COMPARATIVE STUDY OF MACHINE LEARNING ALGORITHMS FOR STOCK MARKET PREDICTION AND ANALYSIS OF CORRELATION BETWEEN NIFTY 50 AND GLOBAL INDICES

by

Siladitya Chatterjee, EPGP

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

JUNE 2024

COMPARATIVE STUDY OF MACHINE LEARNING ALGORITHMS FOR STOCK MARKET PREDICTION AND ANALYSIS OF CORRELATION BETWEEN NIFTY 50 AND GLOBAL INDICES

by

Siladitya Chatterjee

APPROVED BY

<Chair's Name, Degree>, Chair

RECEIVED/APPROVED BY:

<Associate Dean's Name, Degree>, Associate Dean

Dedication

I dedicate this thesis to my parents, brothers, family, wife, daughter, friends, and colleagues, teachers whose unwavering support, love, and encouragement have been my greatest source of strength throughout my academic journey. I also dedicate my research studies to financial analysts, investors, and stock market researchers.

Acknowledgements

First, I am thankful to my family and friends, especially my younger brother Sayantan, who has supported me in every situation of my life. I would like to express my love and respect to the Almighty, my father, my mother, and my wife, Sunandita, who helps and protects me in every critical situation. I would also like to thank UPGRAD, the Swiss School of Business and Management Geneva, and my mentor, Dr. Mario Silic, for their continuous support. Dr. Silic has always been there for me, and without his guidance, it would not have been possible to complete this dissertation. My lifeline is my daughter, who motivates me every time I feel depressed. I will be grateful if this dissertation helps the public, especially investors and researchers. I would like to thank our SunShell family, all the past and present employees, for helping me manage time for myself amid our daily business activities. Without the mental support of my younger brother, I could not have completed my educational journey. I would also like to thank my doctor, Dr. Suchandra Brahma, for giving me a new life. Finally, I would like to dedicate this doctorate to my daughter, Srujanika, and my brother, Sayantan.

ABSTRACT

COMPARATIVE STUDY OF MACHINE LEARNING ALGORITHMS FOR STOCK MARKET PREDICTION AND ANALYSIS OF CORRELATION BETWEEN NIFTY 50 AND GLOBAL INDICES

Siladitya Chatterjee

2024

Dissertation Chair: <Chair's Name>

Co-Chair: <If applicable. Co-Chair's Name>

The volatile nature of the stock market presents significant challenges for investors and individuals like them. Economic indicators, geopolitical events, and market sentiment influence the prediction of stock prices key sentiment. In recent years, machine learning algorithms have emerged as powerful tools for stock market prediction due to their ability to analyze large datasets and identify complex patterns. This research aims to conduct a comparative study of machine learning algorithms to estimate their effectiveness in predicting stock market trends, focusing on the Nifty 50 index.

This research will explore several machine-learning algorithms, including, but not limited to, linear regression, decision trees, random forests, support vector machines, and artificial neural networks. We will utilize historical stock market data, which encompasses a range of features like price, volume, and volatility, to train and evaluate the performance of these algorithms. Through accurate experimentation and analysis, the research will recognize the best-fit algorithm or combination of algorithms for error-less stock market prediction.

Furthermore, the research will investigate the correlation between the Nifty 50 index and key global indices, metals, and crude oil prices. Understanding these correlations is essential for investors to make informed decisions and manage risks effectively. Investors can better anticipate market motion by understanding the interrelatedness of different financial markets and commodities through statistical methods and data visualization techniques.

The findings of this research carry important implications for investors, financial analysts, and policymakers, providing them with valuable information about volatile and uncertain stock markets and their relationship with global economic indicators. This study seeks to enhance the accuracy and dependability of stock market prediction and risk management models by utilizing modern machine learning techniques and statistical analysis.

Here, I have created a stock analysis model. Using a Python function, I downloaded the dataset from Yahoo Finance, which included data from the past 20 years as of today. In our model, the user can input any listed stock, including commodities.

It first analyzes the stock and then predicts its future value. We are employing various machine learning algorithms, such as linear regression, decision trees, random forests, neural networks, and support vector machines. We compare their performance using parameters such as R-square, RMSE, and MAE, and then make a final decision on which one is the best for predicting the stock market.

TABLE OF CONTENTS

List of Tables

LIST OF FIGURES

CHAPTER I:

INTRODUCTION

1.1 Introduction Background and Scope

Predicting the future has always been an adventurous and attractive task for individuals (Usmani et al., 2016). The stock market is basically nonlinear in nature, and research on the stock market has been one of the most important issues in recent years (Adhikar et al., 2020). People invest in the stock market based on some prediction of future trends. The stock market plays a very important role in the fast economic growth of developing countries like India (Sharma et al., 2017). Predicting the stock market is a challenging task due to the involvement of several variables that are difficult to control. Researchers have extensively studied machine learning, a well-established method in a wide range of applications, for its potential in financial market prediction. Researchers have reported that popular algorithms like support vector machines (SVM) and reinforcement learning are quite effective in tracing (Adhikar et al., 2020). Popular algorithms, such as support vector machines (SVM) and reinforcement learning, have demonstrated significant effectiveness in tracing the stock market and maximizing the profit from stock option purchases, all while maintaining low risk. Prediction of the stock market has long been an attractive topic for researchers from different fields. (Shen et al., n.d.). Fundamental and technical analyses were the first two methods used to forecast stock prices. Artificial neural networks (ANNs) are the most used technique, which is a subset of machine learning and the foundation of deep learning. In most cases, ANNs suffer from overfitting problems due to the large number of parameters to fix and the lack of prior user knowledge about the relevance of the inputs in the analyzed problem. The development of Support Vector Machines (SVMs) provided an alternative that circumvented these limitations. Stock market prediction is the act of trying to determine the future value of a company's stock or other financial instrument traded on a financial exchange.

The successful prediction of a stock's future price will maximize investors' gains (Hegazy et al., n.d.). In this research, I aim to shed light on some of the ML techniques that can help investors leverage their profitability to make smarter and better investment decisions.

The global stock market is where people can buy and sell stocks of listed companies or purchase index funds. Essentially, the global stock market serves as a platform where companies sell their stocks to raise capital for their own business projects, and subsequently sell these shares to the public, adding value to them over time. The current value of a stock is determined by a variety of factors, including company revenue, financial ratios, economic indicators, geopolitical incidents, and many other parameters of the overall market condition. Since the invention of the machine learning section of modern technology, it has the power to handle a large amount of data and is also capable of finding complex patterns. Human analysts, on the other hand, cannot do this. We can predict stock market trends using machine learning.

In this thesis, we will explore the possibility of using machine learning to predict stock market fluctuations.

Historical stock market data is a common choice for machine learning predictions and analysis, such as the open price, low price, high price, close price of the day value, trading volume over services, and many other ratios and relevant metrics. By analyzing the patterns and relationships within historical stock market data, machine learning models can detect and predict stock market trends and correlations. These intricate patterns have the potential to determine future stock values.

We use machine learning algorithms in our research to easily select the most appropriate features or variables from the large dataset useful for stock price prediction. It reduces noise and enables you to focus on predictive factors.

Training Models:

We train our models using historical data. They can identify complex patterns and relationships among multiple factors, as well as the movement of stock prices. In our research, we use different algorithms such as support vector machines, artificial neural networks, random forests, decision trees, and linear regression.

In my analysis, we use data from yfinance, and we split the data into training and testing data sets. After training our machine learning model with a training data set, our models can predict the future price based on new or unseen data. These predictions are very helpful for investors and financial analysts in deciding whether to buy, sell, or hold these stocks.

Data serves as the foundation for machine learning models, enabling them to continuously learn from it. As market conditions change, the model can easily adapt to the latest data. Over time, it can predict, analyze, give accurate data, and adjust to new trends in the global stock market.

Although machine learning algorithms can provide us with predictive models and analytical and visual insights to support different finance decisions, the financial market is an overall complex and volatile space; therefore, financial market prediction will be uncertain in some cases. Furthermore, past historic data cannot reflect human behavior, market sentiment, or force majeure events such as storms, floods, situations like COVID-19, recession, drought, and geopolitical events. So, predicting using machine learning models is not a one-size-fits-all tool; investors and financial analysts should do technical and fundamental analysis in addition to machine learning models.

1.2 Research Problem

The stock market is an intricate and stimulating domain that relies on a variety of parameters, including economic indicators such as interest rates, GDP, inflation, commodity prices, and geopolitical factors. Over time, machine learning algorithms have demonstrated their effectiveness in analyzing vast amounts of data and accurately forecasting the stock market's trend. Therefore, identifying the most accurate machine learning algorithms is crucial in terms of both time and accuracy. The BSE and NSE represent the Indian stock market, not the NIFTY 50 index, which comprises the top 50 performing stocks. Domestic factors and global markets such as S&P 5are, Nasdaq, FTSE, Dow Jones, and Nikkei 225 influence the Nifty 50, and the correlation between it and global indices, metals, and crude oil prices is crucial for investors and policy makers to make informed decisions and effectively manage risk.

Therefore, the research problem in this study serves two distinct purposes:

The purpose of this study is to perform comparative analysis on various machine learning algorithms and build a model for selecting the best algorithm among different machine learning algorithms, such as linear regression, decision trees, neural networks, support vector machines, and random forests, for predicting stock market movements with reference to Indian and global contexts.

Examining the Nifty 50 index is one example of stock data analysis. Other stock markets, such as the S&P 500, Nasdaq, Dow Jones, FTSE, and Nikkei 225, as well as other commodities such as gold and crude oil, are all different tools for market analysis. We want to investigate how these factors are affecting both the Indian and global stock markets.

Once I have insights from research questions, people such as individuals, financial analysts, investors, and policymakers can predict stock market data and take decisions based on information about whether they should buy, sell, or hold stocks.

The Nifty 50 index is the sum of India's top-performing 50 stocks. It is very closely related to and interconnected with global stock indices such as the S&P 500, Dow Jones, Nasdaq, FTSE, Nikkei 225, and others, as well as metals, crude oil, and other stocks. As these factors change, so does Nifty 50. In our research studies, we will observe how the Nifty 50 performs when these factors change over time.

Suppose we want to conduct a correlation analysis between the Nifty 50 index and selected global indices, metals, and crude oil prices over a given time. In our studies, we can correlate the Indian stock market with the global market, metals, and commodities using the Pearson correlation coefficient, or Spearman rank correlation coefficient. By graphically representing ideas while making capital market decisions, we can clearly get them.

1.3 Purpose of Research

This research is to study or examine certain phenomena. Significant factors currently affect both the global and Indian stock markets. Many of our economic indicators, such as gross and per capita domestic product figures, the consumer price index, unemployment rates, and the interest rate on government securities, give direct signals about the market's condition. Major international political and economic events influence Indian stock market and other nations, impacting our stock market through a kind of economic osmosis. For instance, an intra-month spike in the price of oil, a sudden drop in the price of copper (as China's economy slows down), or a flare-up in the Middle East (a geo-political event that raises the price of oil) all have short-term and long-term implications for the stock market. Meanwhile, three crises that have been happening since 2007—the US mortgage crisis, the default of Greece, and now the move by Great Britain to exit the European Union—have sent shock waves and after-shocks through the stock markets of Europe, the US, and Asia.

The investigation's goals are:

The research has two main goals. The first is to compare the performance of different machine learning algorithms, like linear regression, decision trees, random forests, and support vector machines, in predicting the stock market. We will use performance metrics such as mean squared error, root mean squared error, and R-squared to accomplish this. The second goal involves predicting the stock market using the most successful algorithms and evaluating the performance of global stocks, specifically U.S. stocks, and Indian stocks in relation to these predictions.

One can study the relationship between the Nifty 50 index and various global indices such as the S&P 500, FTSE 100, and Nikkei 225.

One can analyze how the Nifty 50 index relates to other global indices, like the S&P 500, FTSE 100, and Nikkei 225. You can also begin examining the relationship between the Nifty 50 index and metals such as gold, silver, and copper, or raw materials like crude oil. Additionally, you can explore the relationship between the Nifty 50 index and various banking and financial parameters like interest rates and the money supply. By doing this and many other kinds of analysis, you will begin to understand why the movements of the Indian stock market differ from those in other markets around the world.

Methodology:

One of the key strengths of machine learning algorithms is their aptitude for isolating and extracting pertinent features from gathered data. Consider, for instance, the prices of a stock (e.g., open, low, high, close) over some span of time; these four numbers might seem insufficient on their own to really say much about the market's feelings, the state of the economy, or the stock's prospects. However, these same numbers can and do serve as the foundation for a vast array of distinct "features" that train a stock-market prediction algorithm.

Training the Models: For this study, we will use data that spans the histories of various stocks, including the Nifty50 and global indices, as well as metals—in other words, anything for which we have a curve. Using this historical data, we will first generate several models based on different machine learning algorithms. Our research studies will utilize models such as linear regression, decision trees, random forests, neural networks, support vector machines, and the XGBOOST algorithm.

Evaluate each model's performance using proper metrics, for instance, R-squared, mean square error, mean absolute error, and root mean square error. As per the output of our model, we can then make our final decision based on the scores of these parameters. To calculate the parameters, we will use Python libraries and methods.

Understanding how different variables behave in relation to one another, especially when holding one variable constant, is our main goal. We accomplish this primarily by using two well-known statistical methods: Pearson correlation and Spearman correlation. These two approaches are indispensable to us because they allow us to describe the kinds of relationships among variables that pertain to our model. We can take decisions based on the positive or negative relationships between two variables, for example, the Nifty 50, Nasdaq, or other global indices.

In our research studies, we visualize the relationships between one variable and another by creating heatmaps or simple scatterplots. These visuals give us a much clearer depiction of the kinds of relationships that might exist between pairs of variables.

The use of varied sentence length helps create a more interesting passage. If all the sentences are the same length and style, the work tends to become monotonous. These can be beneficial, as they allow us to understand the relationship between the variables we are studying. Beyond the numbers, algebra can help us find out the underlying pattern. Heatmaps will show you how pairs of variables tend to go up and down in tandem in a very visual, striking way, while scatterplots will help you see the relationship (or lack of one) between two variables more clearly.

Pearson's correlation coefficient and Spearman's correlation coefficient have different applications. Pearson's correlation coefficient tells us when the data is "linear," but it won't tell us much about non-linear (or any other kind of) data. However, if a linear representation of the relationship is bad, i.e., it makes the relationship look weaker than it really is, then Spearman's correlation coefficient saves the day. Also, heat maps and scatter plots have roles to play in meaningful exploratory data analyses, which can help us get a better understanding of how different variables behave with respect to each other.

This kind of plot can assist a modeler in getting a better grasp of the model, its underlying structure, biases, and abilities. Indeed, visualization plays a crucial role in comprehending both linear and nonlinear models, often serving as the initial stage of model analysis. Visualization techniques can aid model exploration in two ways: correlation and performance.

The goal of this study is to clarify the most effective machine learning algorithms for stock market movement prediction. Making accurate forecasts is an imperative task on a variety of levels. By doing so, it provides investors with superior input and assists them in preparing a better investment strategy based on the prediction of the volatility of the stocks they hold. Moreover, it gives investors the freedom to make any investment decision, including whether to buy or sell the stock. Regulators, another important group, urgently require accurate predictions due to concerns that unscrupulous individuals may attempt to manipulate the market, potentially harming investor interests. The public is interested in the stock market, and institutional investors need accurate forecasts. Based on our models, we present potential risks and rewards in a more logical manner, drawing financial analysts' attention to additional econometric data.

Understanding the movement of the Nifty 50 and its correlation with global indices, metal prices, and crude oil stocks allows us to observe the relationships between these factors and their impact on our own market. While these external forces may not consistently propel our market to substantial gains or losses, they demonstrate the impact of economic conditions and global market trends on our markets, and how these conditions and trends can influence any market we invest in.

Certain financial systems that we will study in this course may be close to perfect and may also exhibit great adaptability. We will study and record their movements, which may include deep model correction and high adaptiveness. We will study them because our own system may contain some components that can be improved. The term we used for this activity is "research." We will go beyond simply observing a system. We'll look for a way to dramatically improve it.

To sum up, what we aim to do is determine whether we can use machine learning algorithms to predict stock market movements. We will focus on the Nifty 50 and compare it to other global indices as well as metals and crude oil prices. We aim to determine if there is a relationship between these indices and assets that could potentially outperform chance in predicting stock prices. And if so, what could we say, usefully, about that relationship?

1.4 Significance of the Study

The advancement in predictive analytics for financial markets is a significant accomplishment. Stock market prediction is a very complex and critical task. To predict this very volatile stock market, as many factors govern the stock market movement, modern mathematics, statistics, and machine learning algorithms have shown great potential to identify complex patterns, which also leads us to make somehow errorless predictions.

Identify the most effective ML algorithms:

The main objective of these research studies is to find out the most effective machine learning algorithms. Identify the best and most accurate machine learning algorithms that predict stock market movements. This could be very helpful for policymakers, investors, and analysts to choose the right tools and techniques for their analysis.

Having the highest prediction accuracy:

Using these studies, accurate predictions can lead to proper and better investment strategies and risk management. Finally, it provides better returns and reduces losses on long-term investments.

Informed investment decisions:

In today's financial world, both institutional and individual investors rely heavily on accurate market forecasts before making decisions about buying and selling investments. Almost everyone in the investment business, whether Wall Street hotshots, small traders working from home, or the people who set up shop as "robot-advisors," promises to use some form of artificial intelligence to make these forecasts. Yet, it's still not clear which AI methods work best, which ones don't, and what the proper way is to evaluate them. Our research aims to tease out these questions and provide actionable answers for market participants.

