insights into which named entities and sentiments of economic articles impact the Indian IT stock market. According to the DoWhy library (a machine-learning-based method), the DATE_MoneyControl, EVENT_MoneyControl, GPE_MoneyControl, LANGUAGE_MoneyControl, LAW_MoneyControl, LOC_MoneyControl, NORP_MoneyControl, PERCENT_MoneyControl, PERSON_MoneyControl, PRODUCT_MoneyControl, QUANTITY_MoneyControl, TIME_MoneyControl, WORK_OF_ART_MoneyControl, CARDINAL_TheHindu, DATE_TheHindu, EVENT_TheHindu, FAC_TheHindu, GPE_TheHindu, LANGUAGE_TheHindu, LAW_TheHindu, LOC_TheHindu, NORP_TheHindu, ORDINAL_TheHindu, ORG_TheHindu, PERCENT_TheHindu, PERSON_TheHindu, QUANTITY_TheHindu, TIME_TheHindu, EVENT, GPE, LANGUAGE, LOC, PERCENT, PRODUCT, QUANTITY, TIME, WORK_OF_ART, POSITIVE_MoneyControl, NEGATIVE_MoneyControl, POSITIVE_TheHindu, NEGATIVE_TheHindu, POSITIVE, NEGATIVE, CARDINAL_MoneyControl, FAC_MoneyControl, ORDINAL_MoneyControl, ORG_MoneyControl, and ORDINAL have had an impact on the Indian IT stock market. Meanwhile, the GRU implementation of nonLinCausality methods (a deep-learning-based method) determined that the LANGUAGE_MoneyControl, PERSON_MoneyControl, TIME_MoneyControl, LANGUAGE_TheHindu, LOC_TheHindu, PRODUCT_TheHindu, EVENT, FAC, LANGUAGE, LOC, MONEY, NORP, ORDINAL, PERCENT, QUANTITY, EVENT_TheHindu, LAW, and ORG had an impact on the Indian IT stock market. Additionally, as per the LSTM implementation of nonLinCausality methods, DATE_MoneyControl, FAC_MoneyControl, LANGUAGE_MoneyControl, LAW_MoneyControl, PERSON_MoneyControl, TIME_MoneyControl, EVENT_TheHindu, LANGUAGE_TheHindu, LOC_TheHindu, PRODUCT_TheHindu, WORK_OF_ART_TheHindu, FAC, LANGUAGE, LAW, LOC, MONEY, PERCENT, PRODUCT, QUANTITY, POSITIVE_MoneyControl, PRODUCT_MoneyControl, ORDINAL_TheHindu, and CARDINAL had an impact on the Indian IT stock market.

Similarly, FAC_MoneyControl, GPE_MoneyControl, PRODUCT_MoneyControl, WORK_OF_ART_MoneyControl, CARDINAL_TheHindu, LANGUAGE_TheHindu, LOC_TheHindu, LANGUAGE, NORP, PERCENT, PRODUCT, POSITIVE, LANGUAGE_MoneyControl, QUANTITY_MoneyControl, GPE_TheHindu, and LAW_TheHindu have a causal impact on Indian IT stock market as per the statistical based (Granger Test) method.

This study considered the findings of all three approaches to obtain a more accurate result. The following table presents a consolidated result of all the causality detection methods used in this research.

Table 5.14 Most impacted economic article variables using ensemble methods.

Independent Variables	Impact on the Indian IT stocks		
	Statistical-based (Granger Test)	Machine-learning-based (DoWhy)	Deep-learning-based (nonLinCausality)
FAC_MoneyControl	Yes	Yes	Yes
GPE_MoneyControl	Yes	Yes	Yes
PRODUCT_MoneyControl	Yes	Yes	Yes
WORK_OF_ART_MoneyControl	Yes	Yes	No
CARDINAL_TheHindu	Yes	Yes	Yes
LANGUAGE_TheHindu	Yes	Yes	Yes
LOC_TheHindu	Yes	Yes	Yes
LANGUAGE	Yes	Yes	Yes
NORP	Yes	Yes	Yes
PERCENT	Yes	Yes	Yes
PRODUCT	Yes	Yes	Yes
POSITIVE	Yes	Yes	No
LANGUAGE_MoneyControl	Yes	Yes	Yes
QUANTITY_MoneyControl	Yes	Yes	Yes
GPE_TheHindu	Yes	Yes	No
LAW_TheHindu	Yes	Yes	Yes
EVENT_TheHindu	Yes	Yes	Yes

DATE_MoneyControl	No	Yes	Yes
EVENT_MoneyControl	No	Yes	Yes
LAW_MoneyControl	No	Yes	Yes
ORG_MoneyControl	No	Yes	Yes
PERSON_MoneyControl	No	Yes	Yes
TIME_MoneyControl	No	Yes	Yes
FAC_TheHindu	No	Yes	Yes
NORP_TheHindu	No	Yes	Yes
PRODUCT_TheHindu	No	Yes	Yes
WORK_OF_ART_TheHindu	No	Yes	Yes
FAC	No	Yes	Yes
LAW	No	Yes	Yes
LOC	No	Yes	Yes
MONEY	No	Yes	Yes
ORDINAL	No	Yes	Yes
ORG	No	Yes	Yes
QUANTITY	No	Yes	Yes
POSITIVE_MoneyControl	No	Yes	Yes
NEGATIVE_MoneyControl	No	Yes	Yes
ORDINAL_TheHindu	No	Yes	Yes
CARDINAL	No	Yes	Yes
EVENT	No	Yes	Yes
PERSON	No	Yes	Yes
MONEY_MoneyControl	No	Yes	Yes
DATE	No	Yes	Yes
GPE	No	Yes	Yes
LOC_MoneyControl	No	Yes	Yes
WORK_OF_ART	No	Yes	Yes
NORP_MoneyControl	No	Yes	No
PERCENT_MoneyControl	No	Yes	No
DATE_TheHindu	No	Yes	No
ORG_TheHindu	No	Yes	No
PERCENT_TheHindu	No	Yes	No
PERSON_TheHindu	No	Yes	No
QUANTITY_TheHindu	No	Yes	No
TIME_TheHindu	No	Yes	No
TIME	No	Yes	No
POSITIVE_TheHindu	No	Yes	No
NEGATIVE_TheHindu	No	Yes	No
NEGATIVE	No	Yes	No