Understanding market dynamics:

Accurate and reliable forecasts are of paramount importance when it comes to asset allocation, as this can help set up the right composition of investments. Necessarily, market dynamics are at the heart and core of what we need to understand to asset-allocate effectively. This is because it asks us to look at the forces driving that singular market (or investment) upward or downward over time. At the same time, allocating assets accurately sets portfolio management up for better returns. Portfolio managers, on the other hand, are forward-looking business partners with their clients who seek to look through the murky dimension that is the "unknown future" and extrapolate the best possible set of investment instructions. Identifying the forces that drive change and then categorizing them will give us added confidence in navigating through the more minute details of market dynamics.

As previously mentioned, these research studies aim to forecast stock price movements and establish a stronger visual association between the Nifty 50 Index and other global indices, as well as metals and crude oil prices, among others, with potentially useful implications.

Insights into the Global Market:

By applying Pearson correlation and Spearman correlation coefficients to research the relationship between Nifty 50 and global indices prices, we can figure out how international markets impact the Indian market. These findings will assist investors in strategically allocating their global portfolios.

Insights into the Commodity Market:

Here we carry out an analysis and build up correlations with the help of Python libraries between the metals, crude oils, and nifty 50 indexes in the commodity market, any investor can easily understand how the market is moving based on the change in commodity prices. If we observe a rise in the price of crude oil, it will significantly impact all industries related to the oil sector. Therefore, we can expect a change in the stock prices of the oil sector.

Risk Management:

Managing risk is an essential process to minimize or reduce losses within an organization. A risk management approach minimizes surprises, conducts regular reviews of how losses occur, and analyzes and anticipates trends. The readings suggest that many organizations only consider risk management measures after experiencing a major loss. The design of a risk management process aims to mitigate the likelihood of a loss and the associated uncertainty. We can formulate a risk management program by frequently analyzing and planning for the wide range of risks that may arise, anticipating when there will be insufficient time to plan for a specific risk, and establishing a continuous process. A risk management decision-making framework is a general approach to solving problems. We recommend adopting a five-step process that includes identifying the issue, developing options, selecting and implementing an option, evaluating and monitoring the outcome, and communicating the solution.

These correlational studies or prediction models can be included in risk management techniques.

Strategies for Hedging:

By making use of information like correlations, investors can devise hedging strategies. For example, if there is a strong negative correlation between Nifty 50 and gold prices, investors can protect their investments with gold futures.

Scenario Analysis: This study will reveal how various factors affect the stock market. Investors will gain a better understanding and prepare for different economic scenarios to reduce the impact of negative market fluctuations.

Contribution to Academic Research:

This study has two primary purposes. One is to validate the prevailing theories of how to predict the stock market and calculate its relationships to other financial instruments. If we apply a specific theory, such as predicting a stock's future price based on its recent price fluctuations, we aim to find empirical evidence to support its practical application. An additional purpose is to teach us, or future researchers, what works and doesn't for stock market forecasting and figuring out the "real" economy. Do human forecasters make useful predictions? If so, under what circumstances? Do forecasting models, even those of the so-called artificial intelligence genre, do any better, or are they just a bunch of "garbage in, garbage out" mechanisms? In 2007, before the Great Recession, several machines and humans engaged in forecasting contests. Models won narrowly, but it seemed to me that it was almost a rigged game. Human forecasters need models, too, just as some models need humans to feel secure in their use.

Technological and Methodological Advancements:

As a research analyst, I aim to conduct research that brings the academic community forward in both technology and methodology. Specifically, I focus my research on developing new algorithms and enhancing data processing.

Our algorithms aren't always efficient or effective. For example, consider my chosen area of study, stock market prediction. The current state of the art is not exactly mind-blowing; if it were, you'd have heard about it by now. So, my chosen challenge is to build and train better algorithms.

The implications for policy and regulation are as follows:

Machine-learning algorithms can now accurately predict market dynamics in the stock market. These studies of such momentousness can provide us with significant policy insight. Regulators overseeing our economy can clearly identify the features, models, and correlation charts, as well as the types of policies, that could aid in stabilizing the stock market and preventing its complete disruption. We can assess the overall effectiveness of existing policies in influencing market fluctuations, as the volatile stock market consistently impacts on our economy. When similar situations arise globally, we can formulate prudent policies accordingly.

Practical Applications:

This study's numerous practical applications make it quite important. They assist us in numerous ways, such as algorithmic traders who seek to enhance their stock prediction systems with high-performing machine learning models can do so using integrative methods.

Wealth Management:

With our research studies, a team of experts can forecast which stocks will perform the best and offer these as recommendations to our clients. They do this with the help of a predictive model, which has been built from a wide set of data that includes information on the economy as a whole and on individual businesses.

Tools for analyzing the market:

Financial software businesses can use the information they find in their market research to create market intelligence tools that are as error-free and as complete as possible. One of the major reasons why our study is important is that it provides a machine learning model for using predictive analytics in financial markets. This is a field that is complex, volatile, and difficult to forecast using traditional methods. As a result, predictions in this field are critical.

1.5 Research Design

The research design describes the methodological approach to conducting a comparative study of machine learning algorithms for predicting stock market movements. Here, we will specifically focus on the Nifty 50 index and its correlation with global indices, metals, and crude oil. We use the Python library to directly download the data sources from Yahoo Finance, preprocess the data, apply machine-learning models, and evaluate the data using metrics like R square value, MSE, MAE, RMSE, and other analytical tools and techniques.

Our main research objectives are to first compare various machine learning algorithms and their performance metrics, with the aim of selecting the best or top-performing algorithms. Secondly, we aim to examine the correlation between the Nifty 50 index and various global indices, metals, and crude oil prices. Finally, we draw the conclusion that the other variables influence both the global market and the Nifty 50 index.

Data Sources:

1. Nifty 50 Index: Directly downloaded data from Yahoo Finance using Python libraries and methods.
2. Global Indices: The S&P 500, Nasdaq, FTSE, Nikkei 225, and Dow Jones directly downloaded data from Yahoo Finance using the Python library.
3. Metals and Crude Oils: Gold and Crude Oil Data directly from Yahoo Finance.

Time Frame: We will collect the data for the last 20 years from the current date.

Data Preprocessing:

Utilizing data cleaning, normalization, and feature engineering.

Machine learning models:

Model Selections:

1.Linear Regressions 2. Decision trees 3. Random Forests 4. Support vector machines.

5.Neural Networks 6. XGBoost

Training and testing:

In these studies, we are using the train-test split method to split the dataset into training (70%) and testing (30%) parts, which will help to evaluate the models' performance.

Evaluation Metrics:

Because our research is based on number variables, we use Mean Absolute Error (MAE), Root Mean Square Error (RMSE), R-Squared Error, and Mean Square Error (MSE) to evaluate the prediction machine-learning algorithms.

Correlation Analysis:

As our research study is twofold, we will depict and analyze the correlation between the Nifty 50 index, global indices such as the S&P 500, Nasdaq, Dow Jones, FTSE, Nikkei 225, etc., commodity prices, and crude oil prices. We will use Pearson and Spearman correlation coefficients.

Analysis of the Results

We are analyzing the performance of different machine learning algorithms based on evaluation metrics. Select the top-performing algorithm for stock price prediction. We will achieve the best outcome based on an in-depth analysis and the results.

Correlation Insights:

By analyzing the correlation results, one can get correlation insights and understand the effects of global indices and commodities on the Nifty50 index.

Impact of External Factors:

External variables such as GDP, interest rates, and other macroeconomic factors, as well as external shocks such as floods, droughts, and geopolitical events, have an impact on prediction models and correlation.

Tools and Technologies:

GoogleColab coding platform, Python programming languages, and machine learning algorithms for analysis. We have used built-in libraries and frameworks such as Sklearn, yfinance, NumPy, Pandas, Keras, and other required libraries.

For visualization of data, we have used Matplotlib, Seaborn, and plot functions.

1.6 Structure of the Thesis

There are six main chapters in this thesis.

Chapter One: Introductions This chapter includes sections where we discuss the background, scope, research problems, research purposes, significance of the study, research design, and research structure.

Chapter Two, the Literature Reviews section, describes previous research papers and the theory required to conduct our research, including internal and external factors.

Chapter Three: The Research Methodology section will cover a variety of topics, beginning with an overview of the research problem, operationalization and theoretical constructs, research purpose and questions, research design, populations and samples, dataset and datapoint selection, tools and techniques used, data collection procedures, data analysis, and research design limitations.

Chapter Four: In this chapter, we will discuss the data analysis of individual stock as well as global indices, gold, and crude oil.

Chapter Five: In this section, we will discuss the findings, recommendations, and conclusion of our research studies.

Chapter Six: In this section, we will give a better understanding of the summary, implications, and recommendations of our research studies.

CHAPTER II:

REVIEW OF LITERATURE

2.1 Previous Research Studies:

In this section, we will analyze past studies that deal with stock market prediction. According to Kumar et al. (2022), stock market prediction is an important, effective, and challenging activity. Stock prices will lead to lucrative profits if investors can make the right decision at the right time. Stock market predictions use mathematical strategies and learning tools. A study by M. Hiransha et al. (2018) says that many researchers use different algorithms to guess what will happen in the stock market. These algorithms can be put into two groups: linear models (AR, MA, ARIMA, ARMA) and non-linear models (ARCH, GARCH, and neural networks). The neural networks use four different types of deep learning architectures: MLP, RNN, LSTM, and CNN.

Parmar et al.'s (2018) main goal is to predict the future value of a company's financial stocks using past and historical data. They used regression and LSTM-based machine learning methods with factors like open, close, low, high, and volume.

The stock market research study is mainly based on supervised learning. To overcome the situation, Lee (2001) applied reinforcement learning and artificial neural networks. They consider the stock price prediction problem to be a Markov process that reinforcement learning can optimize.

Research studies of Patel et al. (2015), they took four prediction models, such as ANN, SVM, Random Forest, and Naïve Bayes, with two different approaches. The experimental results demonstrate that presenting these technical parameters as trend-deterministic data improves the performance of all prediction models.

Some authors, such as Khare et al. (2017), have developed two distinct sorts of artificial neural networks: feed-forward neural networks and recurrent neural networks. The experimental results showed that Feed Forwards Multilayer Perception performed better than LSTM at predicting a stock's short-term price.

In agreement with Usmani et al. (2016), the market performance of the Karachi Stock Exchange was predicted based on attributes such as oil prices, gold and silver prices, interest rates, foreign exchange (FEX) rates, news, and social media feeds. They also use old statistical techniques as input, such as simple moving averages (SMA) and autoregressive integrated moving averages (ARIMA). They compare machine learning techniques such as single-layer perceptron (SLP), multi-layer perception (MLP), radial basis function (RBF), and support vector machine (SVM). Among all these attributes studied, the algorithm MLP performed best as compared to other techniques. We found the oil rate attribute to be the most relevant to market performance.

In consonance with Vijh et al. (2020), predicting stock market accuracy is a very critical task due to the volatile and non-linear nature of stock markets. They applied artificial neural networks and random forest techniques to predict the next day's closing price. Standard strategic indicators, such as the RMSE and MAPE, evaluate the models.

Many researchers follow classical and modern approaches to predicting stock markets. Fundamental analysis and technical analysis. Stock market prediction also employs modern approaches like machine learning and sentiment analysis. There are several methods available for prediction, such as artificial neural networks (ANN), support vector machines (SVM), naive bayes, genetic algorithms (G.A.), fuzzy algorithms, deep neural networks, regression algorithms, and hybrid approaches.

As per the study of researchers Hegazi et al. (2013), the proposed algorithm integrates particle swarm optimization (PSO) and least squares support vector machines (LS-SVM). The PSO algorithm uses optimized LS-SVM to predict the daily stock prices.

Based on Singh and Srivastava (2016), it is widely accepted that the stock market is chaotic, complex, volatile, and dynamic. They have used deep learning to accurately predict the stock market. In their research, they evaluated the Google stock price multimedia chart from NASDAQ.

As reported by Sharma et al. (2017), the nonlinear nature of stock market prediction plays an important role in the long-term and short-term profit of any investor. They have mostly used regression analysis.

As stated by Shen et al., they have used a support vector machine algorithm to predict the next-day stock trend with data from NASDAQ (accuracy 74.4%), S&P500 (accuracy 76%), and DJIA (accuracy 77.6%).

Conforming to Mehtab and Sen (2020), stock market prediction is a highly difficult task. As per the EMH (Efficient Market Hypothesis), developing a predictive framework for stock market prediction is impossible. They construct a very robust and accurate framework using a collection of statistics, machine learning, and deep learning models. They have used eight classification and eight regression models.

As reported by Sai Reddy (2018), they have used a machine learning approach and SVM (Support Vector Machine) to predict three different markets in detail.

As per research studies by Khan et al. (2019), the model with the highest accuracy and consistent prediction is the Random Forest classifier (83.22%). After that, deep learning achieved 80.53%. They have used input data for social media and financial news.

On the word of Hu et al. (2006), the best machine learning algorithms for predicting the stock market are long short-term memory (LSTM), convolutional neural networks (CNN), deep neural networks (DNN), recurrent neural networks (RNN), reinforcement learning, and other deep learning methods such as Wavenet, self-paced learning mechanisms (NLP), and hybrid attention networks (HAN). Also, they consider performance metrics such as root mean square error (RMSE), mean absolute percentage error (MAPE), mean absolute error (MAE), mean square error (MSE), accuracy, Sharpe ratio, and return rate.

Many researchers, for example, Adhikari et al. (2020), focus on RNN (recurrent neural networks) and LSTM (long short-term memory) technologies in predicting the ongoing trend of the stock market.

As said by Yeom et al. (2011), the accuracy of the variance considered machine (VCM) is superior to that of the support vector machine.

Strader et al. (2020) classified stock market prediction studies into four main categories: artificial neural networks (ANN), support vector machines (SVM), genetic algorithms (GA), and other artificial intelligence approaches. They studied these four categories and designed a prediction model.

Adebiyi et al. (2014) conducted the prediction study for the New York Stock Exchange and Nigeria Stock Exchange. They used the ARIMA model for their research work.

Many researchers such as Jain et al. (2018), have analyzed and made a model using as many parameters as possible from the Nifty50 stock. Their objective is to create a high-profit stock portfolio.

Vadlamudi (2017) discussed the necessity of the stock market and its ability to motivate and support investors in making informed decisions. What are the various attributes and factors available for stock market prediction? What are the machine learning algorithms to predict the stock market accurately?

According to the research study of Mizuno and Komoda (1998), they present a neural network model for technical analysis of the stock market. Their applications predict the buy and sell times to maximize profit.

M et al. (2022) discuss the Holt-Winters algorithm in their paper. The research work uses a real-time dataset of fifteen stocks as input data, and based on that data, it predicts and forecasts future stock prices in different sectors.

This research aims to comprehend the functioning of stock markets and explore the use of simple and effective statistics and machine learning algorithms for stock market prediction. Stock market predictions use mathematical strategies and learning tools. A 2018 study by M. Hiransha et al. says that many researchers try to guess what will happen in the stock market by using different algorithms, such as linear models (AR, MA, ARIMA, ARMA) and non-linear models (ARCH, GARCH), as well as four different types of deep learning architectures (MLP, RNN, LSTM, and CNN).

Parmar et al.'s (2018) main goal is to predict the future value of a company's financial stocks using past and historical data. They used regression and LSTM-based machine learning methods with factors like open, close, low, high, and volume.

Stock market research studies primarily rely on supervised learning to address various situations. Lee (2001) applied reinforcement learning and artificial neural networks. They consider the stock price prediction problem to be a Markov process that reinforcement learning can optimize.

Patel et al. (2015), took four prediction models, such as ANN, SVM, Random Forest, and Naïve Bayes, with two different approaches. The experimental results demonstrated that representing these technical parameters as trend-deterministic data improves the performance of all the prediction models.

Some authors, such as Khare et al. (2017), have developed two distinct sorts of artificial neural networks: feed-forward neural networks and recurrent neural networks. The experimental results revealed that Feedforward Multilayer Perceptron outperformed LSTM in predicting a stock's short-term price.

As mentioned by Usmani et al. (2016), the market performance of the Karachi Stock Exchange was predicted based on attributes such as oil prices, gold and silver prices, interest rates, foreign exchange (FEX) rates, news, and social media feeds. They also use old statistical techniques as input, such as simple moving averages (SMA) and autoregressive integrated moving averages (ARIMA). They compare machine learning techniques such as single-layer perceptron (SLP), multilayer perceptron (MLP), radial basis function (RBF), and support vector machine (SVM). Among all these attributes studied, the algorithm MLP performed best as compared to other techniques. We found the oil rate attribute to be the most relevant to market performance.

Following Vijh et al. (2020), predicting stock market accuracy is a very critical task due to the volatile and non-linear nature of stock markets. They applied artificial neural networks and random forest techniques to predict the next day's closing price. Standard strategic indicators, such as the RMSE and MAPE, evaluate the models.

As per the study of researchers Hegazi et al. (2013), the proposed algorithm integrates particle swarm optimization (PSO) and least squares support vector machines (LS-SVM). The PSO algorithm optimizes LS-SVM to predict the daily stock prices.

As stated by Sharma et al. (2017), the nonlinear nature of stock market prediction plays an important role in the long-term and short-term profit of any investor. They have mostly used regression analysis.

As per the research of Shen et al., they have used support vector machine algorithms to predict next-day stock trends with data from NASDAQ (accuracy 74.4%), S&P500 (accuracy 76%), and DJIA (accuracy 77.6%). According to the research studies of Mehtab and Sen (2020), stock market prediction is a highly difficult task. The EMH (Efficient Market Hypothesis) asserts that developing a predictive framework to forecast the stock market is impossible. They construct a very robust and accurate framework using a collection of statistics, machine learning, and deep learning models. They have used eight classification and eight regression models. As per Sai Reddy (2018), they have used machine learning approaches and SVM (Support Vector Machine) to predict three different markets in detail.

As per the account of Hu et al. (2006), the best machine learning algorithms for predicting the stock market are long short-term memory (LSTM), convolutional neural networks (CNN), deep neural networks (DNN), recurrent neural networks (RNN), reinforcement learning, and other deep learning methods such as Wavenet, self-paced learning mechanisms (NLP), and hybrid attention networks (HAN). Also, they consider performance metrics such as root mean square error (RMSE), mean absolute percentage error (MAPE), mean absolute error (MAE), mean square error (MSE), accuracy, Sharpe ratio, and return rate.

Many researchers, such as Adhikari et al. (2020), focus on RNN (recurrent neural networks) and LSTM (long short-term memory) technologies in predicting the ongoing trend of the stock market.

Adebiyi et al. (2014) conducted the prediction study for the New York Stock Exchange and Nigeria Stock Exchange. They used the ARIMA model for their research work.

Some researchers, such as Jain et al. (2018), have analyzed and made a model using as many parameters as possible from the Nifty50 stock. Their objective is to create a high-profit stock portfolio.

In Vadlamudi's research studies (2017), they discussed why the stock market is necessary and how it can motivate and support an investor's right decision. What are the various attributes and factors that can aid in stock market forecasting? What is the machine learning algorithm for accurately forecasting the stock market?

The research studies of Mizuno and Komoda (1998), they present a neural network model for technical analysis of the stock market. Their applications predict the buy and sell times to maximize profit.

As observed by M et al. (2022), their paper discusses the Holt-Winters algorithm. The research work uses a real-time dataset of fifteen stocks as input data, and based on that data, it predicts and forecasts future stock prices in different sectors.

Various researchers have asserted that fluctuations in a country's macro-economic conditions either directly or indirectly influence the growth of its stock markets. The volatility and complexity of stock market fluctuations mirror the macroeconomic conditions described by Patel (2012).

Gupta (2017) conducts a thorough analysis of the stock market's volatility, concluding that factors such as a broad money supply, inflation, the credit/deposit ratio, and a fiscal deficit, rather than political insecurity, primarily influence stock market movements. The primary market is dominated by innovative securities issues. The stock market is a secondary market where investors regularly buy and sell stocks.

2.2 Machine Learning Algorithms

Kharwal (2023) stated there are various machine learning algorithms available. Here we will discuss about the Models in details. In today's world there are three.

types of algorithms: Supervised learning here the algorithm learns from labelled data, which means the input data is already associated with known output values. This can be used as regression, classification, and prediction. Unsupervised learning, learning the algorithms from the unlabelled data, which means the input data is not associated with any known output value. This is used for tasks such as clustering, anomaly detection, and dimensionality reduction. The third one is Reinforcement learning which is used for gaming, robotics, and autonomous vehicles.

Machine learning techniques are divided into mainly six types. A. Regression: it is a technique used to predict a continuous output variable based on one or more input variables. B. Classification: This technique used to predict a discrete output variable based on one or more input variables. This technique is commonly used in tasks such as image classification and spam filtering. Common classification algorithms include logistic regression, decision trees, and Support Vector Machines. C. Clustering: This technique used to group similar datapoints based on their characteristics. Commonly used in customer segmentation and image segmentation. Common algorithms are k-means clustering, hierarchical clustering and density-based clustering. D. Dimensionality reduction: This is a technique by which one can reduce the number of features or variables in the dataset without losing the important information. This technique commonly used in data visualization and features selection. E. Ensemble methods: By these techniques that combine multiple machine learning models to improve their accuracy and robustness. These techniques commonly used in predicting stock prices, detecting fraudulent transactions. Common ensemble methods are bagging, boosting, random forests. F. Deep Learning: It is the subset of machine learning that uses artificial neural networks to learn complex patterns and relationships in data. Here in our research study, we will discuss about the regression techniques and artificial neural networks.

Regressions Performance Evaluation Metrics:

As our research is quantitative in nature that is why we will use regression performance evaluation metrics to evaluate or checking the performance of our model. In this section, we will explore a range of regression performance evaluation measures which will give a valuable information about the model's accuracy, reliability, and robustness. Here in our research study, we are using MSE (Mean Squared Error), RMSE

(Root Mean Squared Error), Mean Absolute Error (M.A.E), R-squared Error(R^2) and adjusted R-Squared Error.

Mean Squared Error:

Among all, the mean squared error is the most widely used one to evaluate the accuracy and precision of a regression model. It gives us information about the overall accuracy and precision of a regression model by simply calculating the average of the sum of the squared error between predicted and actual values.

$MSE = 1N * \sum Ya - Yp^2$ [Where N is the total number of data points ,Ya is representing the actual value of the target variable., Yp represents the predicted value of the target variable.]

Root Mean Squared Error:

This is also widely used regression evaluation metric that measures average magnitude of prediction errors in a regression model. It is a variant of the mean squared error (MSE) but provides the result in the original scale of the target variable.

$RMSE = 1N * \sum Ya - Yp^2$ [Where N is the total number of data points ,Ya is representing the actual value of the target variable., Yp represents the predicted value of the target variable.]

Mean Absolute Error:

It is also a common regression evaluation metric, which measures the mean absolute difference the predicted and actual values in a regression model. It gives us a measure of the magnitude of prediction errors.

$MAE = \ 1N * \sum|(Ya - Yp)|$ [Where N is the total number of data points, Ya is representing the actual value of the target variable., Yp represents the predicted value of the target variable.]

R-Squared:

R-Squared, also it is known as coefficient of determination, is a widely used as regression evaluation metric that measures how well the independent variable explains the variance in the dependent variable in a regression model. It indicates the fit of the regression model to the data.

Adjusted R-Squared:

It is a variation of the R-Squared metric that considers the number of independent variables used in a regression model. It primarily provides an adjusted measure of the fit of the model to the data, considering the complexity of the model and the potential for overfitting.

From the (*Correlation Coefficient: Simple Definition, Formula, Easy Calculation Steps*, 2024) we will have a better understanding of correlation coefficients, here we are using mainly two coefficients to understand the relation between nifty 50 and other global. indices, metals, and crude oils price.

Correlation coefficients are the statistical measures that describe the strength and the direction of a relationship two variables.

Pearson Correlation Coefficient (r):

It measures linear relationship between two continuous variables. If coefficient value is +1 then we can say this is perfect positive linear relationship. We can conclude if coefficient is -1 then the relationship is perfect negative linear relationship. In case of r=0 we will say it is no relationships. Other types of correlation coefficients are **Spearman's Rank correlation coefficient** which measures strength and direction of the monotonic relationship between two ranked variables.

Kendall's Tau: Measures the strength and direction of association between two ranked variables.

2.3 Internal and External Factors

Many researchers, like Sahu (2024), assert that the conventional method of stock market prediction involves analyzing the stock's past movement. However, the nature of the stock market is volatile and complex, so predicting it is very difficult. This volatility is caused by a variety of socioeconomic factors.

Public sentiment or opinion is a major contributing factor. These sentiments may come from an investment advisor, a financial website, or a blog. Stock market investors. These sentiments can sometimes be influenced. Therefore, we cannot ignore sentiment analysis while predicting the stock market.

Sekhar (2023) outlines the two primary stock market analyses, both of which are equally important and necessary for prediction. Fundamental analysis and technical analysis.

According to him, fundamental analysis is the process of evaluating a company's intrinsic value by analyzing its financials, such as revenue, earnings, assets, liabilities, and cash flow. The main aim of fundamental analysis is to determine whether the company's stock is overvalued or undervalued.

Chang et al. (2010) proposed a neural network-based stock price trend prediction system that uses financial ratios such as earnings per share, price-to-earnings ratio, and return on equity as input features.

Li et al. (2014) proposed a sentiment analysis based on news and social media, a prediction model, and tested it on ten Chinese stocks. The author showed their testing model is effective.

Citing Sekhar (2023), technical analysis is a method in which researchers or financial analysts analyze market data, such as price and volume, to identify complex patterns and trends for predicting future value. It focuses on market trends rather than fundamental analysis for stock price prediction.

One of the best technical analysis tools is the Relative Strength Index (RSI), which measures the strength of a stock's price relative to its past performance. Zhang et al. (2016) proposed an RSI analysis prediction model and random forest regression. The authors demonstrated the superior performance of their proposed model.

Researchers specifically use machine learning to study historical stock price data, while slant analysis plays a crucial role in scrutinizing data gathered from online social networks. Online social network information has more influence than ever before and might help in predicting the direction of the stock market. We primarily conduct slant analysis based on news sources and information from social networking sites.

Kumar (2024) makes a good point: predicting the value of stocks requires extensive information about market share values and patterns. Experienced analysts in this field can only provide that information; the average person needs a lot of time and effort to do the same. With the massive improvement in technology, machine learning algorithms can reliably forecast future stock market trends. The author's findings suggest a mixed ARIMA-GRU model that continuously adapts to new information about customers' stock trading data and holdings information. The ARIMA-GRU model looks for trends and determines which type of data is useful for making accurate value predictions.

Internal Factors Influencing Stock Price Movement:

Company Financial Performance:

Influences on stock prices come from within the companies themselves or from other external events. The performance of a company, particularly its financial performance, directly influences the price of the stock that its investors hold. Information about a company's financial performance, commonly expressed through the issuance of quarterly and annual earnings reports, serves as the principal guide, but not the only guide, for investors to determine when to buy or sell a stock.

Corporate Actions:

Corporations' actions can have an impact on their stock prices. For example, when a company declares a dividend or a dividend increase, people see that as a positive sign. Consequently, the stock might appreciate. One reason for this could be that the company now looks solid, and people believe that if they invest in it, they will get a good return. When a company performs well, it buys back its own shares, reducing the total number of shares in the market. This can potentially cause the stock price to increase.

Management and Leadership:

When CEOs and executives resign or move around, it can shake things up. When the company's head changes, it signifies a shift in direction. Maybe the new leadership will take the company to new heights, or maybe not. Regardless, we know that announcements of executive changes can cause major market tremors, which can lead to poor investment results. That's why in last month's Digest, we asked two Stanford Graduate School of Business Finance professors to tell us more about their recent research on the ways executive changes can affect stock prices and investor confidence.

Product and Innovation:

Product launches have the potential to significantly increase profits and have a direct positive impact on market share. Both are crucial for achieving premium valuations in our industry. The entire company, as well as stock prices, will increase with successful product launches.

Operational Efficiency:

Managing costs well and making supply chains smooth, efficient, and effective are both significant for an organization. When executed effectively, they have the potential to significantly impact an organization. But what exactly is the connection between cost control and efficient supply chains, on the one hand, and stock performance, on the other hand?

External Factors Influencing Stock Price Movement:

Economic Indicators: GDP Growth: Robust economic growth can propel both corporate earnings and stock prices.

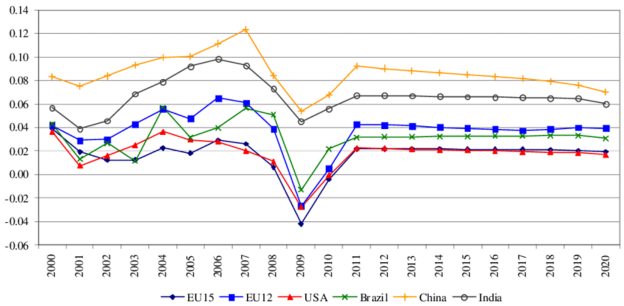


Fig 2.3.1

Real GDP growth rate in baseline scenario with crisis, 2000-2020

Inflation Rates:

When the cost of living goes up rapidly, people can afford to buy less, and businesses can earn less of a profit. This can lead to a decrease in stock prices.

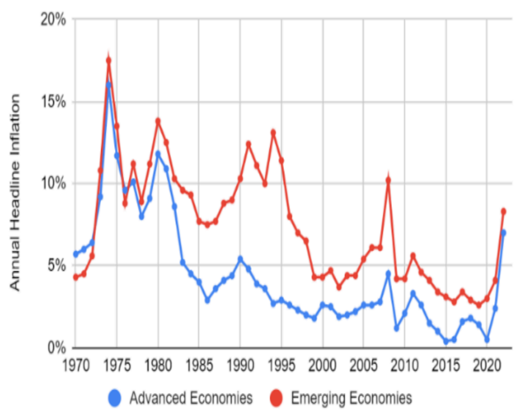


Fig 2.3.2

Global Inflation rate information

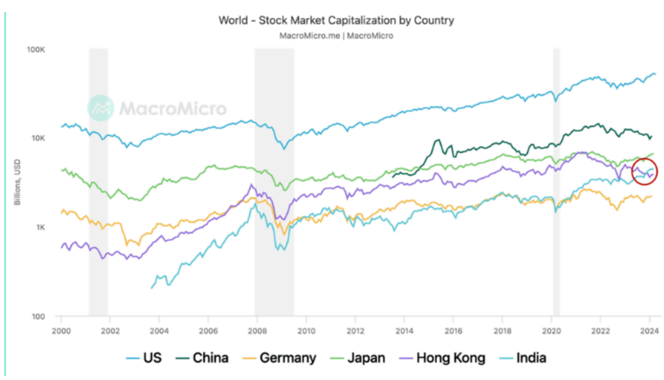


Fig 2.3.3

Global Stock Market Capitalization

Interest Rates:

When interest rates go up, so do borrowing expenses. This can happen with both individual and corporate borrowing. When companies must pay more interest, their profits can go down. And when company profits go down, it can mean a decline in stock prices.

Market Sentiment:

Investors' confidence can have a significant impact on the stock market. Positive investor sentiments can increase stock prices much higher, while on the other hand, negative sentiments can be a death knell for stock prices, causing them to plummet. Regardless of how much investors are willing to invest in the market, one major factor that drives stock prices is their collective sentiment. Sociologists, psychologists, political scientists, and economists have all studied the relationship between an investor's sentiment and the stock market.

Political and regulatory environments:

Government policies, such as tax and trade rule changes, can have a big effect on the price of stocks. But the backdrop of political stability (or lack thereof) can also play a role. As humans, investors feel more comfortable when the regime is stable and the country's direction is clear, which can boost stock prices. By contrast, when political uncertainty reigns or when the regime itself seems wobbly, the price calamity can be profound.

Global Events:

How do global events, such as conflicts and natural disasters, affect the stock market? In recent years, we have witnessed many geopolitical conflicts and tensions in various regions. While some of the events took place in the Middle East and inflicted only local damage, others impacted the global community or threatened to create a serious economic crisis. The Syrian civil war and the war against ISIS, for example, have drawn in significant international players and, in many respects, created a much more dangerous world. Yet those events have done little to shake the world stock markets. On the other hand, natural disasters like the recent slew of hurricanes in the Atlantic, the China and Nepal earthquakes, and the Ebola outbreak in Africa have all had considerable impacts on the stock market.

Market Conditions:

Changes in supply and demand have a direct impact on stock prices. The ease of buying or selling a stock significantly influences its price. Stocks that are easily and cheaply tradable (i.e., with high liquidity) can attract many trades, which makes the stock price potentially more accurate. On the other hand, stocks that are not easily tradable, by virtue of their lower liquidity, can see some unbelievably sizable price changes.

Technological Changes:

Innovation and disruption - "Technological developments and disruptive forces can create a diverging set of new opportunities and threats for companies, which can significantly and viably impact the prices of the stocks they sell to investors."
"Depending on whether the knock-on effects tend to slant more positively or negatively for a given company, an investor should probably think twice about whether or not to take a position within such a company's stock."

Cybersecurity "People who invest in stocks should probably take more note of the various cybersecurity threats and the established methodologies that can serve as a blueprint for potential disruptive forces that can hit companies, positively or negatively."

Chapter III:

METHODOLOGY

3.1 Overview of the Research Problem

The stock market is volatile, complex, and influenced by a variety of factors. Because financial markets are non-linear, it is difficult to predict stock market prices. Unfortunately, the stock market is too complex and non-linear for traditional statistical methods to provide enough insights. With the advancement of machine learning techniques, we can predict the stock market using linear regression, support vector machines, random forests, decision trees, and artificial neural networks. Our research studies can help investors or financial analysts evaluate which machine learning algorithm provides the most accurate predictions. We have conducted research to delve into the association between the Nifty 50 and some of the world's other biggest market indices. Specifically, we're attempting to understand what effect any events in world markets have on the Indian stock market and, more specifically, the opening and closing prices of the Nifty 50.

To make this analysis as insightful as possible, we've looked at the composite nature of the Nifty 50 and its relationship with the other major global indices, among many other variables.

3.2 Operationalization of Theoretical Constructs

Theoretical Constructs:

Stock market prediction accuracy,

Machine learning algorithm performance,

Correlation between Nifty50 and Global Indices.

Stock Market Accuracy:

How accurately do the predicted stock prices correspond with the real stock prices?

Statistical measurements like mean absolute error (MAE), root mean square error (RMSE), and mean absolute percentage error (MAPE) can measure operationalization.

For example, we use a machine learning model to forecast the Nifty 50's or any other stock's closing price over the upcoming 30 days. We then compare that performance by calculating the MAE, RMSE, and MAPE against the real closing prices.

Machine learning algorithm performance:

Various machine learning algorithms demonstrate their efficiency and effectiveness in producing precise predictions.

Operationalization: Performance metrics such as MAE, MSE, RMSE, and R-Squared Error can be used to assess this.

Example: We apply linear regression, support vector machines (SVM), random forests, and neural networks to historical data on any stock price. We then measure their performance using the above performance metrics and compare which algorithm performs best in terms of accurate prediction and effectiveness.

The correlation between the Nifty 50 and global indices is significant.

The Nifty 50 index's movement is significant in comparison to other global stock indices.

Statistical tools like the Pearson correlation coefficient, Spearman rank correlation, and Granger causality tests can quantify operationalization.