CARDINAL_MoneyControl	No	Yes	No
ORDINAL_MoneyControl	No	Yes	No
MONEY_TheHindu	No	Yes	No

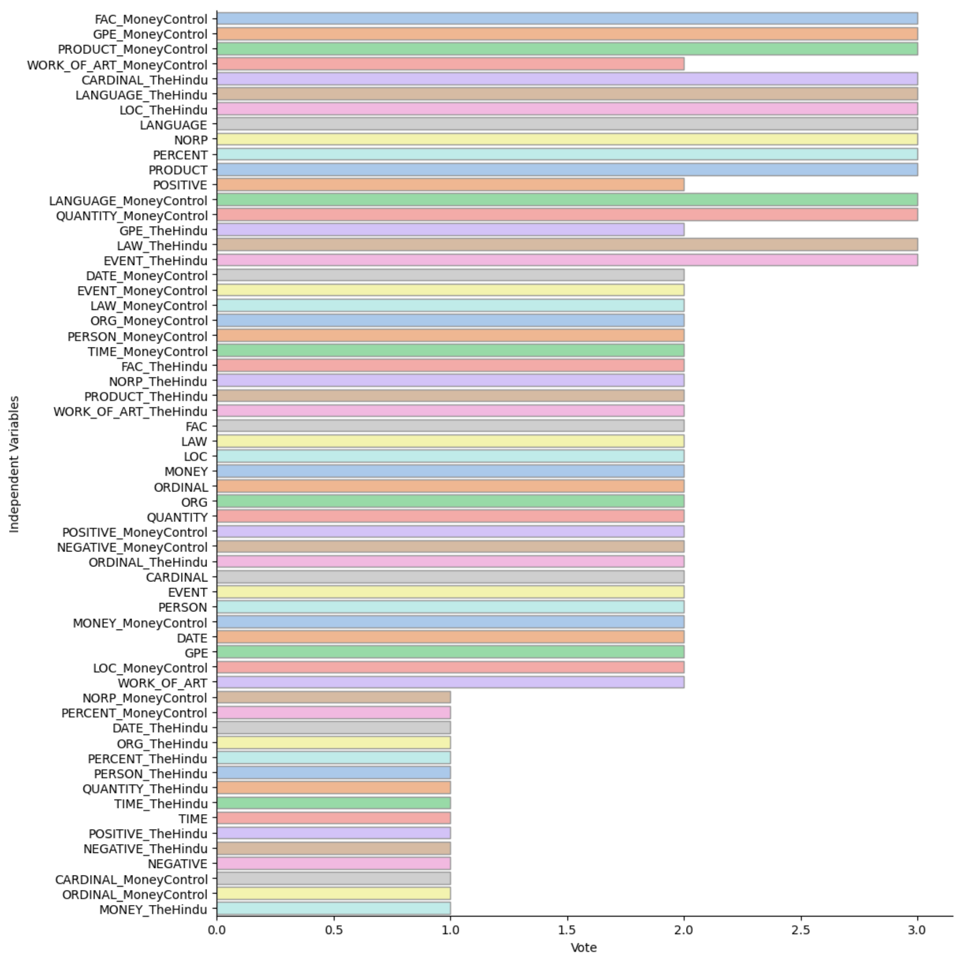

Figure 5.14 Most impacted economic articles variables vs. Vote count for ensemble methods.

The results of the statistical, machine learning and deep-learning methods employed in this research provide strong evidence that entities such as FAC_MoneyControl, GPE_MoneyControl, PRODUCT_MoneyControl, WORK_OF_ART_MoneyControl, CARDINAL_TheHindu, LANGUAGE_TheHindu, LOC_TheHindu, LANGUAGE, NORP, PERCENT, PRODUCT, POSITIVE, LANGUAGE_MoneyControl, QUANTITY_MoneyControl, GPE_TheHindu, LAW_TheHindu, EVENT_TheHindu, DATE_MoneyControl, EVENT_MoneyControl, LAW_MoneyControl, ORG_MoneyControl, PERSON_MoneyControl, TIME_MoneyControl, FAC_TheHindu, NORP_TheHindu, PRODUCT_The Hindu, WORK_OF_ART_TheHindu, FAC, LAW, LOC, MONEY, ORDINAL, ORG, QUANTITY, POSITIVE_MoneyControl, NEGATIVE_MoneyControl, ORDINAL_TheHindu, CARDINAL, EVENT, PERSON, MONEY_MoneyControl, DATE, GPE, LOC_MoneyControl, and WORK_OF_ART have a significant impact on the Indian IT stock market. As depicted in Figure 4.3, a clear inverse relationship is observed between the Indian IT stock market and the named entities and sentiments identified by the machine-learning, deep-learning, and statistical methods, except positive sentiment in articles from The Hindu and The Financial Times, which means that an increase in the positive sentiments in economic articles from The Hindu and The Financial Times is likely to increase the Nifty IT price. In contrast, an increase in other entities apart from positive sentiments from The Hindu and Financial Times will lead to a decrease in the Indian IT stock market.