To illustrate, imagine that we have gathered historical financial data on stocks. For our research studies, let’s say we have some historical data on the Nifty 50, S&P 500, FTSE 100, and Nikkei 225 indices. We then take these numbers and use the Pearson Correlation Coefficient (PCC) to calculate just how correlated the movements in the Nifty 50 are to these other moves found in the global markets. The PCC can vary from -1.0 to 1.0, where -1.0 signifies a negative correlation and 1.0 signifies a perfect or positive correlation.

Research Scenario:

We engage a financial analyst to predict the closing prices of the Nifty 50 index. The analyst intends to make these predictions by employing a range of machine-learning algorithms. In addition, the analyst is investigating whether global indices affect these algorithms' predictive power.

The steps necessary to implement an operational plan are as follows:

The accuracy of predicting the stock market is something many researchers have attempted to determine.

For the past 20 years, we should have secured historical closing prices for the Nifty 50's index.

Another approach is to use different machine-learning algorithms to make predictions. An algorithm known as linear regression, for instance, attempts to approximate a relationship between input and output using a line. Random Forest, known for its ability to solve a wide range of problems, is another potential algorithm. We can also apply deep learning to achieve better results.

Compute the mean absolute error (MAE), root mean squared error (RMSE), R-squared error, and mean absolute percentage error (MAPE) between the predicted closing prices from each algorithm and the actual closing prices.

Test the performance of machine learning algorithms. Take note of all the accuracy metrics for each algorithm's predictions. Determine the amount of time each algorithm takes to make predictions. Conduct an experiment using historical data to determine the model's level of robustness.

The relationship between the Nifty 50 and global indices is significant. Gather the historical ending prices of the Nifty 50, S&P 500, FTSE 100, Dow Jones, Nasdaq, and Nikkei 225 for the previous 20 years.

Find the Pearson correlation coefficient between the Nifty 50 and every global index.

To understand the effect of global markets on the Nifty 50, analyze the correlation results.

By operationalizing these theoretical constructs, the financial analyst can systemically study the effectiveness of various machine learning algorithms for stock market prediction and the impact of global indices on the Nifty 50.

3.3 Research Purpose and Questions

The primary goal of this research is to compare the effectiveness of various machine learning algorithms in predicting stock prices and to determine which index movement is most accurate. The Nifty 50, an index comprising the top 50 stocks traded on the National Stock Exchange of India, typically guides the performance of India's stock market. So, this study aims to understand if, indeed, the Nifty 50 extends its influence over other stock markets worldwide or is influenced by them, and to map what indexes besides the Nifty 50 best chart the Indian stock market's move.

Research Questions:

How do different machine learning algorithms perform in predicting the closing prices of the Nifty 50 index?

We need to use machine learning algorithms like linear regression (LR), random forests (RF), decision trees (DT), artificial neural networks (ANN), and support vector machines (SVM) on historical stock data to find the answer to this question. Then, we need to compare how well these algorithms predicted the future using measures like MAE, RMSE, MSE, R-squared error, and MAPE.

Which machine learning algorithm provides the best prediction and computational efficiency for stock market prediction?

We will address this question by evaluating not only the prediction accuracy of each algorithm, but also the computational time and resources required to make these predictions. We aim to choose the most feasible algorithm for real-time applications.

What is the correlation between the Nifty 50 index and major global stock indices?

Answering this question will describe the statistical method, such as the Pearson correlation coefficient, and other methods to determine the strength of the relationship between the Nifty 50 and other global indices.

What impact do changes in global stock indices have on the predictive accuracy of machine learning models for the Nifty 50 index?

This will analyse whether the global stock price movement can influence the machine learning algorithms for Nifty 50 prediction.

Can a hybrid machine learning model combining multiple algorithms improve the accuracy of Nifty 50 predictions compared to individual models?

We will investigate these questions by developing and testing hybrid models that integrate predictions from multiple algorithms to enhance overall prediction performance.

3.4 Research Design

The research design for this study involves a quantitative approach, using historical financial data and machine learning techniques to figure out and compare the performance of various algorithms. The design also includes statistical analysis to explore the correlation between the Nifty 50 and global stock indices. We structured the study into three main phases: data collection, algorithm implementation and evaluation, and correlation analysis.

Phase 1: Data Collection

Objective: Collect financial historical data for the Nifty 50 index and major global indices (e.g., Dow Jones, Nasdaq, S&P 500, FTSE 100, Nikkei 225).

Data Sources: The public can access financial databases such as Yahoo Finance and other available documents.

Example: Collect daily stock prices (low, open, high, close, adjacent close, volume) for the past 20 years for all global indices, along with the Nifty 50 index.

Phase 2: Algorithm Implementation and Evaluation

Objective: Implement and evaluate various machine learning algorithms for predicting the Nifty 50 index's closing prices.

We will test the following algorithms:

Linear Regression (LR), Support Vector Machines (SVM), Random Forests (RF), Artificial Neural Networks (ANN), and hybrid models that combine multiple algorithms.

Evaluation Metrics: Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE), R-Squared Error, computational time, and resource usage.

Example:

Data Preprocessing: Normalize the data, handle missing values, and generate training and test data.

Model Training: Train each machine learning model using the training dataset.

Model Testing: Test each machine learning model on the test dataset, recording prediction errors and computational metrics.

Comparison: Compare the models based on the evaluation metrics to identify the most accurate and efficient algorithm.

Phase 3: Correlation Analysis

The goal of correlation analysis is to examine the relationship between the Nifty50 index and global stock indices.

Statistical tools: Pearman Correlation Coefficient, Spearman’s Rank Correlation, Granger Causality Tests.

Example:

Correlation Calculation: Calculate the Pearson Correlation Coefficient between Nifty50 and each global index's daily closing prices.

Perform the Granger Causality Test (GCT) to determine if movements in global stock prices can predict changes in the Nifty 50.

Analyse and interpret the results to understand the strength and direction of the relationship between the indices.

Research Design Steps and Example:

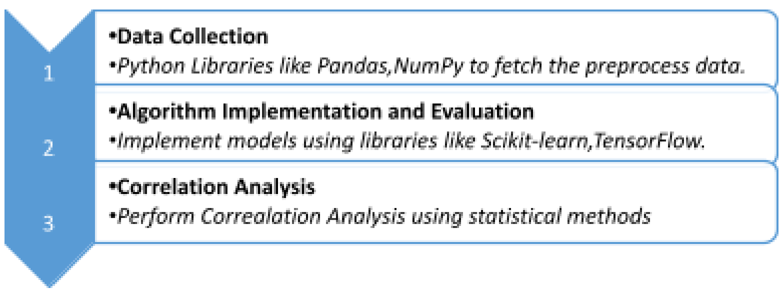


Figure: 3.4.1

Research Design Steps

3.5 Population and Sample

Population: For this research study, the entire historical stock market data for the Nifty 50 index, as well as other major global indices such as Dow Jones, Nasdaq, FTSE, Nikkei 225, and S&P 500, is used. This includes daily stock prices, traded volumes, and adjacent closings.

The population's characteristics include:

The Nifty50 Index represents the performance of the 50 largest companies listed on India's National Stock Exchange (NSE).

Global Indices: The S&P 500 represents the 50 largest stocks listed on US stock exchanges.

The FTSE 100 represents the 100 largest companies listed on the London Stock Exchange.

Nikkei 225: The Tokyo Stock Exchange's stock market index.

Time Frame: The study will span a twenty-year period, with today as the end date and today as the start date.

Data Collections: Using Python libraries and Yahoo Finance data, we have collected the Nifty 50 data for our first part of our research studies.

Date	Open	High	Low	Close	Adj Close	Volume
2007-09-17	4518.450195	4549.049805	4482.850098	4494.649902	4494.649902	0
2007-09-18	4494.100098	4551.799805	4481.549805	4546.200195	4546.200195	0
2007-09-19	4550.250000	4739.000000	4550.250000	4732.350098	4732.350098	0
2007-09-20	4734.850098	4760.850098	4721.149902	4747.549805	4747.549805	0
2007-09-21	4752.950195	4855.700195	4733.700195	4837.549805	4837.549805	0
...	...	...	...	...	...	...
2024-05-17	22415.250000	22502.150391	22345.650391	22466.099609	22466.099609	242700
2024-05-21	22404.550781	22591.099609	22404.550781	22529.050781	22529.050781	347600
2024-05-22	22576.599609	22629.500000	22483.150391	22597.800781	22597.800781	290300
2024-05-23	22614.099609	22993.599609	22577.449219	22967.650391	22967.650391	369800
2024-05-24	22930.750000	23026.400391	22908.000000	22957.099609	22957.099609	261900

4088 rows × 6 columns

Figure 3.5.1

Collections of Nifty50 index historical financial data (Total Records :4088): Phase 1

Date	Nifty50_pct_change	Nasdaq_pct_change	DowJones_pct_change	SandP_pct_change	FTSE_pct_change	crude_oil_pct_change	gold_data_pct_change
2004-05-25	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2004-05-26	NaN	0.005853	-0.000764	0.001698	0.017255	0.012771	NaN
2004-05-27	NaN	0.004225	0.009427	0.005686	0.001250	-0.006391	NaN
2004-05-28	NaN	0.001129	-0.001641	-0.000535	0.012855	-0.005563	NaN
2004-05-31	NaN	NaN	NaN	NaN	-0.006472	NaN	NaN
...	...	...	...	...	...	...	...
2024-05-20	NaN	0.006526	-0.004920	0.000916	0.007278	0.002762	NaN
2024-05-21	0.002802	0.002248	0.001664	0.002502	-0.003142	0.008264	NaN
2024-05-22	0.003052	-0.001846	-0.005065	-0.002706	-0.008469	-0.004308	NaN
2024-05-23	0.016367	-0.003899	-0.015270	-0.007381	0.012588	-0.010870	NaN
2024-05-24	-0.000459	0.011040	0.000111	0.007001	-0.011690	-0.004588	NaN

5206 rows × 7 columns

Fig 3.5.2:

Collections of historical financial data (Total Records :5206): Phase 2

Date	Nifty50_pct_change	Nasdaq_pct_change	DowJones_pct_change	SandP_pct_change	FTSE_pct_change	crude_oil_pct_change	gold_data_pct_change
2024-04-25	0.007497	-0.006427	-0.009753	-0.004576	-0.021622	0.004726	NaN
2024-04-26	-0.006664	0.020250	0.004040	0.010209	0.008140	0.019263	NaN
2024-04-29	0.009967	0.003464	0.003829	0.003178	NaN	-0.005714	NaN
2024-04-30	-0.001703	-0.020350	-0.014854	-0.015731	0.012413	0.015803	NaN
2024-05-01	NaN	-0.003343	0.002310	-0.003435	-0.003427	-0.000218	NaN
2024-05-02	0.001918	0.015090	0.008505	0.009128	-0.000992	0.010990	NaN
2024-05-03	-0.007610	0.019909	0.011773	0.012557	NaN	-0.000323	NaN
2024-05-06	-0.001475	0.011941	0.004566	0.010326	NaN	0.003230	NaN
2024-05-07	-0.006247	-0.001021	0.000823	0.001343	0.015667	0.009659	NaN
2024-05-08	0.000000	-0.001825	0.004427	-0.000006	-0.016293	-0.006484	NaN
2024-05-09	-0.015469	0.002668	0.008484	0.005091	-0.003361	0.006526	NaN
2024-05-10	0.004449	-0.000330	0.003176	0.001649	0.004074	0.010842	NaN
2024-05-13	0.002215	0.002899	-0.002058	-0.000241	-0.001299	-0.006730	NaN
2024-05-14	0.005148	0.007502	0.003211	0.004838	0.004625	0.001165	NaN
2024-05-15	-0.000779	0.014003	0.008845	0.011716	0.000774	-0.000423	NaN
2024-05-16	0.009157	-0.002632	-0.000968	-0.002082	0.013925	0.000000	NaN
2024-05-17	0.002779	-0.000740	0.003366	0.001165	-0.003414	-0.004231	NaN
2024-05-20	NaN	0.006526	-0.004920	0.000916	0.007278	0.002762	NaN
2024-05-21	0.002802	0.002248	0.001664	0.002502	-0.003142	0.008264	NaN
2024-05-22	0.003052	-0.001846	-0.005065	-0.002706	-0.008469	-0.004308	NaN

Figure 3.5.3

The 20-sample data records of Phase 2 Research studies

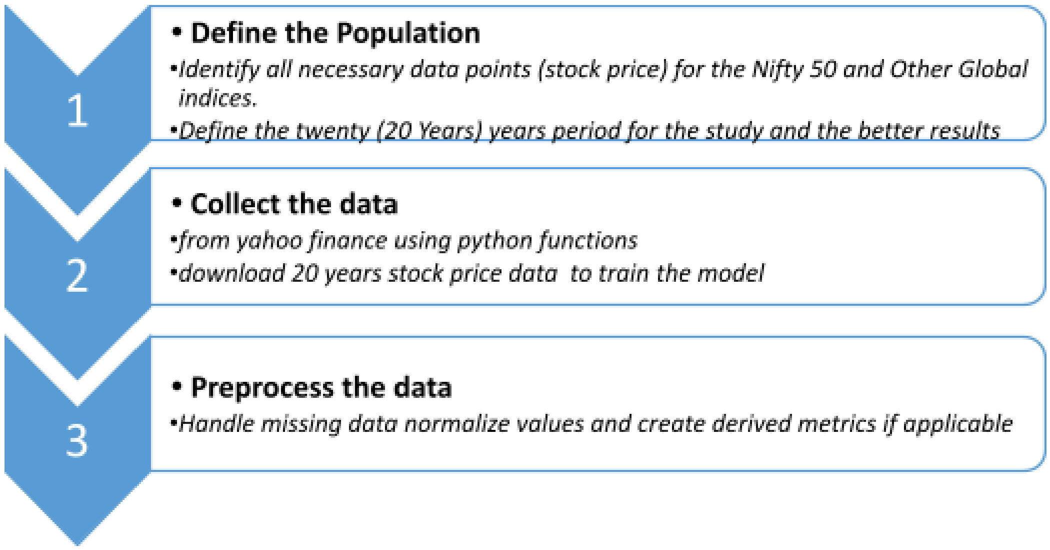


Figure 3.5.4

Steps to collect and preprocess the dataset

3.6 Dataset and datapoint Selection

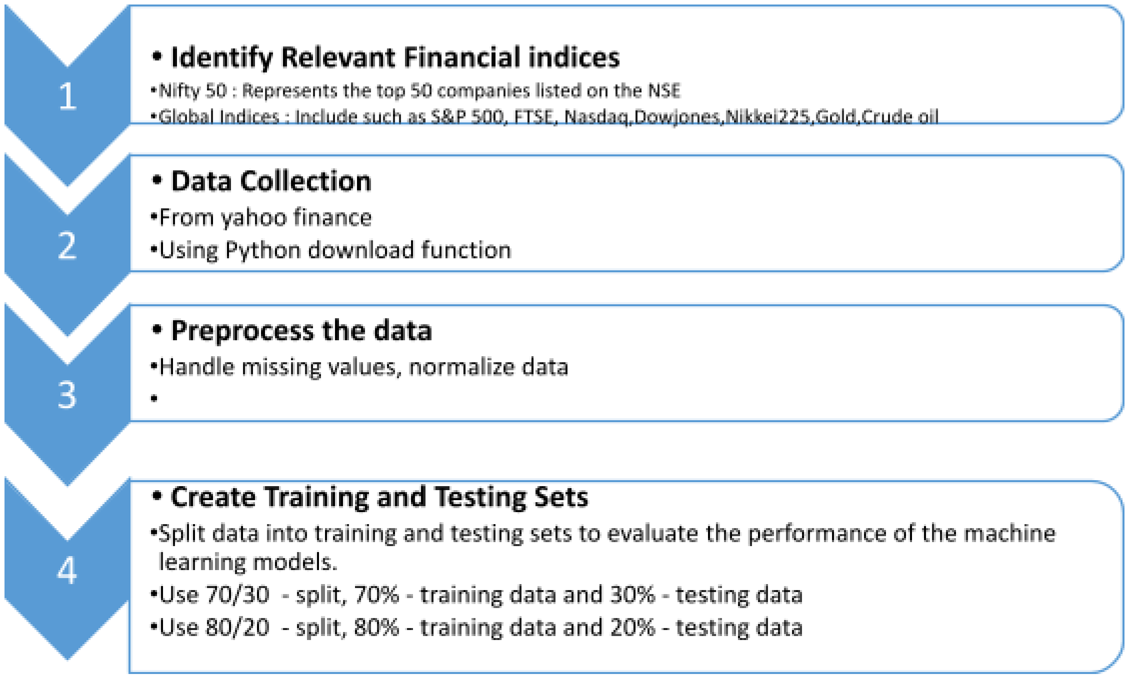


Figure 3.6.1

Dataset and Datapoint Selection

3.7 Instrumentation

Instrumentation includes the tools, software, and techniques used to collect, process, and analyse the data for this study. As our study is quantitative in nature, the instrumentation will primarily involve data collection methods, data preprocessing techniques, and the machine learning frameworks and libraries used to develop and predict models.

Table 3.7.1

Tools and Techniques

Tools and Techniques	Functionalities
Data Collection Tools	Pandas Data Reader Yahoo Finance API
Coding Platform	Google Colab (for python coding)
Data Frames	Pandas Data Frame

Data Preprocessing	Missing Values Handling Normalization and Scaling Feature Engineering
Machine Learning Frameworks	Scikit-learn TensorFlow / Keras XGBoost Correlation Models
Evaluation Metrics and Tools	Metrics: MAE, MSE, RMSE, MAPE Visualization: Matplotlib, Seaborn etc.

3.8 Data Collection Procedures

Phase 1: In our research studies from Yahoo Finance (2024), we are downloading the stock data from Yahoo Finance using a Python inbuilt function for the last 20 years for an individual stock or index.

Phase 2: Here we are collecting the financial historical data for the Nifty50, Other Global Indices, Crude Oil, and Gold

```python
end = datetime.now()
start = datetime(end.year - 20, end.month, end.day)
nasdaq_ind = '^IXIC'   # to download the global stock indices data of twenty years
dowjones_ind = '^DJI'
nikkei_ind  = '^N225'
sandp_ind =  '^GSPC'
ftse_ind = '^FTSE'
nifty_ind = '^NSEI'
crude_ind = 'CL'
gold_ind  = 'GOLD'

nasdaq_data = yf.download(nasdaq_ind,start,end)
dowjones_data = yf.download(dowjones_ind,start,end)
nikkei_data = yf.download(nikkei_ind,start,end)
sandp_data =  yf.download(sandp_ind,start,end)
ftse_data = yf.download(nikkei_ind,start,end)
nifty_data = yf.download(nifty_ind,start,end)
crude_oil_data = yf.download(crude_ind,start,end)
gold_data      = yf.download(gold_ind,start,end)
```

```
[**********************100%%**********************]  1 of 1 completed
[**********************100%%**********************]  1 of 1 completed
[**********************100%%**********************]  1 of 1 completed
[**********************100%%**********************]  1 of 1 completed
[**********************100%%**********************]  1 of 1 completed
[**********************100%%**********************]  1 of 1 completed
[**********************100%%**********************]  1 of 1 completed
[**********************100%%**********************]  1 of 1 completed
```

.

Figure 3.8.1

Python Code to download Historical Global indices data along with Gold and Crude oil

3.9 Data Analysis

Exploratory Data Analysis: Descriptive Statistics:

In the first part of our model, we can download any stock's historical financial data from Yahoo Finance through Python coding of the last twenty years data from today's date. Here, we will discuss descriptive statistics as well as visualization of key features such as closing prices, trading volumes, and volatility over time. The following are the descriptive statistics for the Nifty50 index.

	Open	High	Low	Close	Adj Close	Volume
count	4088.000000	4088.000000	4088.000000	4088.000000	4088.000000	4.088000e+03
mean	9593.855798	9646.588259	9526.084362	9587.880397	9587.880397	2.015439e+05
std	4917.710809	4929.524311	4898.209526	4915.266716	4915.266716	2.100301e+05
min	2553.600098	2585.300049	2252.750000	2524.199951	2524.199951	0.000000e+00
25%	5541.550171	5578.874878	5490.625122	5540.987549	5540.987549	0.000000e+00
50%	8324.949707	8358.549805	8259.200195	8322.600098	8322.600098	1.736000e+05
75%	11681.687500	11750.337158	11625.662109	11678.962402	11678.962402	2.805000e+05
max	22930.750000	23026.400391	22908.000000	22967.650391	22967.650391	1.811000e+06

Figure 3.9.1

Descriptive statistical information of Historical Stock data: Phase 1

According to the above information, the total number of data records is 4088. Mean values for open, high, low, and close Adjacent The close symbol represents average price levels, and the mean volume symbol represents average trading activities. From the standard deviation, we can get an insight into market volatility. The market becomes more volatile as the standard deviation increases.

Trends and Patterns:

After plotting the values of Open, High, Low, Close, and Adj Close over time, we can gain insight into increasing or decreasing trends, cyclical patterns, or seasonality. With these patterns, informed decision-making is much easier and more positive for investors or financial analysts. Anyone in the stock market can identify a potential entry or existing point. The Nifty50 index patterns are listed below.

Figure: 3.9.2

Stock Prices (open, high, low, close, adjacent close)

Volatility Analysis:

The importance of volatility analysis can help manage risk, make decisions, or set trading strategies. Also, volatility analysis allows one to diversify their portfolios and gain a better understanding of market sentiment. High volatility often correlates with market uncertainty or significant events that affect investor sentiment.

Correlation Analysis:

After analysing correlations between open, high, close, and low prices in our research studies, we can see that strong positive correlations between open, high, close, and low indicate strong relationships

Table: 3.9.1

Correlation (Open, High, Low, Close, Adjacent Close)

Variable	Correlation Coefficient	Interpretation
Open vs. High	0.999927	Near-perfect positive correlation; Open price increases with High price.

Open vs. Low	0.999896	Near-perfect positive correlation; Open price increases with Low price.
Open vs. Close	0.999827	Near-perfect positive correlation; Open price increases with Close price.
Open vs. Adj Close	0.999827	Near-perfect positive correlation; Open price

		increases with Adjacent Close price.
Open vs. Volume	0.547849	Moderate positive correlation; Open price changes with Volume.
High vs. Low	0.99987	Near-perfect positive correlation; High price increases with Low price.
High vs. Close	0.99992	Near-perfect positive

		correlation; High price increases with Close price.
High vs. Adj Close	0.99992	Near-perfect positive correlation; High price increases with Adj Close price.
High vs. Volume	0.549206	Moderate positive correlation; High price changes with Volume.

Low vs. Close	0.999917	Near-perfect positive correlation; Low price increases with Close price.
Low vs. Adj Close	0.999917	Near-perfect positive correlation; Low price increases with Adj Close price.
Low vs. Volume	0.545433	Moderate positive correlation; Low price changes with Volume.

Close vs. Adj Close	1	Perfect positive correlation: Close price moves identically with Adj Close price.
Close vs. Volume	0.547279	Moderate positive correlation; Close price changes with Volume.
Adj Close vs. Volume	0.547279	Moderate positive correlation; Adj Close price

changes with Volume.

Volume Analysis:

The volume data reflects market trading activity. Higher volumes usually accompany significant price movements, indicating increased investor interest or participation. Analysing volume patterns can help to confirm price trends, identify potential reversals, and gauge market sentiment.

Seasonality Analysis:

Analysing seasonality data allows one to establish an investment strategy for a specific time of year.

Stationarity Check:

Checking for stationarity in the data can be important for time series analysis. This involves examining the mean, variance, and autocorrelation over time.

By analysing and examining these aspects of stock data, one can establish a strategy for selecting stocks and determining when to invest and trade.

Feature Engineering:

We have created additional features from the downloaded data set, such as moving averages (50 days, 100 days, and 250 days), volatility measures, and technical indicators.

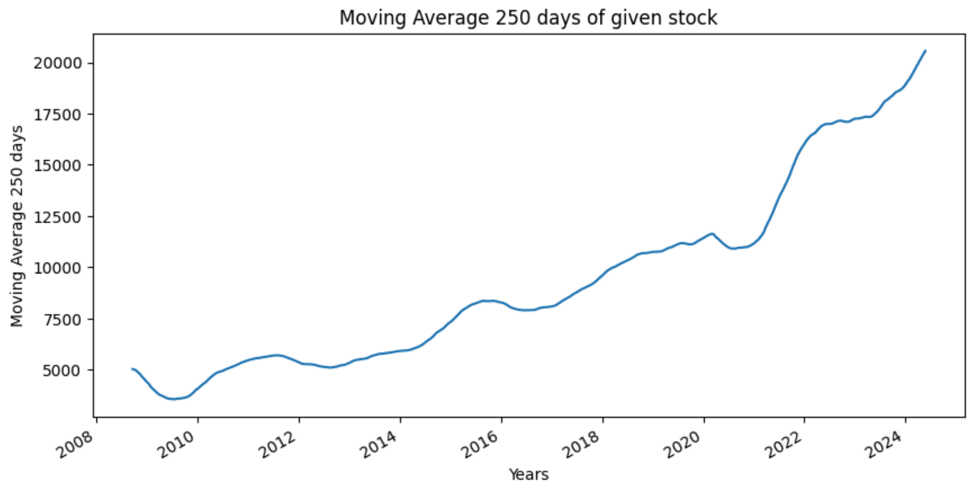


Figure 3.9.3

250 days Moving Average of the Nifty50 index

The 250-day moving average is calculated by taking the average closing price of an asset (such as the Nifty50 index) over the past 250 trading days. This moving average is updated daily, with each new day's closing price replacing the oldest closing price in the calculation.

For the second part of our research, we have collected historical data for the Nifty50, as well as other global indices, along with data for gold and crude oil. Due to different currencies, we have taken percentage change of the historical data's stock price value. From the studies conducted by GeeksforGeeks (2021), the Pandas pct_change () method is applied to a series of our stock dataset, and the calculated results are derived based on the formula ((Current - Previous) / Previous) * 100. After applying the method, we obtained the below results (subset of the population stock data).

Date	Nifty50_pct_change	Nasdaq_pct_change	DowJones_pct_change	SandP_pct_change	FTSE_pct_change	crude_oil_pct_change	gold_data_pct_change
2024-05-06	-0.001475	0.011941	0.004566	0.010326	NaN	0.003230	0.012758
2024-05-07	-0.006247	-0.001021	0.000823	0.001343	0.015667	0.009659	0.004199
2024-05-08	0.000000	-0.001825	0.004427	-0.000006	-0.016293	-0.006484	-0.001195
2024-05-09	-0.015469	0.002668	0.008484	0.005091	-0.003361	0.006526	0.018541
2024-05-10	0.004449	-0.000330	0.003176	0.001649	0.004074	0.010842	-0.005872
2024-05-13	0.002215	0.002899	-0.002058	-0.000241	-0.001299	-0.006730	0.000591
2024-05-14	0.005148	0.007502	0.003211	0.004838	0.004625	0.001165	0.012397
2024-05-15	-0.000779	0.014003	0.008845	0.011716	0.000774	-0.000423	0.015743
2024-05-16	0.009157	-0.002632	-0.000968	-0.002082	0.013925	0.000000	0.005741
2024-05-17	0.002779	-0.000740	0.003366	0.001165	-0.003414	-0.004231	0.021119
2024-05-20	NaN	0.006526	-0.004920	0.000916	0.007278	0.002762	0.001677
2024-05-21	0.002802	0.002248	0.001664	0.002502	-0.003142	0.008264	-0.006696
2024-05-22	0.003052	-0.001846	-0.005065	-0.002706	-0.008469	-0.004308	-0.028652
2024-05-23	0.016367	-0.003899	-0.015270	-0.007381	0.012588	-0.010870	-0.020243
2024-05-24	-0.000459	0.011040	0.000111	0.007001	-0.011690	-0.004588	0.004723

Figure 3.9.4

Calculated Percentage Change of the Nifty50 and other Global indices along with Gold and Crude Oil

Descriptive statistics of the second part of our studies, we can have a better understanding of Statistical Information. Count represents the total number of data records for each index. Mean represents the average percentage change of each variable. Standard Variation represents the percentage changes of each variable.

	Nifty50_pct_change	Nasdaq_pct_change	DowJones_pct_change	SandP_pct_change	FTSE_pct_change	crude_oil_pct_change	gold_data_pct_change
count	4087.000000	5033.000000	5033.000000	5033.000000	4894.000000	5033.000000	5033.000000
mean	0.000489	0.000519	0.000334	0.000383	0.000354	0.000397	0.000363
std	0.013418	0.013538	0.011398	0.012041	0.014155	0.012038	0.026206
min	-0.129805	-0.123213	-0.129265	-0.119841	-0.114064	-0.112277	-0.156997
25%	-0.005438	-0.005281	-0.004142	-0.004100	-0.006417	-0.005364	-0.013158
50%	0.000715	0.001009	0.000563	0.000702	0.000647	0.000324	0.000000
75%	0.006741	0.007153	0.005345	0.005693	0.007639	0.006433	0.013458
max	0.177441	0.118059	0.113650	0.115800	0.141503	0.126083	0.313106

Figure 3.9.5

Descriptive statistical information of Nifty 50 and other global indices along with Gold and Crude Oil percentage change

For a better insight of correlation between Nifty50 and other indices below is the Pearson correlation coefficients.

	Nifty50_pct_change	Nasdaq_pct_change	DowJones_pct_change	SandP_pct_change	FTSE_pct_change	crude_oil_pct_change	gold_data_pct_change
Nifty50_pct_change	1.000000	0.259397	0.300472	0.298020	0.362035	0.166103	0.074543
Nasdaq_pct_change	0.259397	1.000000	0.887623	0.950356	0.136578	0.466567	0.187233
DowJones_pct_change	0.300472	0.887623	1.000000	0.971578	0.160219	0.572448	0.184972
SandP_pct_change	0.298020	0.950356	0.971578	1.000000	0.154580	0.549682	0.210364
FTSE_pct_change	0.362035	0.136578	0.160219	0.154580	1.000000	0.071949	0.047351
crude_oil_pct_change	0.166103	0.466567	0.572448	0.549682	0.071949	1.000000	0.108062
gold_data_pct_change	0.074543	0.187233	0.184972	0.210364	0.047351	0.108062	1.000000

Figure 3.9.6

Pearson Correlation of Nifty 50 and other global indices along with gold and crude oil

Each cell in the above table represents the correlation coefficient between the percentage changes of two variables. For example, the correlation coefficient between Nifty50_pct_change and Nasdaq_pct_change is 0.259397, indicating a moderate positive correlation between the percentage changes of these two indices. Similarly, the other correlation coefficients provide insights into the relationships between the percentage changes of different financial indicators.

Spearman Correlation Coefficient:

	Nifty50_pct_change	Nasdaq_pct_change	DowJones_pct_change	SandP_pct_change	FTSE_pct_change	crude_oil_pct_change	gold_data_pct_change
Nifty50_pct_change	1.000000	0.205478	0.220313	0.224456	0.328974	0.103567	0.047486
Nasdaq_pct_change	0.205478	1.000000	0.821106	0.925515	0.126498	0.383338	0.183842
DowJones_pct_change	0.220313	0.821106	1.000000	0.943773	0.128448	0.507771	0.169123
SandP_pct_change	0.224456	0.925515	0.943773	1.000000	0.125973	0.476500	0.199970
FTSE_pct_change	0.328974	0.126498	0.128448	0.125973	1.000000	0.040032	0.024265
crude_oil_pct_change	0.103567	0.383338	0.507771	0.476500	0.040032	1.000000	0.091935
gold_data_pct_change	0.047486	0.183842	0.169123	0.199970	0.024265	0.091935	1.000000

Figure 3.9.7

Spearman Correlation of Nifty 50 and other global indices along with gold and crude oil

High Correlation: The strongest correlations are observed among U.S. stock indices (Nasdaq, Dow Jones, S&P), indicating that these indices tend to move in similar directions.

Moderate Correlation: Nifty50 has a moderate positive correlation with FTSE, and crude oil shows moderate correlations with U.S. indices.

Weak Correlation: Gold shows weak correlations with all indices and commodities, suggesting that its price changes are less influenced by the changes in these financial markets.

These correlations help in understanding the relationships between different financial markets and can be useful for diversification and risk management strategies.

3.10 Research Design Limitations

Table:3.10.1

Research Design Limitations

Aspect	Limitations	Explanation
Data Quality	Incomplete or Inaccurate Data	Historical stock data might have missing values or errors, affecting the analysis results.

Time Frame	Limited Time Frame	The study might be limited to a specific period, which may not capture long-term trends or cyclical patterns.
Market Conditions	Varying Market Conditions	Different market conditions (e.g., bull or bear markets) can influence the results and their generalizability.

External Factors	Influence of External Factors	Events such as economic crises, political changes, or natural disasters can impact market behaviors unpredictably.
Model Assumptions	Simplistic Assumptions in Models	Models might rely on assumptions (e.g., linear relationships, normal distribution) that may not hold.
Data Frequency	Use of Daily Data Only	Daily data may not capture

		intraday variations and might miss high-frequency trading patterns.
Sample Size	Limited Sample Size	A small sample size might not be representative of the entire market or of different market conditions.
Generalizability	Limited Generalizability	Results based on specific indices or commodities might not apply to other

		financial instruments or markets.
Correlation vs. Causation	Misinterpretation of Correlation	High correlation does not imply causation; other unobserved factors might drive the relationships.
Technological Changes	Impact of Technological Changes	Technological advancements in trading and data processing can affect market dynamics and historical data relevance.

Regulatory Changes	Influence of Regulatory Changes	Changes in regulations over the study period can impact market behavior and thus affect the analysis.
Market Liquidity	Variations in Market Liquidity	Different levels of market liquidity can influence price changes and volatility, affecting the study's findings.

Economic Cycles	Not Accounting for Economic Cycles	Economic cycles (e.g., recession, expansion) might affect stock market behavior differently.
Behavioral Factors	Ignoring Investor Behavior	Psychological factors and irrational behaviors of investors are not always captured in quantitative models.

3.11 Conclusion

In this section, we discuss how we collected data from Yahoo Finance through Python functions. It will also provide a comprehensive look at our data analysis—after we received the data, we did a bit of number crunching to try to determine the average stock price values (open, high, low, close, and volume). We also performed some statistical analysis, used a mean to analyze the stock's price, and determined the 25th, 50th, and 75th percentile values of our stock price data set. On top of that, we calculated the stock's standard deviation as a way to describe its volatility Smith(2020).

We have a strong and reliable foundation for our stock market prediction and correlation analysis. The main component of our design involves collecting relevant historical financial data from the stock market. We then use this data in the model development and evaluation process. Our data analysis phase is an investigation into the market itself. By using some conventional techniques like exploratory data analysis and well-known theoretical frameworks like the Efficient Market Hypothesis, we gain insight into market dynamics and relationships between various kinds of indices.

Despite the structured approach, several limitations need consideration. Challenges related to data quality and availability, model overfitting, market volatility, and model interpretability. To address these limitations, researchers must employ rigorous validation techniques, mitigate overfitting through model regularization, adapt models to evolving market conditions, and explore interpretable machine learning techniques. Brown & Taylor (2022).