This finding addresses the research question of *"How do currency exchange rates affect the Indian IT stock market"* and supports the hypothesis that an increase in the positive sentiments of the Hindu and the Financial Times articles leads to an increase in the Indian IT stock market. In contrast, an increase in all other representations of economic articles except the positive sentiments of the Hindu and Financial Times articles leads to a decrease in the Indian IT stock market.

In conclusion, the findings of this study indicate that the French, American, British, and German IT stock markets positively impact the Indian IT stock market. Precisely, an increase in the selected foreign IT stock markets corresponds with an increase in the Indian IT stock market. Additionally, an increase in currency exchange rates such as USD_INR, JPY_INR, and SGD_INR also increases in the Indian IT stock market. In contrast, an increase in the MUR-INR currency exchange rate leads to a decrease in the Indian IT stock market. Furthermore, an increase in positive sentiments, as reflected in articles from The Hindu and The Financial Times, corresponds with an increase in the Indian IT stock market.

In contrast, an increase in other representations of economic articles apart from positive sentiments in The Hindu and Financial Times corresponds with a decrease in the Indian IT stock market. Therefore, based on the results of this research, it can be inferred that foreign IT stock markets, currency exchange rates, and economic articles from different sources do indeed impact the Indian IT stock market. It is worth noting that this research also has certain limitations, which are discussed in the following section of this chapter.

5.6 Limitation and Future work

This study employed machine-learning, deep-learning, and statistical-based techniques to investigate the causal relationship between the foreign IT stock market, currency exchange rate, and economic articles from different sources and the Indian IT stock market. The results obtained from these methods indicate a significant impact on the

Indian stock market by the above factors. However, it is essential to acknowledge that this research has its limitations, which are discussed below.

This study utilized machine-learning-based, deep-learning-based, and statistical-based causality detection methods to investigate the impact of foreign IT stock markets, currency exchange rates, and economic articles on the Indian IT stock market.

Additionally, bivariate analysis was employed to gain insight into the relationship between these variables and the Indian IT stock market, providing a visual representation of their association. However, this research should have included statistical methods to quantify the strength of the relationship between these factors. Therefore, a potential avenue for future research could involve using statistical methods to quantify the strength of the relationship between the foreign IT stock market, currency exchange rate, economic articles, and the Indian IT stock market.

This research focuses on a specific sector of the Indian stock market, namely the Information Technology sector. As such, the results of this study are limited to the Information Technology segment of the Indian stock market. However, the causality detection framework introduced in this research can be easily replicated in other stock market segments. Therefore, a potential avenue for future research could involve applying the same causality detection framework to detect the impactful features of other stock market segments.

A limitation of this study is the limited number of words per article analyzed. By relying solely on article headlines, the study ignored the rest of the information in each article and risked an unknown impact on the results generated by the stock market. Hence, future research could include a complete analysis of the economic news articles, which would provide a more comprehensive context and increase the reliability of the results.

This research explores the impact of foreign IT stock market movements, currency exchange rates, and economic articles on the Indian IT stock market. It is determined that changes in any of these variables can provide indications of shifts in the Indian IT stock market. However, a thorough examination of the individual impact of each variable and their combined effect on the Indian IT stock market would require a significant amount of analysis.

While this may be manageable for portfolio managers, it may need to be more convenient for individual investors. To alleviate this issue, a future avenue of research could involve the development of a machine-learning model based on the findings of this thesis, which would streamline the analysis process and make it more accessible for individual investors.

5.7 Conclusion

The present study examines the impact of foreign IT stock markets, currency exchange rates, and economic articles on the Indian IT stock market. Unlike prior research, which primarily relied on the Granger Causality test and considered the Indian stock market, this study employs advanced techniques such as deep-learning-based, machine-learning-based, and statistical-based methods to analyze non-linear datasets.

The results of this multi-model approach reveal that foreign IT stock markets such as France, America, Britain, and Germany positively impact the Indian IT stock market. Additionally, an increase in currency exchange rates such as USD_INR, JPY_INR, and

SGD_INR leads to an increase in the Indian IT stock market. In contrast, an increase in the MUR-INR currency exchange rate corresponds with a decrease in the Indian IT stock market. Furthermore, positive sentiments reflected in articles from The Hindu and The Financial Times correspond with an increase in the Indian IT stock market, whereas other representations of economic articles apart from positive sentiments in The Hindu and Financial Times correspond with a decrease in the Indian IT stock market.

The implication of this research study can provide valuable insights for portfolio managers to devise profitable portfolios and assist investors in making intelligent investment decisions in the Indian IT stock market.

Secondly, the research study will help improve academic performance, financial knowledge, and saving and investment habits of people and businesses who work in the financial markets.

Furthermore, the study implications serve as a synoptic analysis for business performance and an indication that comparative analysis in the global indices can improve the international financial stock market.

Finally, the overall study contributes to the existing literature by providing a more robust and confident analysis of the impact of foreign and domestic factors on the Indian IT stock market.