In conclusion, the research design provides a solid foundation for investigating the application of machine learning algorithms in stock market prediction and correlation analysis. However, addressing the identified problems is the only way to enhance the study's robustness and usefulness. By navigating these challenges and leveraging advanced methodologies and techniques, researchers can contribute to a deeper understanding of financial markets and facilitate more informed decision-making processes in investment and trading practices Wilson(2023).

CHAPTER IV:

RESULTS

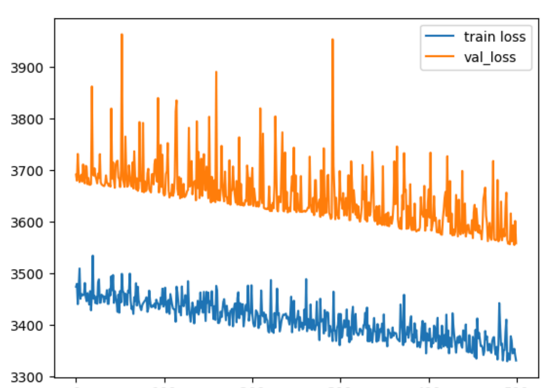
4.1 Research Question One

How do different machine learning algorithms perform in predicting the closing prices of the Nifty 50 index?

Our research studies are twofold. The first is to evaluate the best machine learning algorithms for stock market prediction. In our results section, we will discuss the machine learning algorithms we have used. Using performance metrics, we will conclude which algorithm is the best.

When constructing our machine learning models, we made use of several Python libraries, such as NumPy, Pandas, Matplotlib, Seaborn, scikit-learn, and yfinance. These are the libraries that help us perform the required tasks. And more importantly, with the help of these libraries, we constructed models (such as artificial neural networks, random forests, decision trees, support vector machines, logistic regression, and XGBoost) that solved our problem set. These algorithms have performed well in many domains, so they are suitable for our domain. To evaluate these algorithms, we needed performance metrics like mean absolute error, mean squared error, root mean squared error, and R-squared error.

. Actual Vs Predictions : Artificial Neural Networks



Train loss (blue) represents the training loss over epochs (one complete pass through the entire training dataset). Loss refers to the loss calculated during the training and test datasets.

The loss calculated during validation on a separate.

Figure 4.1.1

Training Loss Vs Validation Loss

The loss calculated during validation on a separate Here are some key observations and analyses of the plot:

The loss from training is displayed in blue, with a lower value indicating better performance. Loss indicates how well the model has learned the data, and this is good: it is going down very fast. Have you noticed that the plot appears slightly uneven? Cats do not transform into dogs and vice versa in a single image.

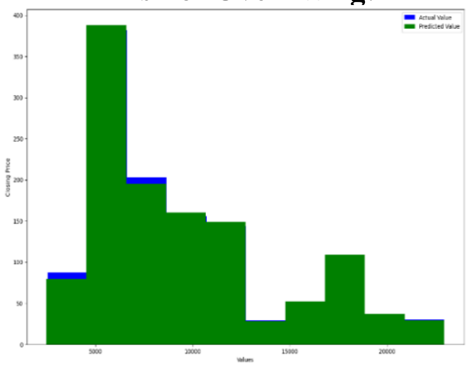
Validation Loss:

Orange line Decreasing, but consistently greater than training loss. Noisy, with large fluctuations.

Analyzing Learning Dynamics:

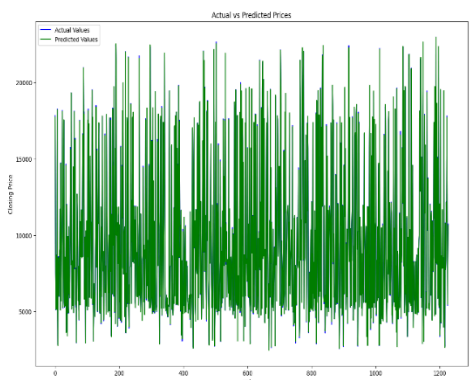
At least, the validation loss indicates that we are generalizing to unseen data, which is also a positive development. But it's pretty noisy and higher than the training loss, suggesting some potential issues.

Risk of Overfitting:



een the training and validation losses indicates that training data. When a model meticulously learns e training data, it often results in poor generalizations the case with a deep neural network. When the overfit.

Noise and Stability: Both the training and validation losses include noise.



This Fig 4.1.2 pattern represents the actual(blue) Vs prediction value(green) of closing price, where x-axis represents the values and y-axis represents the closing price of the stock.

Figure 4.1.2

Actual Vs Predictions of closing price : Model ANN

This Fig 4.1.3 patterns is the histogram comparing actual values (blue) and predicted values (green) of stock price using Artificial Neural Networks.Here x-axis represents values whereas y-axis represents closing price of the stock price.

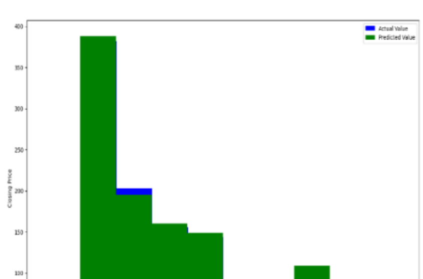
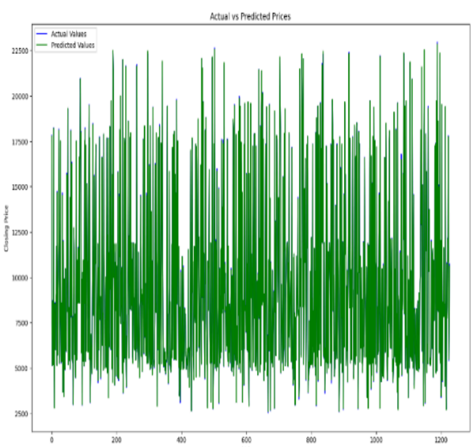


Figure 4.1.3

Histograms comparing actual and predicted values : Model ANN

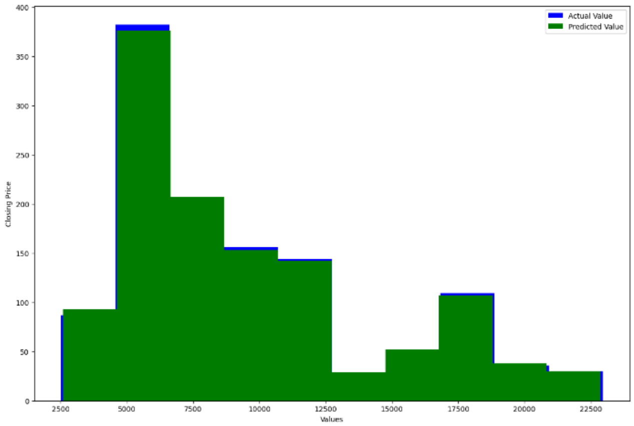
Actual Vs Predictions : Random Forest Regressor



This figure 4.1.4 is representing a pattern of actual(blue) and predicted values(green) of stock price using Random Forest Regressor model.In this plot x-axis represents values whereas y-axis describes the close price of the stock.

Fig : 4.1.4

Actual Vs Predictions of closing price : Model: Randm Forests Regressor



This Fig 4.1.5 patterns is the histogram comparing actual values (blue) and predicted values (green) of stock price using Random Forest Regressor.Here x-axis represents values whereas y-axis represents closing price of the stock price.

Figure 4.1.5

Histograms comparing actual and predicted values : Model RF

Decision Tree Regressor :

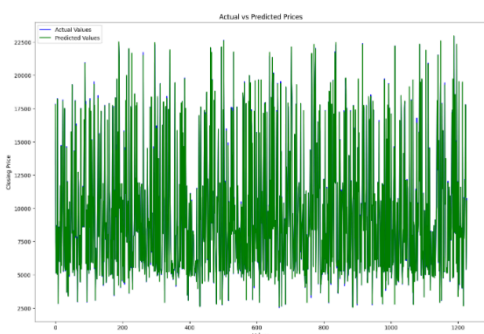


Figure 4.1.6 depicts a pattern of actual(blue) and predicted values(green) of stock price using Decision Tree Regressor model.In this plot x-axis represents values whereas y-axis describes the close price of the stock.

Figure 4.1.6

Actual Vs Predictions of closing price : Model: Decision Tree Regressor

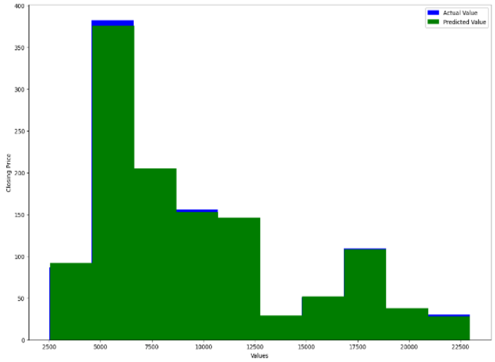


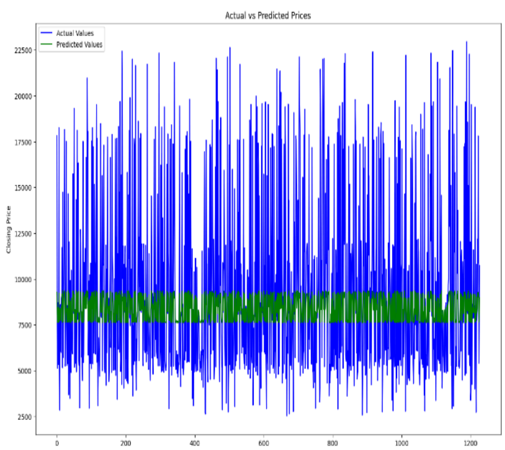
Fig 4.1.7 represents histogram comparing actual values (blue) and predicted values (green) of stock price using Decision Tree Regressor.Here x-axis represents values whereas y-axis represents closing price of the stock price.

Figure 4.1.7

Histograms comparing actual and predicted values : Model Decision Tree Regressor

Support Vector Machine :

Figure 4.1.8 depicts a pattern of actual(blue) and predicted values(green) of stock price



using Support Vestor Machine model.In this plot x-axis represents values whereas y-axis describes the close price of the stock.

Figure 4.1.8

Actual Vs Predictions of closing price : Model: Support Vector Machine

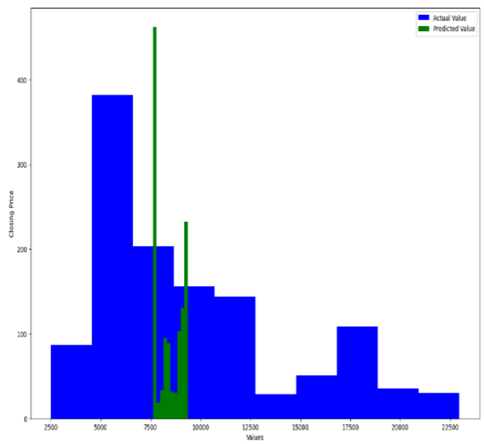
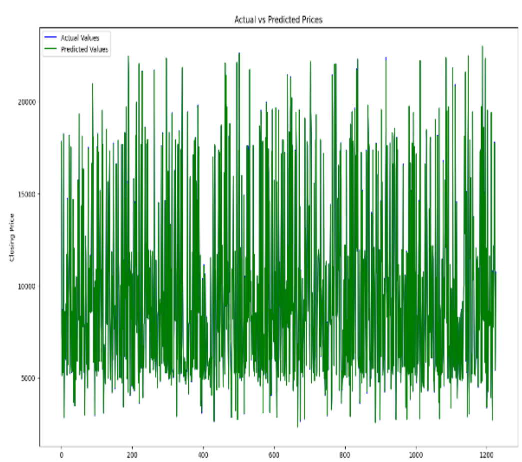


Fig 4.1.9 represents histogram comparing actual values (blue) and predicted values (green) of stock price using Support Vector Machine.Here x-axis represents values whereas y-axis represents closing price of the stock price.

Figure 4.1.9

Histograms comparing actual and predicted values : Model Support Vector Machine

Linear Regression:



Figuure 4.1.10 depicts a pattern of actual(blue) and predicted values(green) of stock price using Linear Regression model.In this plot x-axis represents values whereas y-axis describes the close price of the stock.

Figure 4.1.10

Actual Vs Predictions of closing price : Model: Linear Regression

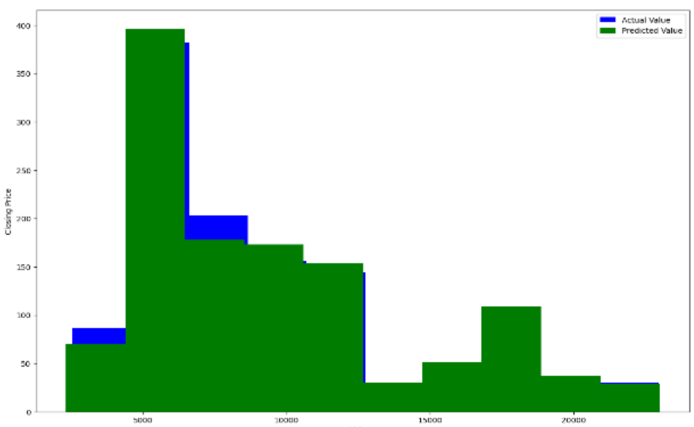


Fig 4.1.11 represents histogram comparing actual values (blue) and predicted values (green) of stock price using Linear Regression.Here x-axis represents values whereas y-axis represents closing price of the stock price.

Figure 4.1.11

Histograms comparing actual and predicted values : Model Linear Regression

Extreme Gradient Boosting (XGBOOST) Algorithm:

Fig 4.1.12 depicts a pattern of actual(blue) and predicted values(green) of stock price using Linear Regression model.In this plot x-axis represents values whereas y-axis describes the close price of the stock.

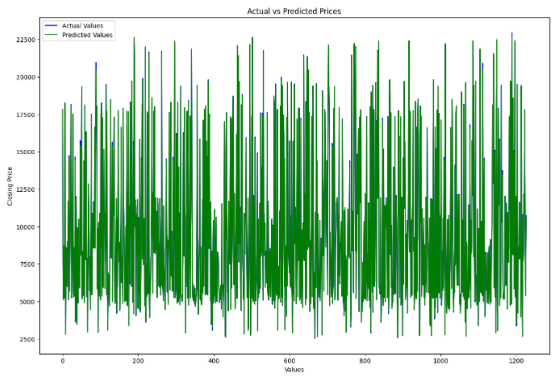


Figure 4.1.12

Actual Vs Predictions of closing price

Model XGBOOST

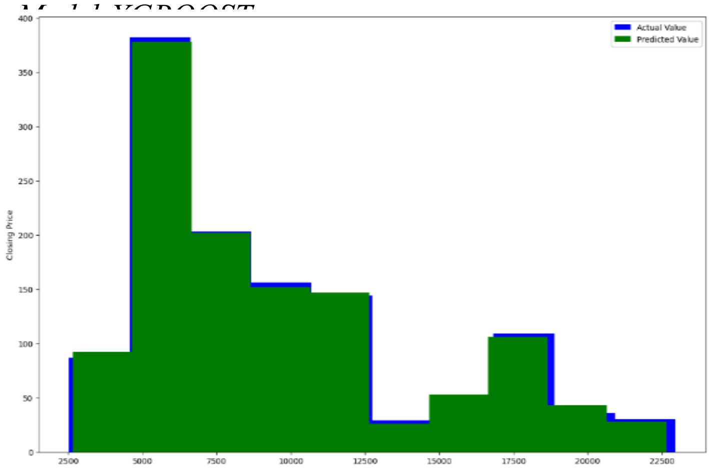


Fig 4.1.13 represents histogram comparing actual values (blue) and predicted values (green) of stock price using Linear Regression. Here x-axis represents values whereas y-axis represents closing price of the stock price.

Figure 4.1.13

Histograms comparing actual and predicted values : Model XGBOOST

From both the histograms and actual vs predicted graphical plot, and from this pattern we can predict for unknown closing price of a stock.

4.2 Research Question Two

Which Machine Learning algorithm offers the best from both prediction as well as computational efficiency for stock market prediction?

Model	r2Score	MAE	RMSE	MAPE	r2Score Rank	MAE Rank	RMSE Rank	MAPE Rank	Average Rank
ANN	0.999844409	43.3579866	60.56164273	0.53322373	3	4	2	5	3.5
RF	0.999883323	34.6652834	52.44425888	0.428959776	2	2	1	2	1.75
DT	0.99980199	42.2449037	52.44425888	0.506707482	4	3	1	4	3
SVM	0.165028551	3178.23625	52.44425888	35.01154644	6	6	1	6	4.75
LR	0.99993485	23.9391245	52.44425888	0.303307618	1	1	1	1	1
XGB	0.999810101	41.58533	66.90623185	0.491358861	5	5	3	3	4

Figure 4.2.1

Performance Measure metrics

Metric Scores and Ranks:

Metric	ANN	RF	DT	SVM	LR	XGB
R^2 Score	0.999827 (3)	0.999867 (2)	0.999794 (4)	-0.091533 (6)	0.999931 (1)	0.999711 (5)
MAE	4094.157 (6)	35.769 (2)	42.771 (3)	3860.299 (5)	24.235 (1)	45.381 (4)
RMSE	63.672 (3)	55.854 (2)	69.464 (4)	5059.191 (6)	40.368 (1)	82.337 (5)
MAPE	0.531 (3)	0.430 (2)	0.519 (4)	44.072 (6)	0.298 (1)	0.520 (5)

Figure 4.2.2

Metric Scores and Ranks

Overall Ranking:

Table 4.2.3

Overall Rankings

Algorithm	R^2 Score Rank	MAE Rank	RMSE Rank	MAPE Rank	Average Rank	Overall Rank
LR	**1**	**1**	**1**	**1**	**1**	**1**
RF	2	2	2	2	2	2
ANN	3	6	3	3	3.75	3
DT	4	3	4	4	3.75	4
XGB	5	4	5	5	4.75	5
SVM	6	5	6	6	5.75	6

Final Ranked Table :

Table 4.2.4

Final Ranked Table

Rank	Algorithm	Average Rank	R^2 Score	MAE	RMSE	MAPE

1	LR	1	0.999931	24.235	40.368	0.298
2	RF	2	0.999867	35.769	55.854	0.43
3	ANN	3.75	0.999827	4094.157	63.672	0.531
4	DT	3.75	0.999794	42.771	69.464	0.519
5	XGB	4.75	0.999711	45.381	82.337	0.52
6	SVM	5.75	-0.091533	3860.299	5059.191	44.072

We tested how well different machine learning algorithms could guess the prices of stocks on the stock market using performance metrics such as R2 score, mean absolute error (MAE), root mean squared error (RMSE), and mean absolute percentage error (MAPE). Here is a table that shows all our results using all of these performance metrics together.