APPENDIX A

DATASET AND CODE USED FOR STUDY

Following are all the artifcats used in order to conduct this study, alongwith the code and analysis sheet.

Foreign stock data

global_indices-20230
221T160356Z-001.zip

Exchange rates

exchange_rates-2023
0221T160400Z-001.z

Economic articles

news_articles-20230
221T160357Z-001.zip

Code used and analysis done for study:

Causal_Inference_DBA.ipynb

Final%20Analysis.xlsx

REFERENCES

Abdullahi, S. (2020). Measuring Co-Movements and Linkages between Nigeria and the UAE Stock Exchanges: Is there Opportunity for Portfolio Building? *Journal of Advanced Research in Economics and Administrative Sciences*, *1*(2), 106–122. https://doi.org/10.47631/jareas.v1i2.124

AbdulRahman Bala, S., & Hassan, A. (2018). Arabian Journal of Business and Management Review. *Arabian Journal of Business and Ar A Management Review*. https://www.researchgate.net/publication/363121226

Agarap, A. F. (2018). *Deep Learning using Rectified Linear Units (ReLU)*. http://arxiv.org/abs/1803.08375

Akbik, A., Schweter, S., Blythe, D., & Vollgraf, R. (2019). *F LAIR : An Easy-to-Use Framework for State-of-the-Art NLP*. 54–59.

Albawi, S., Abed MOHAMMED, T., & Al-zawi, S. (2017). Understanding of a convolutional neural network; Understanding of a convolutional neural network. *2017 International Conference on Engineering and Technology (ICET)*. https://doi.org/10.1109/ICEngTechnol.2017.8308186

Ali, K., & Al-Hameed, A. (2022). Spearman's correlation coefficient in statistical analysis. *Int. J. Nonlinear Anal. Appl*, *13*, 2008–6822. https://doi.org/10.22075/ijnaa.2022.6079

Anuradha, A. (2018). Impact of Exchange Rate Fluctuation on Stock Market Volatility - a Study To Predict the Economic Scenario in India. *International Journal of Pure and Applied Mathematics*, *118*(July), 4309–4316.

Bahdanau, D., Cho, K. H., & Bengio, Y. (2015). Neural machine translation by jointly learning to align and translate. *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, 1–15.

Bertani, A., di Paola, G., Russo, E., & Tuzzolino, F. (2018). How to describe bivariate data. *Journal of Thoracic Disease, 10*(2), 1133–1137. https://doi.org/10.21037/jtd.2018.01.134

Bertsimas, D., Pawlowski, C., & Zhuo, Y. D. (2018). From predictive methods to missing data imputation: An optimization approach. *Journal of Machine Learning Research*, *18*, 1–39.

Bhandari, P. (2021). *A step-by-step guide to data collection*. Scribber. https://www.scribbr.com/methodology/data-collection/

Bhattacharjee, A., & Das, J. (2020). The Relationship between USD-INR Exchange Rate and Indian Stock Market. *An International Peer Reviewed Research Journal*, *8*(January 2021), 1–6.

BiObserver. (2014). *Visualization of en:Granger causality. Regenerated from Liu and Bahadori, "A Survey on Granger Causality, A Computational View."* Own Work, CC BY-SA 3.0. https://commons.wikimedia.org/w/index.php?curid=33470670

Biswal, A. (2022). *Convolutional Neural Network Tutorial*. Simplilearn. https://www.simplilearn.com/tutorials/deep-learning-tutorial/convolutional-neural-network#what_is_convolutional_neural_network

Biswas, S., Ghosh, A., Chakraborty, S., Roy, S., & Bose, R. (2020). Scope of sentiment analysis on news articles regarding stock market and GDP in struggling economic condition. *International Journal of Emerging Trends in Engineering Research*, *8*(7), 3594–3609. https://doi.org/10.30534/ijeter/2020/117872020

Blackstone, A. (2018). Principles of sociological inquiry: Qualitative and quantitative methods. In *Saylor Academy Open Textbooks.*

Bouikidis, C. (2018). Focusing on the Fundamentals: A Simplistic Differentiation Between Qualitative and Quantitative Research. In *Article in Nephrology nursing* *journal: journal of the American Nephrology Nurses' Association*. https://www.researchgate.net/publication/328250766

BusinessToday. (2018). How the fall in rupee will affect IT companies. *BusinessToday*. https://www.businesstoday.in/current/corporate/how-the-fall-in-rupee-will-affect-it-companies/story/282244.html

Casteel, A., & Bridier, N. L. (2021). Describing Populations and Samples in Doctoral. *International Journal of Doctoral Studies*, *16*, 339–362.

Chen, B., Deng, W., & Du, J. (2017). Noisy softmax: Improving the generalization ability of DCNN via postponing the early softmax saturation. *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, *2017-January*, 4021–4030. https://doi.org/10.1109/CVPR.2017.428

Chen, H., Harinen, T., Lee, J.-Y., Yung, M., & Zhao, Z. (2020). *CausalML: Python Package for Causal Machine Learning*. http://arxiv.org/abs/2002.11631

CHEN, J. (2021). *Stock Market*. https://www.investopedia.com/terms/s/stockmarket.asp

Chivukula, A. S., Li, J., & Liu, W. (2018). Discovering granger-causal features from deep learning networks. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *11320 LNAI*, 692–705. https://doi.org/10.1007/978-3-030-03991-2_62

Cho, K., van Merriënboer, B., Bahdanau, D., & Bengio, Y. (2014). On the properties of neural machine translation: Encoder–decoder approaches. *Proceedings of SSST 2014 - 8th Workshop on Syntax, Semantics and Structure in Statistical Translation*, 103–111. https://doi.org/10.3115/v1/w14-4012

Creswell, J. W., & Creswell, J. D. (2018). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches* (Fifth edition). Sage.

Daniel, E. (2016). The Usefulness of Qualitative and Quantitative Approaches and Methods in Researching Problem-Solving Ability in Science Education Curriculum. *Journal of Education and Practice*, *7*(15). www.iiste.org

Datta, L. (2020). *A Survey on Activation Functions and their relation with Xavier and He Normal Initialization*. http://arxiv.org/abs/2004.06632

Deo, M., & Arun Prakash, P. (2017). a Study on Integration of Stock Markets: Empirical Evidence From National Stock Exchange and Major Global Stock Markets. *ICTACT Journal on Management Studies*, *03*(02), 479–485. https://doi.org/10.21917/ijms.2017.0065

Dickey, D. A., & Fuller, W. A. (1979). Distribution of the Estimators for Autoregressive Time Series With a Unit Root. *Journal of the American Statistical Association*, *74*(366), 427. https://doi.org/10.2307/2286348

Dix, P. (2021). *Why Time Series Matters For Metrics, Real-Time Analytics and Sensor Data*.

Favorskaya, M. N., & Andreev, V. v. (2019). The study of activation functions in deep learning for pedestrian detection and tracking. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, *42*(2/W12), 53–59. https://doi.org/10.5194/isprs-archives-XLII-2-W12-53-2019

Frank Wilcoxon. (1945). Individual Comparisons by Ranking Methods. *Biometrics Bulletin*, *1*(6), 80–83. https://doi.org/10.2307/3001968

Gangadharan, S. R., & Yoonus, C. A. (2012). Global Financial Crisis and Stock Market Integration: A Study on the Impact of Global Financial Crisis on the Level of Financial Integration between the US and Indian Stock Markets. *Asia-Pacific*

Journal of Management Research and Innovation, *8*(2), 101–110. https://doi.org/10.1177/2319510X1200800203

Gao, S. E., Lin, B. S., & Wang, C. M. (2019). Share price trend prediction using CRNN with LSTM structure. *Proceedings - 2018 International Symposium on Computer, Consumer and Control, IS3C 2018*, 10–13. https://doi.org/10.1109/IS3C.2018.00012

Gholamalinezhad, H., & Khosravi, H. (n.d.). *Pooling Methods in Deep Neural Networks, a Review*.

Goel, K., & Gupta, R. (2011). Impact of Globalization on Stock Market: Development in India. *Delhi Business Review*, *12*(1), 69–84. https://doi.org/10.51768/dbr.v12i1.121201106

Granger, C. W. J. (1969). Investigating Causal Relations by Econometric Models and Cross-spectral Methods. *Econometrica*, *37*(3), 424. https://doi.org/10.2307/1912791

Gulzar, S., Mujtaba Kayani, G., Xiaofeng, H., Ayub, U., & Rafique, A. (2019). Financial cointegration and spillover effect of global financial crisis: a study of emerging Asian financial markets. *Economic Research-Ekonomska Istrazivanja* , *32*(1), 187–218. https://doi.org/10.1080/1331677X.2018.1550001

Gunzler, D., Chen, T., Wu, P., & Zhang, H. (2013). Introduction to mediation analysis with structural equation modeling. *Shanghai Archives of Psychiatry*, *25*(6), 390–394. https://doi.org/10.3969/j.issn.1002-0829.2013.06.009

Haukoos, J. S., & Lewis, R. J. (2015). The propensity score. In *JAMA - Journal of the American Medical Association* (Vol. 314, Issue 15, pp. 1637–1638). American Medical Association. https://doi.org/10.1001/jama.2015.13480

Heale, R., & Twycross, A. (2015). Validity and reliability in quantitative studies. *Evidence-Based Nursing*, *18*(3), 66–67. https://doi.org/10.1136/eb-2015-102129

Hicham, E. B., & Salah-Ddine, K. (2021). Stock market forecasting model from multi news data source using a two-level learning algorithm. *Turkish Journal of Computer and Mathematics Education*, *12*(5), 1022–1030. https://doi.org/10.17762/turcomat.v12i5.1746

Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*. https://doi.org/10.1162/neco.1997.9.8.1735

Honnibal, M. (2015). *spaCy*. https://spacy.io/usage/linguistic-features

Hu, Z., Zhang, J., & Ge, Y. (2021). Handling Vanishing Gradient Problem Using Artificial Derivative. *IEEE Access*, *9*, 22371–22377. https://doi.org/10.1109/ACCESS.2021.3054915

Huang, L., Qin, J., Zhou, Y., Zhu, F., Liu, L., & Shao, L. (2020). *Normalization Techniques in Training DNNs: Methodology, Analysis and Application*. http://arxiv.org/abs/2009.12836

Huang, Y., Fu, Z., & Franzke, C. L. E. (2020). Detecting causality from time series in a machine learning framework. *Chaos*, *30*(6). https://doi.org/10.1063/5.0007670

Hwang, E., & Kim, Y. H. (2019). Interdependency between the Stock Market and Financial News. *Proceedings - 2019 IEEE International Conference on Big Data, Big Data 2019*, 6074–6076. https://doi.org/10.1109/BigData47090.2019.9006533

Imbens, G. W., & Rubin, D. B. (2015). *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*. Cambridge University Press. https://doi.org/DOI: 10.1017/CBO9781139025751