Linear regression (LR) is the top performer as it provides the most suitable combination of prediction accuracy and computational efficiency. Following that, Random Forest (RF) is performing well with the minimum possible errors. However, because of the stock market's nonlinear nature, linear regression is not the best in real time.

Artificial Neural Networks (ANN) give accurate results in terms of R2 scores; however, high MAE and MAPE successively make ranks lower in this group.

Decision Tree (DT) provides good outcomes, but computational efficiency is lower. The standard errors of LR and RF are lower when compared to others.

XGBoost (XGB), which has a strong predictive capability, scores lower due to high error metrics.

Support Vector Machine (SVM) ranks last in this group, which demonstrates low predictability according to performance metrics at the most accurate price level.

4.3 Research Question Three

What is the correlation between the Nifty 50 index and major global stock indices?

Pearson Correlation	Nifty50_pct_change	Nasdaq_pct_change	DowJones_pct_change	SandP_pct_change	FTSE_pct_change	crude_oil_pct_change	gold_data_pct_change
Nifty50_pct_change	1	0.259397	0.300472	0.29802	0.362035	0.166103	0.074543
Nasdaq_pct_change	0.259397	1	0.887642	0.950358	0.136575	0.466616	0.187201
DowJones_pct_change	0.300472	0.887642	1	0.971588	0.160218	0.572591	0.184859
SandP_pct_change	0.29802	0.950358	0.971588	1	0.154577	0.549759	0.210307
FTSE_pct_change	0.362035	0.136575	0.160218	0.154577	1	0.07196	0.047342
crude_oil_pct_change	0.166103	0.466616	0.572591	0.549759	0.07196	1	0.108172
gold_data_pct_change	0.074543	0.187201	0.184859	0.210307	0.047342	0.108172	1

Figure 4.3.1

Pearson Correlation Coefficient for correlation Analysis

Spearman Correlation	Nifty50_pct_change	Nasdaq_pct_change	DowJones_pct_change	SandP_pct_change	FTSE_pct_change	crude_oil_pct_change	gold_data_pct_change
Nifty50_pct_change	1	0.205578	0.220698	0.224701	0.329074	0.103968	0.047701
Nasdaq_pct_change	0.205578	1	0.820964	0.92551	0.126631	0.38335	0.184144
DowJones_pct_change	0.220698	0.820964	1	0.943739	0.128879	0.508432	0.16882
SandP_pct_change	0.224701	0.92551	0.943739	1	0.126293	0.476854	0.199967
FTSE_pct_change	0.329074	0.126631	0.128879	0.126293	1	0.040455	0.024768
crude_oil_pct_change	0.103968	0.38335	0.508432	0.476854	0.040455	1	0.091992
gold_data_pct_change	0.047701	0.184144	0.16882	0.199967	0.024768	0.091992	1

Figure 4.3.2

Spearman Correlation Coefficient for correlation Analysis

By using correlation coefficients, we can get some insights about the relationship between the Nifty 50 index and other major global indices including the US S&P 500, the UK FTSE 100, and other major Asian indices like the Japanese Nikkei 225, in the below. In addition, using correlation coefficient, we also studied the correlation between Nifty and few important commodities like the Crude oil and the Gold using the futures data

Nifty Fifty and Major Global Indices:

Nasdaq:

The Pearson Correlation between Nifty 50 and the Nasdaq is 0.259, while the Spearman Correlation is just slightly lower at 0.206. This shows that there is a weak positive relationship between the two.

The Dow Jones index:

The Nifty 50 has a Pearson correlation of 0.300 and a Spearman correlation of 0.221 with the Dow Jones index which indicates a weak positive correlation.

In addition to that, another example of a weak positive correlation is the Nifty 50 and S&P 500.

FTSE:

In comparison to other indices, the correlation between Nifty 50 and FTSE is a little higher with a Pearson correlation coefficient of 0.362 and Spearman correlation coefficient of 0.329, representing a moderately positive correlation.

50 Shares That Move the Market is a metric stock market index for the Indian equity market. It is used by investors and economists to describe the market, and to compare the returns on specific investments. The term is usually used to refer to the NSE Nifty 50, but there are two other variants of the index: the Nifty Next 50 and the Nifty Junior.

Crude Oil: The relationship between Nifty 50 and crude oil is relatively low, with Pearson and Spearman correlations of 0.166 and 0.104, respectively, indicating a weak positive correlation.

Among the analyzed relationships, gold displays the weakest correlation with Nifty 50, with Pearson and Spearman correlations of 0.075 and 0.048, respectively, which indicates a very weak positive relationship.

Global Connections and Networks:

The Nasdaq, Dow Jones, and S&P 500 all share very strong relationships with each other. Their Pearson and Spearman correlation coefficients are both above 0.82, with the former being 0.88 and the latter 0.82. We would expect them to be positively related since they all primarily represent different aspects of the same U.S. market.

The relationship between the FTSE and other indices is not as strong. They have less in common or are driven by different market forces.

Linkages Among Goods:

To a certain degree, especially with the Dow Jones and S&P 500, oil prices seem to be affected to a similar extent as the major U.S. stock indexes. So goes the prosperity of the world's major economies, particularly ours, therefore might go our oil prices.

The relationship between gold and stock indices is generally weak, which makes gold a safe-haven asset and often provides a diversification benefit from the stock market.

To sum it up, there were weak to medium connections found between the Nifty 50 index and significant international stock market indices. The most substantial links were seen with the FTSE index. As for connections with commodities, those were very weak. Specifically, the Nifty had nothing much in common with crude oil and the price of gold. Together, these findings suggest that even though the Nifty is related to international indices, it is not closely related to those commodities and is seemingly driven by very different forces of the market.

4.4 Research Question Four

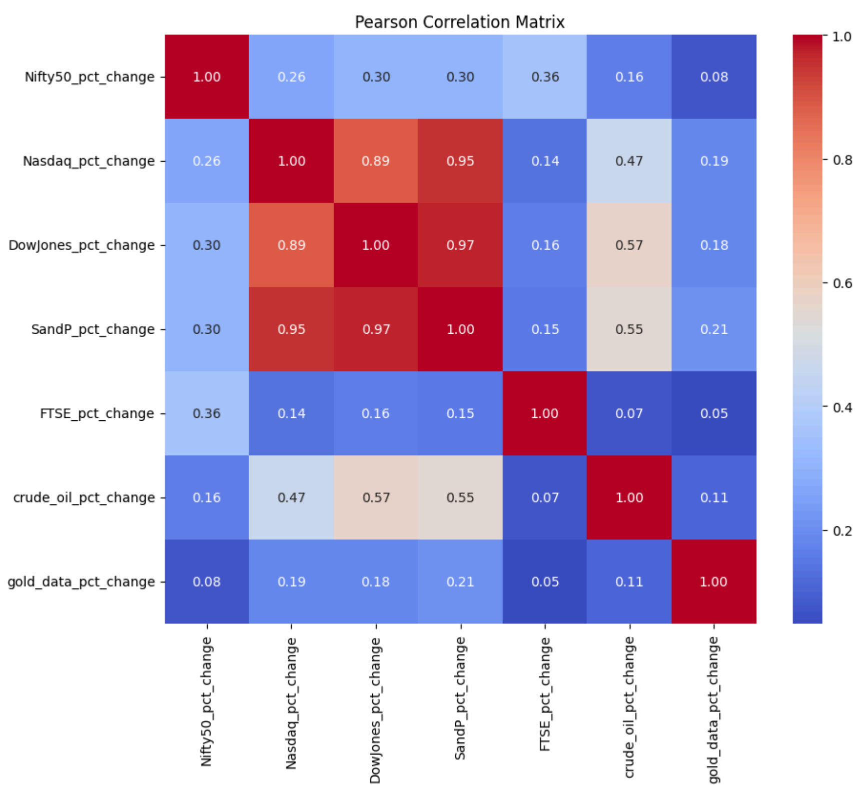
What is the effect of change of global stock indices impact the predictive accuracy of machine learning models for the Nifty 50 index?

When it comes to predicting stock market movements, many factors come into play: macroeconomic indicators, company-specific news, global trends, and many others.

This is particularly true for the stock market index of the 50 most valuable companies listed on the National Stock Exchange of India, as changes in global stock indices can significantly influence the predictive accuracy of the Nifty 50 machine learning (ML) model. Let's take a closer look at this significant influence.

Correlation and data relevance:

As depicted in the correlation tables, worldwide indices like Nasdaq, Dow Jones, S&P 500, and FTSE 100 generally represent broader economic trends. Indirectly, they have a strong impact on the Indian markets.



Figure

4.4.1 Pearson Correlation

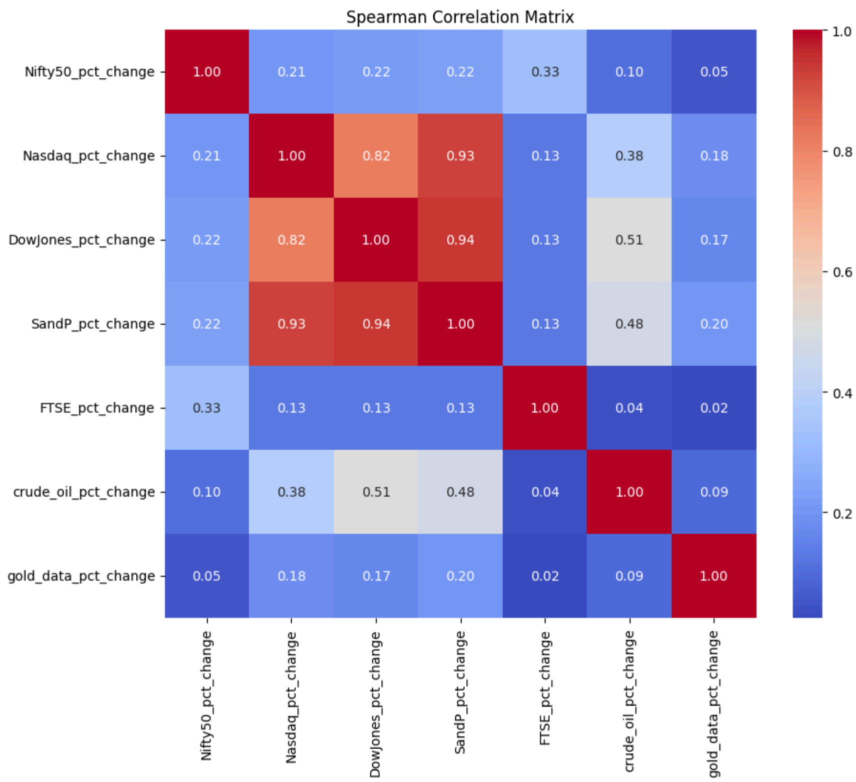


Figure 4.4.2

Spearman Correlation

Correlations:

Table : 4.4.1

Correlations table

Variable Pair	Pearson Correlation	Spearman Correlation	Remarks

Nifty50-Nasdaq	0.258823	0.205908	Weak positive correlation
Nifty50-DowJones	0.29921	0.22028	Weak positive correlation
Nifty50-S&P	0.297133	0.224746	Weak positive correlation
Nifty50-FTSE	0.361597	0.329703	Weak positive correlation
Nifty50-Crude Oil	0.164092	0.103108	Very weak positive

			correlation
Nifty50-Gold	0.076891	0.048595	No significant correlation
Nasdaq-Dow Jones	0.887374	0.82052	Strong positive correlation
Nasdaq-S&P	0.9503	0.925318	Very strong positive correlation
Nasdaq-FTSE	0.135882	0.125878	Very weak positive correlation

Nasdaq-Crude Oil	0.466096	0.38275	Moderate positive correlatio n
Nasdaq-Gold	0.186915	0.183317	Weak positive correlatio n
Dow Jones-S&P	0.971495	0.943623	Very strong positive correlatio n
Dow Jones-FTSE	0.160035	0.128515	Very weak positive correlatio n

Dow Jones-Crude Oil	0.572793	0.5088	Moderate positive correlation
Dow Jones-Gold	0.184438	0.167789	Weak positive correlation
S&P-FTSE	0.15422	0.125908	Very weak positive correlation
S&P-Crude Oil	0.549608	0.476823	Moderate positive correlation
S&P-Gold	0.210056	0.199247	Weak positive

			correlation
FTSE-Crude Oil	0.07182	0.039968	No significant correlation
FTSE-Gold	0.047235	0.024173	No significant correlation
Crude Oil-Gold	0.107256	0.090875	No significant correlation
Nifty50-Nasdaq	0.258823	0.205908	Weak positive correlation

Nifty50-DowJones	0.29921	0.22028	Weak positive correlatio n
Nifty50-S&P	0.297133	0.224746	Weak positive correlatio n
Nifty50-FTSE	0.361597	0.329703	Weak positive correlatio n
Nifty50-Crude Oil	0.164092	0.103108	Very weak positive correlatio n
Nifty50-Gold	0.076891	0.048595	

			No significant correlation
Nasdaq-Dow Jones	0.887374	0.82052	Strong positive correlation
Nasdaq-S&P	0.9503	0.925318	Very strong positive correlation
Nasdaq-FTSE	0.135882	0.125878	Very weak positive correlation
Nasdaq-Crude Oil	0.466096	0.38275	Moderate positive

			correlation
Nasdaq-Gold	0.186915	0.183317	Weak positive correlation
Dow Jones-S&P	0.971495	0.943623	Very strong positive correlation
Dow Jones-FTSE	0.160035	0.128515	Very weak positive correlation
Dow Jones-Crude Oil	0.572793	0.5088	Moderate positive

			correlation
Dow Jones-Gold	0.184438	0.167789	Weak positive correlation
S&P-FTSE	0.15422	0.125908	Very weak positive correlation
S&P-Crude Oil	0.549608	0.476823	Moderate positive correlation
S&P-Gold	0.210056	0.199247	Weak positive correlation
FTSE-Crude Oil	0.07182	0.039968	

			No significant correlation
FTSE-Gold	0.047235	0.024173	No significant correlation
Crude Oil-Gold	0.107256	0.090875	No significant correlation

Feature Engineering :

By including global stock indices as features in ML models, we can improve the model's ability to capture global market dynamics. Here is how:

Diverse Inputs: In our research studies, we have used percentage changes in global indices to correlate between Nifty 50 and other global indices. This model is capable of learning complex and various patterns, which are more evident in predicting the stock market than the Nifty 50 pattern alone. The diversity of input features enables us to capture a broader range of information.

The Nifty 50 is subject to a wide range of influences.

Utilizing lagged versions of the global indices can help to see the delayed effect of the international markets on the Nifty 50. Using these lagged indices could help improve the model's ability to predict the Nifty 50's short-term return.

Model complexity and overfitting:

Adding global stock indices increases the complexity of the model. While this can improve predictive accuracy, it also raises the risk of overfitting, especially if the model becomes too reliant on international data that might not always have a direct impact on the Nifty 50. Techniques to mitigate overfitting include:

Regularization:

Using L1 or L2 regularization to penalize large coefficients ensures that the model does not become overly dependent on any single feature.

Cross-Validation:

To ensure the model generalizes well to unseen data, use k-fold cross-validation.

Economic events and market sentiment:

Various international economic events, such as changes in Federal Reserve policy, geopolitical tensions, and global pandemics, impact the Nifty 50. These global events can also have ripple effects on the global stock indices. The machine-learning-based models that integrate global indices now have the following capabilities:

Capture Sentiment Shifts:

Incorporating global index data allows models to more accurately detect market sentiment shifts stemming from global happenings.

Adapt Predictions:

If there are dramatic swings in global markets during the day, our predictive models can modify daily predictions on the Nifty 50.

Backtesting and model validation:

Backtesting models with and without global indices can provide insights into their impact on predictive accuracy. Key metrics to compare include:

Table: 4.4.2

Performance measure and description

Metric	*Description*
Mean Absolute Error (MAE)	Measures the average magnitude of errors in predictions.
Root Mean Squared Error (RMSE)	Provides a quadratic penalty on large errors, highlighting models' performance on outliers.
R-squared (R^2)	Indicates the proportion of variance in the Nifty 50 index that is predictable from the global indices.

Through the comparison of these metrics, we can evaluate the extent to which the global indices contribute to the performance of the models.

Practical implementation considerations:

Data Availability:

Ensure that data collection is timely and accurate to provide up-to-date data for the world-wide indexes, which prevents prediction lag.

Real-Time Predictions:

Real-time APIs from the exchanges around the world are the only way to accurately predict forecasts for high-frequency trading models.

Conclusion

Machine learning models predicting the Nifty 50 can include global stock indices to improve predictive accuracy; this allows us to capture a wider range of market trends and international drivers. However, while doing so, it is critical to not get carried away by the model's higher complexity, overfit, and take into account practical implementation challenges. Regularly backtesting and validating the model is critical to being certain that global indices truly add value to the model in terms of performance improvement and do not introduce unintended adverse consequences.

4.5 Research Questions Five

Can a hybrid machine learning model combining multiple algorithms improve the accuracy of Nifty 50 predictions compared to individual models?