INDIA: FOREIGN INVESTMENT. (n.d.). https://santandertrade.com/en/portal/establish-overseas/india/foreign-investment

Isaac, E., & Chikweru, E. (2018). *Test for Significance of Pearson's Correlation Coefficient ( )*. www.seahipaj.org

Jamal, P., Ali, M., Faraj, R. H., Ali, P. J. M., & Faraj, R. H. (2014). Data Normalization and Standardization: A Technical Report. In *Machine Learning Technical Reports* (Vol. 1, Issue 1). https://docs.google.com/document/d/1x0A1nUz1WWtMCZb5oVzF0SVMY7a_58KQulqQVT8LaVA/edit#

Jerez, J. M., Molina, I., García-Laencina, P. J., Alba, E., Ribelles, N., Martín, M., & Franco, L. (2010). Missing data imputation using statistical and machine learning methods in a real breast cancer problem. *Artificial Intelligence in Medicine*, *50*(2), 105–115. https://doi.org/10.1016/j.artmed.2010.05.002

Johansen. (1991). Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models. *Econometrica*, *59*(6), 1551–1580. https://doi.org/https://doi.org/10.2307/2938278

Joshi, K., H. N, B., & Rao, J. (2016). Stock Trend Prediction Using News Sentiment Analysis. *International Journal of Computer Science and Information Technology*. https://doi.org/10.5121/ijcsit.2016.8306

Joshi, M. C. (2013). Factors Affecting Indian Stock Market. *SSRN Electronic Journal*, *November*. https://doi.org/10.2139/ssrn.2238539

Khan, W., Ghazanfar, M. A., Azam, M. A., Karami, A., Alyoubi, K. H., & Alfakeeh, A. S. (2022). Stock market prediction using machine learning classifiers and social media, news. *Journal of Ambient Intelligence and Humanized Computing*, *13*(7), 3433–3456. https://doi.org/10.1007/s12652-020-01839-w

Kirongo A. Chege, & Odoyo C. Otieno. (2020). Research Philosophy Design and Methodologies: A Systematic Review of Research Paradigms in Information Technology. *Global Scientific Journals*, *8*(5). www.globalscientificjournal.com

Kishor, N., & Singh, R. P. (2017). Contagion Effect among the BRICS Stock Market Indices. *Journal of Poverty, Investment and Development*, *31*(ISSN 2422-846X), 32–46. https://www.researchgate.net/publication/317209759_Contagion_Effect_among_the_BRICS_Stock_Market_Indices

Kluyver, T., Ragan-Kelley, B., Pérez, F., Granger, B., Bussonnier, M., Frederic, J., Kelley, K., Hamrick, J., Grout, J., Corlay, S., Ivanov, P., Avila, D., Abdalla, S., & Willing, C. (2016). Jupyter Notebooks—a publishing format for reproducible computational workflows. *Positioning and Power in Academic Publishing: Players, Agents and Agendas - Proceedings of the 20th International Conference on Electronic Publishing, ELPUB 2016*, 87–90. https://doi.org/10.3233/978-1-61499-649-1-87

Kotsiantis, S. B., & Kanellopoulos, D. (2006). Data preprocessing for supervised leaning. *International Journal of …*, *1*(2), 1–7. https://doi.org/10.1080/02331931003692557

Kucharavy, D., Damand, D., Gamoura, S., & Barth, M. (2020). *Supporting Strategic Decision-Making in Manufacturing 4.0 with mix of qualitative and quantitative data analysis SUPPORTING STRATEGIC DECISION-MAKING IN MANUFACTURING 4.0 WITH MIX OF QUALITATIVE AND QUANTITATIVE DATA ANALYTICS*. https://hal.science/hal-03017736

Kumar, A. (2019). Understanding the Interdependency between the Nifty 50 Future Index and the Advanced Future Stock Market through Econometrics. *International Journal of Recent Technology and Engineering*, *8*(4), 3660–3664. https://doi.org/10.35940/ijrte.d7880.118419

Kumar, B. S., Ravi, V., & Miglani, R. (2020). Predicting Indian Stock Market Using the Psycho-Linguistic Features of Financial News. *Annals of Data Science*, 1–45. https://doi.org/10.1007/s40745-020-00272-2

Kumar Sharma, R., Geneva, S., & Annan, D. (2022). "IMPACT OF FOREIGN MARKET AND CURRENCY EXCHANGE RATE ON INDIAN IT STOCKS." *Global Journal of Business and Integral Security*.

Kwatra, N. (2018). How-news-impacts-stock-market-movements. *Livemint*. https://www.livemint.com/Money/9QZwywOBgxrtZ4xydfpofJ/How-news-impacts-stock-market-movements.html

Li, X., Xie, H., Wang, R., Cai, Y., Cao, J., Wang, F., Min, H., & Deng, X. (2016). Empirical analysis: stock market prediction via extreme learning machine. *Neural Computing and Applications*. https://doi.org/10.1007/s00521-014-1550-z

L.Lima, M., P. Nascimento, T., Labidi, S., S. Timbo, N., v. L. Batista, M., N. Neto, G., A. M. Costa, E., & R. S. Sousa, S. (2016). Using Sentiment Analysis for Stock Exchange Prediction. *International Journal of Artificial Intelligence & Applications*, *7*(1), 59–67. https://doi.org/10.5121/ijaia.2016.7106

Manu, K. S., & Bhaskar, P. (2018). Effect of Exchange Rates Volatility on Stock Market Performance. *Asian Journal of Management*, *9*(4), 1337. https://doi.org/10.5958/2321-5763.2018.00212.3

Manu, K. S., & Menda, V. L. (2017). Dynamics of indian stock market integration with global stock markets. *Asian Journal of Management*, *8*(3), 559. https://doi.org/10.5958/2321-5763.2017.00090.7

Marcinkevičs, R., & Vogt, J. E. (2021). *Interpretable Models for Granger Causality Using Self-explaining Neural Networks*. http://arxiv.org/abs/2101.07600

Menzies, T., Kocagüneli, E., Minku, L., Peters, F., & Turhan, B. (2015). Using Goals in Model-Based Reasoning. *Sharing Data and Models in Software Engineering*, 321–353. https://doi.org/10.1016/B978-0-12-417295-1.00024-2

Meyer-Baese, A., & Schmid, V. (2014). Foundations of Neural Networks. *Pattern Recognition and Signal Analysis in Medical Imaging*, 197–243. https://doi.org/10.1016/B978-0-12-409545-8.00007-8

Mohanty, S., & Pathak, P. (2017). Interlinkages of Asian and US Stock Market: A Study of selected Indices. *Asian Journal of Research in Banking and Finance*, *7*(10), 1. https://doi.org/10.5958/2249-7323.2017.00114.6

Mohith, M., & Sangeetha, R. (2019). Impact of Foreign Exchange Rates on the Sectoral Indices of the Nse-. *International Journal of Research and Analytical Reviews*, *6*(July), 1199–1208.

Nauta, M., Bucur, D., & Seifert, C. (2019). Causal Discovery with Attention-Based Convolutional Neural Networks. *Machine Learning and Knowledge Extraction*, *1*(1), 312–340. https://doi.org/10.3390/make1010019

Ncube, G., & Mingiri, K. F. (2015). Stock Market Integration in Africa: The Case of the Johannesburg Stock Exchange and Selected African Countries. *International Business & Economics Research Journal (IBER)*, *14*(2), 367. https://doi.org/10.19030/iber.v14i2.9169

Nikkinen, J., & Peltomäki, J. (2020). Crash Fears and Stock Market Effects: Evidence From Web Searches and Printed News Articles. *Journal of Behavioral Finance*, *21*(2), 117–127. https://doi.org/10.1080/15427560.2019.1630125

Nimon, K. F. (2012). Statistical assumptions of substantive analyses across the general linear model: A mini-review. In *Frontiers in Psychology* (Vol. 3, Issue AUG). https://doi.org/10.3389/fpsyg.2012.00322

Nwankpa, C., Ijomah, W., Gachagan, A., & Marshall, S. (2018). *Activation Functions: Comparison of trends in Practice and Research for Deep Learning*. http://arxiv.org/abs/1811.03378

Oprescu, M., Syrgkanis, V., Battocchi, K., Hei, M., & Lewis, G. (2019). *EconML : A Machine Learning Library for Estimating Heterogeneous Treatment Effects*. *NeurIPS*.

Pagolu, V. S., Reddy, K. N., Panda, G., & Majhi, B. (2017). Sentiment analysis of Twitter data for predicting stock market movements. *International Conference on Signal Processing, Communication, Power and Embedded System, SCOPES 2016 - Proceedings*. https://doi.org/10.1109/SCOPES.2016.7955659

Panda, A. K., & Nanda, S. (2017). Market linkages and conditional correlation between the stock markets of South and Central America. *Journal of Financial Economic Policy*, *9*(2), 174–197. https://doi.org/10.1108/JFEP-08-2016-0063

Paramanik, R. N., & Singhal, V. (2020). Sentiment analysis of Indian stock market volatility. *Procedia Computer Science*, *176*, 330–338. https://doi.org/10.1016/j.procs.2020.08.035

Patel, D., & Patel, R. (2012). the Study on Co-Movement & Interdependency of Indian Stock Market With Selected Foreign Stock. *Researchers World: Journal of Arts, Science & …*, *3*(2(3)), 9–13.

Pearl, J. (2010). An introduction to causal inference. In *International Journal of Biostatistics* (Vol. 6, Issue 2). https://doi.org/10.2202/1557-4679.1203

Rao, G. V. K. (2019). AN EMPIRICAL INVESTIGATION OF CO-INTEGRATION BETWEEN SENSEX AND THE SELECTED WORLD STOCK MARKETS. *International Journal of Advance and Innovative Research*, *6*(1), 24–32.

Richardson. (2007, April). *Beautiful soup documentation*.

Rohilla Shalizi, C. (n.d.). *Advanced Data Analysis from an Elementary Point of View*.

Rosli, M. M., Tempero, E., & Luxton-Reilly, A. (2016). What is in our datasets? Describing a structure of datasets. *ACM International Conference Proceeding Series, 01-05-Febr*(May). https://doi.org/10.1145/2843043.2843059

Rosoł, M., Młyńczak, M., & Cybulski, G. (2022). Granger causality test with nonlinear neural-network-based methods: Python package and simulation study. *Computer Methods and Programs in Biomedicine, 216,* 106669. https://doi.org/10.1016/J.CMPB.2022.106669

Ross, P. T., & Bibler Zaidi, N. L. (2019). Limited by our limitations. *Perspectives on Medical Education*, *8*(4), 261–264. https://doi.org/10.1007/s40037-019-00530-x

Samadder, S., & Bhunia, A. (2018). Integration Between Indian Stock Market and Developed Stock Markets. *Journal of Commerce & Accounting Research*, *7*(1), 13–23. https://www.researchgate.net/publication/323918127_INTEGRATION_BETWEEN_INDIAN_STOCK_MARKET_AND_DEVELOPED_STOCK_MARKETS

Schmidhuber, J. (2015). Deep Learning in neural networks: An overview. *Neural Networks, 61*, 85–117. https://doi.org/10.1016/j.neunet.2014.09.003

Shah, D., Isah, H., & Zulkernine, F. (2019). Predicting the Effects of News Sentiments on the Stock Market. *Proceedings - 2018 IEEE International Conference on Big Data, Big Data 2018*, 4705–4708. https://doi.org/10.1109/BigData.2018.8621884

Sharma, A., & Kiciman, E. (2020). *DoWhy: An End-to-End Library for Causal Inference*. http://arxiv.org/abs/2011.04216

Sharma, R. K. (2020). Comparison of stock price prediction models using news articles, currency exchange rates and global indicator performance. *Journal of Advanced*

Research in Dynamical and Control Systems, *12*(7 Special Issue), 1668–1676. https://doi.org/10.5373/JARDCS/V12SP7/20202273

Sharma, S., Sharma, S., & Athaiya, A. (2020). Activation Functions in Neural Networks. *International Journal of Engineering Applied Sciences and Technology*, *04*(12), 310–316. https://doi.org/10.33564/ijeast.2020.v04i12.054

Sherif, V. (2018). Evaluating Preexisting Qualitative Research Data for Secondary Analysis. *Forum Qualitative Social Research, 19*.

Singh, R. G. (2015). An Empirical Investigation of Linkages Between India and Major Asian Stock Markets. *The Indian Journal of Commerce*, *68*(1), 30–39.

Srivastava, M., & Sharma, G. D. (2016). Risk and return linkages among stock markets of selected Asian countries. *TSME Jourrnal of Management*, *6*(1 & 2), 1–16.

Staiano, J., & Guerini, M. (2014). *DepecheMood: a Lexicon for Emotion Analysis from Crowd-Annotated News*. Association for Computational Linguistics. http://nie.mn/QuD17Z

SÜRÜCÜ, L., & MASLAKÇI, A. (2020). VALIDITY AND RELIABILITY IN QUANTITATIVE RESEARCH. *Business & Management Studies: An International Journal*, *8*(3), 2694–2726. https://doi.org/10.15295/bmij.v8i3.1540

Taherdoost, H. (2018). Validity and Reliability of the Research Instrument; How to Test the Validation of a Questionnaire/Survey in a Research. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3205040

Tank, A., Covert, I., Foti, N., Shojaie, A., & Fox, E. B. (2021). Neural Granger Causality. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1–14. https://doi.org/10.1109/TPAMI.2021.3065601

Thehindubusinessline. (2020, August 15). How the IT sector has emerged as a pillar of modern India. *Thehindubusinessline*.

https://www.thehindubusinessline.com/news/national/how-the-it-sector-has-emerged-as-a-pillar-of-modern-india/article32357389.ece

Tripathi, V., & Sethi, S. (2010). Integration of Indian Stock Market with Major Global Stock Markets. *Asian Journal of Business and Accounting*, *3*(1), 117–134.

Valentini, G. (2012). *Ensemble methods: A review*. https://www.researchgate.net/publication/230867318

van Rossum, G., & Python Development Team. (2018). Python Tutorial Release 3.7.0. *Python Software Foundation*, 1–155.

Vargas, M. R., dos Anjos, C. E. M., Bichara, G. L. G., & Evsukoff, A. G. (2018). Deep Leaming for Stock Market Prediction Using Technical Indicators and Financial News Articles. *Proceedings of the International Joint Conference on Neural Networks*. https://doi.org/10.1109/IJCNN.2018.8489208

Verma, I., Dey, L., & Meisheri, H. (2017). Detecting, quantifying and accessing impact of news events on indian stock indices. *Proceedings - 2017 IEEE/WIC/ACM International Conference on Web Intelligence, WI 2017*, *August*, 550–557. https://doi.org/10.1145/3106426.3106482

Wang, Z., Ho, S. B., & Lin, Z. (2019). Stock market prediction analysis by incorporating social and news opinion and sentiment. *IEEE International Conference on Data Mining Workshops, ICDMW*, *2018-Novem*, 1375–1380. https://doi.org/10.1109/ICDMW.2018.00195

Weisberg, S. (2005). *Applied Linear Regression Third Edition* (Vol. 528). John Wiley & Sons. www.copyright.com.

Wicht, B. (2017). *Deep Learning feature Extraction for Image Processing*. https://doi.org/10.13140/RG.2.2.11700.35207

Woiceshyn, J., & Daellenbach, U. (2018). Evaluating inductive vs deductive research in management studies. *Qualitative Research in Organizations and Management: An International Journal*, *13*(2), 183–195. https://doi.org/10.1108/qrom-06-2017-1538

Wong, C., & Ko, I. Y. (2017). Predictive Power of Public Emotions as Extracted from Daily News Articles on the Movements of Stock Market Indices. *Proceedings - 2016 IEEE/WIC/ACM International Conference on Web Intelligence, WI 2016*, 705–708. https://doi.org/10.1109/WI.2016.0126

Young, A. (2022). Consistency without Inference: Instrumental Variables in Practical Application. *European Economic Review*, *147*. https://doi.org/10.1016/j.euroecorev.2022.104112