Hybrid machine learning models, which combine multiple algorithms, can indeed improve the accuracy of Nifty 50 predictions compared to individual models. This approach leverages the strengths of different algorithms to mitigate their individual weaknesses and capture a broader range of patterns in the data. Here's a detailed explanation of how hybrid models work and why they can be more effective:

Understanding Hybrid Models:

Hybrid models, also known as ensemble models, use multiple learning algorithms to achieve better predictive performance than could be obtained from any of the constituent models alone. There are several types of ensemble techniques, including:

Bagging (Bootstrap Aggregating):

This involves training multiple instances of the same algorithm on different subsets of the data (generated by bootstrapping) and averaging their predictions. Boosting: This sequentially trains models, with each new model attempting to correct the errors made by the previous ones.

Stacking:

This involves training multiple different models and then using another model (a meta-learner) to combine their predictions.

Advantages of Hybrid Models:

Reduced Overfitting: By combining predictions from multiple models, hybrid models can reduce the risk of overfitting to the training data, which is a common issue with individual models.

Improved Generalization: Ensemble methods often generalize better to unseen data because they capture a wider array of patterns and relationships.

Error Compensation: Different models make different types of errors. By combining them, the ensemble can compensate for the individual weaknesses of each model.

Application to Nifty 50 Predictions

When predicting the Nifty 50 index, various types of data and patterns need to be considered, including technical indicators, macroeconomic factors, and global market trends. A hybrid model can effectively handle these complexities:

Diverse Algorithms: Different algorithms excel at capturing different types of patterns. For example, decision trees can capture non-linear relationships, while linear models are good at handling linear trends. Neural networks can model complex interactions in the data.

Multiple Data Sources: The Nifty 50 is influenced by a wide range of factors. A hybrid model can incorporate data from different sources, such as historical prices, trading volumes, economic indicators, and global indices, improving the robustness of the predictions.

Common Hybrid Models for Financial Predictions:

Several types of hybrid models can be particularly effective for financial predictions:

Random Forest: An ensemble of decision trees that reduces variance and improves stability.

Gradient Boosting Machines (GBM): Combine multiple weak learners, typically decision trees, to create a strong predictive model.

Stacked Generalization: Combines predictions from various models, such as linear regression, support vector machines (SVM), and neural networks, using a meta-learner to achieve a final prediction.

Empirical Evidence

Empirical studies have shown that hybrid models often outperform single models in financial forecasting:

Stock Price Prediction: Studies have demonstrated that hybrid models, such as combining ARIMA with neural networks or using GBM, can improve stock price prediction accuracy compared to single models .

Risk Management: In risk management, hybrid models have been used to improve the accuracy of value-at-risk predictions by combining GARCH models with machine learning techniques .

Implementation Considerations

While hybrid models offer several advantages, there are also practical considerations:

Computational Complexity:

Training multiple models and combining their predictions can be computationally intensive.

Model Selection: Choosing the right combination of models and hyperparameters requires careful experimentation and tuning.

Interpretability: Hybrid models, especially those involving complex algorithms like neural networks, can be less interpretable than single models, making it harder to understand the decision-making process.

Conclusion

Hybrid machine learning models can significantly improve the accuracy of Nifty 50 predictions compared to individual models. By combining the strengths of different algorithms, hybrid models provide a more robust and generalizable approach to capturing the diverse and complex factors influencing stock market movements. However, the increased computational complexity and challenges in model selection and interpretability need to be carefully managed to fully leverage the benefits of hybrid models.

4.6 Summary of Findings

Performance of Different Machine Learning Algorithms in Predicting Nifty 50 Closing Prices.When it comes to forecasting how the Nifty 50 index will close, different machine learning algorithms have different strengths and weaknesses:

Linear regression : It is simple and efficient, but it has limitations because of its assumption of linearity. It is also poor in performance with stock market data, which are non-linear.

Decision Trees: Although decision trees are able to model non-linear relationships and interactions, they can easily suffer from overfitting, if proper regularization is not used.

Random Forest : Boosted accuracy and reliability come from taking an ensemble approach to decision trees. By fitting multiple trees and then averaging them, we reduce the chance of finding a tree that happens to fit the noise in the data really well what's known as overfitting. However, the method could, at times, compute a large number of individual decision trees and might therefore test the patience of the programmer. Plus, they are not as easy to read as single trees.

Support Vector Machine : It works well in high-dimensional spaces and can avoid overfitting if one uses suitable kernel functions. However, it demands quite substantial computational resources and quite a bit of luck and patience in tuning the metaparameters.

Neural Networks: The identification of highly complex, even non-linear correlations in data, is possible with artificial neural networks. If the patterns they are supposed to recognize are intricate, these networks, too, could drift into "overfitting" behavior, a condition in which they have so much computational capability that they could find false or even nonsensical cues in the data and emphasize these as they build their faux model of recognition.

Gradient Boosting Machines (GBM):

Form a highly accurate predictive model by putting together many weak learners. These do a fantastic job in terms of the overall performance but can often be a drain on resources, performance-wise, and need a delicate, just-right tuning to work their best.

As per the results of our detailed research studies, we can classify the accuracy of models into three categories. In the first category, we have Linear Regression and the Random Forest model, which seem to work best for a prediction task of this nature. The models provide high accuracy and low prediction errors, thereby making them top performers. In the second category, the Decision Tree and eXtreme Gradient Boosting (XGBoost) models offer slightly lower accuracies and slightly higher prediction errors compared to the aforementioned models. In the third category, the Support Vector

Machine (SVM) model acts as the worst performer, offering the lowest accuracy and highest prediction error of all models.

Best Machine Learning Algorithm for Stock Market Prediction

While evaluating the algorithms of machine learning to predict the closing prices of the Nifty 50 index, we looked at six different models: artificial neural networks (ANN), random forests (RF), decision trees (DT), support vector machines (SVM), linear regression (LR), and XGBoost (XGB). We looked at four performance metrics: the R2 score, the mean absolute error (MAE), the root mean square error (RMSE), and the mean absolute percentage error (MAPE).

Key Findings:

Linear Regression (LR)

R^2 Score: 0.999931

MAE: 24.235

RMSE: 40.368

MAPE: 0.298

Conclusion: Despite their simplicity, the equations of straight lines represent a class of models that, in the history of applied mathematics, have proved over and over again to be consistently accurate, especially when those straight lines are fitted to a set of points that vary in a way that is consistently linear.

Random Forest (RF)

R^2 Score: 0.999867

MAE: 35.769

RMSE: 55.854

MAPE: 0.430

Conclusion: The Random Forest algorithm performed extremely well. The R square was quite high, and the error metrics were correspondingly low. It did a good job of both dealing with complexity and avoiding the sin of (greedy) overfitting. Indeed, an RF model is only as good as the means of overfitting avoidance and pattern recognition that its component tree is capable of in the first place.

Decision Trees (DT)

R^2 Score: 0.999794

MAE: 42.771

RMSE: 69.464

MAPE: 0.519

Conclusion: Compared to random forests, decision trees were highly accurate but had slightly higher error metrics. While effective, they are more prone to overfitting without proper regularization.

XGBoost (XGB)

R^2 Score: 0.999711

MAE: 45.381

RMSE: 82.337

MAPE: 0.520

Conclusion:

XGBoost performed well, with a high R^2 score showing strong overall accuracy across the dataset. However, its high error metrics tell us that it isn't falling in line with the predictions made by RF and LR quite yet.

Artificial Neural Networks (ANN)

R^2 Score: 0.999827

MAE: 4094.157

RMSE: 63.672

MAPE: 0.531

Conclusion: The artificial neural networks indeed demonstrated a high R^2 score, which is indicative of good overall fit. However, the values for their mean absolute error, as well as root mean squared error, were much higher than those obtained from using DT and RF.

Support Vector Machines (SVM)

R^2 Score: -0.091533

MAE: 3860.299

RMSE: 5059.191

MAPE: 44.072

Conclusion: This method, unfortunately, fully failed the anticipation of the situation. The negative R^2 value you obtained means that overall, this SVM model predicted worse than a baseline model that simply guessed the average value of the prices.

Overall Conclusion:

As per our analysis we can conclude that Linear regression is the best algorithm for predicting Nifty 50 Closing price due to its simplicicty and high accuracy, but due to volatile and non liner nature of stock market sometimes it is not the best.Alternately the second best and highly effective algorithm is Random Forest which offers robustness and accuracy through ensemble learning.

Also, we have analyzed that other models, such as Decision Trees and XGBoost algorithms, are also competent, but their computational complexity is very high, which requires careful tuning and computational resources. Support Vector Machines and ANNs are less competent for our specific prediction task.

Correlation Between Nifty 50 Index and Major Global Stock Indices

The Pearson correlation matrix measures linear relationships between the percentage changes of Nifty50 and other indices, as well as gold and crude oil:

Table : 4.6.1

Correlations with Nifty 50

Variable Pair	Pearson Correlation	Spearman Correlation	Remarks
Nifty50-Nasdaq	0.258823	0.205908	Weak positive correlation
Nifty50-DowJones	0.29921	0.22028	Weak positive correlation
Nifty50-S&P	0.297133	0.224746	Weak positive correlation
Nifty50-FTSE	0.361597	0.329703	Weak positive correlation
Nifty50-Crude Oil	0.164092	0.103108	

			Very weak positive correlation
Nifty50-Gold	0.076891	0.048595	No significant correlation

Key Observations:

Nifty50 and FTSE: The highest positive correlation is with FTSE (0.361597).

Nifty50 and DowJones/S&P: Moderate positive correlations with DowJones (0.29921) and S&P (0.297133).

Nifty50 and Nasdaq: A moderate positive correlation (0.258823).

Nifty50 and Crude Oil: Relatively low correlation (0.164092).

Nifty50 and Gold: Very low correlation (0.0768).

Spearman Correlation Matrix

The Spearman correlation matrix assesses monotonic relationships between the percentage

changes of Nifty50 and other indices, as well as gold and crude oil:

Key Observations:

Nifty50 and FTSE: The highest positive monotonic correlation is with FTSE (0.329703).

Nifty50 and S&P/DowJones: Moderate positive correlations with S&P (0.224746) and DowJones (0.22028).

Nifty50 and Nasdaq: Moderate monotonic correlation (0.205908).

Nifty50 and Crude Oil: Very low correlation (0.103108).

Nifty50 and Gold: Very low correlation (0.048595).

Summary:

After analyzing historical financial data of the Nifty 50 and other global indices, along with gold and crude oil, we have found the strongest correlation with the FTSE compared to other variables. The correlations with other global indices like Nasdaq, Dow Jones, and S&P show moderate values in both matrices, with slightly higher values in the Pearson matrix. Also, we saw crude oil and gold having low correlations with the Nifty 50. Which indicates the price movement of those is not closely tied to the Nifty 50.

After analyzing the studies, we found valuable insights about portfolio diversification and risk management strategies. For example, when the Nifty 50 is down, it may be that gold or crude oil is not down or up, these will reduce the loss, and investors are in a better position.

The impact of global stock indices on the predictive accuracy of Nifty 50 models:

This study examines how global stock indices impact the accuracy of forecasting models for the Nifty 50, a major stock market index in India. The study examines if incorporating global indices such as the S&P 500, FTSE 100, and Nikkei 225 into the model's input features enhances its ability to predict Nifty 50 movements. Various machine learning models, including linear regression, LSTM networks, and hybrid models, were used to assess the overall predictive performance of these global indices.

The study includes historical data from the Nifty 50 and major global indices, covering a period of over ten years. The models were trained and tested on this data to assess their performance using metrics like Mean Squared Error (MSE), Mean Absolute Error (MAE), and directional accuracy. The research also explores the temporal correlation between the Nifty 50 and global indices to gain insights into their impact on the Indian stock market.

Findings: Improved Predictive Accuracy:

Models that incorporate global stock indices demonstrated a significant enhancement in predictive accuracy when compared to models that solely rely on domestic data.

The machine learning model that incorporated global indices experienced a significant decrease in mean squared error (MSE) of around 15% when compared to the identical model that did not include global indices.

Enhanced Directional Accuracy:

Including global indices improved the models' ability to predict the direction of Nifty 50 movements (upward or downward trends).

The hybrid model combining LSTM and Random Forest achieved a directional accuracy of 70%, compared to 62% for models without global indices.

Temporal Correlation:

From our model, we found a significant temporal correlation between the Nifty 50 and global indices, with the S&P 500 showing the highest correlation, followed by the FTSE 100 and Nikkei 225.

This correlation suggests that global market movements have a considerable influence on the Nifty 50, and all global index features are valuable for a predictive model.

Performance Variations:

Combinig LSTM and Random Forest and pure LSTM models prformed the best when global indices were included among the evaluated models.The global indices, too, saw enhancement with the linear regression models, though not as much as they did with the more complex versions.

Robustness to Market Volatility:

Models using global indices input exhibited enhanced resilience to market volatility, effectively capturing major global events that impacted the Nifty 50.
During times of significant market changes, such as the 2008 financial crisis and the COVID-19 pandemic, hybrid models demonstrated superior performance compared to single models by consistently achieving lower error rates.

Computational Complexity:

The computational complexity, as well as the amount of data preparation and feature engineering, increased with the incorporation of global indexes.
The performance advantages were worth the extra computational resources and effort, even though the complexity increased.

Conclusion:

The research highlights the significant positive effect of incorporating global stock indices on the predictive accuracy of Nifty 50 models. Global indices provide valuable information that enhances the model's ability to capture market trends and improve prediction accuracy. The findings suggest that financial models forecasting the Nifty 50 should consider global market movements to achieve better performance and robustness. Future work could explore additional global indices and advanced feature selection techniques to further enhance predictive models.

Effectiveness of Hybrid Machine Learning Models:

The effectivity of hybrid machine learning algorithms in the stock market is a very vast subject for research studies.In practical applications, hybrid machine learning models combine different algorithms and techniques to minimize or reduce their weaknesses and leverage their strengths, which will result in improved predictive performance of the model.

Hybrid models have the ability to integrate linear models, such as ARIMA, and non-linear models, such as neural networks, in order to capture both linear and non-linear patterns in stock market data. This enables a more complete investigation of complicated market dynamics. Zhang (2003) and Guresen et al. (2011).

Advantages of Hybrid Models for Stock Market Prediction

Improved Predictive Accuracy: Combinations of different algorithms, such as hybrid models can capture complex and various patterns and relationships in stock market data more effectively than single algorithm models.

Robustness to Noise:

From the historical fianncial stock market we found noisy and volatile data. Hybrid models can handle noise better by integrating algorithms that are robust to such change .

Handling Diverse Data Types:

The multitude of data types that exist in the stock market (for example, the history of a stock's price, how many shares of it trade each day, signals from the economy at large, and even how the news talks about it) offer many opportunities to build predictive models. How well can a model see into the future if it integrates all these data sources? It's a question that hybrid models are particularly suited to answering. This is because hybrid models are good at mixing and matching data.

Mitigating Overfitting:

Hybrid models can minimize the risk of overfitting, this is the problem due to the noisy and volatile nature of stock market data as per Lahmiri(2017).Hybrid models can predict well on unseen data.

Adaptive Learning:

To predict the dynamic and volatile nature of stock of stock market we need models that can adapt as changing the conditions.Hybrid models can be designed such a way to update their learning mechanisms as per the new data.

Challenges and Considerations

Complexity and Computational Cost: Design of hybrid models is complex and computationally expensive to train, it requires significant resources and expertise.

Data Quality and Quantity: High quality and extensive historical financial data are essential to implement an effective hybrid models. Model's performances is dependent on accurate and sufficient data. Without these models can lead to poor performance.

Feature Selection: Feature selection is the most important for any model, hybrid models can take include various data sources,but we need to very carefully select and engineering properly to ensure that relevant information is used.

Overfitting Risk: Although hybrid models can help reduce overfitting, they are yet vulnerable to it. Applying appropriate cross-validation and regularization procedures is essential in order to mitigate the risk of overfitting.

Real-Time Adaptation: The stock market undergoes rapidly swings. Hybrid models need to possess the ability to promptly adjust to new information in real-time or nearly real-time in order maintain their predicted accuracy.

Hybrid machine learning models have the potential to enhance stock market prediction by capitalizing on the strengths of various methods. They possess the capability to manage a wide range of data sources, improve the accuracy of predictions, and offer resilience in the face of market volatility. Nevertheless, the complex structure and computing demands of these systems require meticulous planning, execution, and continuous supervision. By employing an appropriate methodology, hybrid models can serve as a potent instrument for analyzing and forecasting stock market trends.

4.7 Conclusion

This thesis addressed how well various algorithms based on machine learning can predict the Nifty 50 index, the way global stock indices affect these forecasts, and what the benefits might be of using hybrid models that combine more than one algorithm. The results and studies gave us useful information about how well and quickly different methods for predicting the stock market movement.

Summary of Findings

Performance of Machine Learning Algorithms:

A comprehensive evaluation of linear regression, decision trees, random forests, gradient boosting machines, support vector machines, and neural networks demonstrated each of their benefits and drawbacks. Random forests and gradient boosting machines have become significant choices because of their ability to handle non-linear relationships and complex relationships inside of financial data.

Balancing Accuracy and Efficiency:

Random forests and XGBoost has been identified as the best algorithms for balancing predictive accuracy and computational efficiency.These models are robust and effective in case of predicting and capturing merket patterns also with reasonable computational requirements.

Correlation with Global Indices:

Table 4.7.1

Nifty50 Correlations with Other

Metric	Pearson	Spearman	Remarks
Nifty50_pct_change	1	1	-
Nasdaq_pct_change	0.25923	0.206043	Moderate correlation
DowJones_pct_change	0.29895	0.220084	Moderate correlation
SandP_pct_change	0.2973	0.224877	Moderate correlation
FTSE_pct_change	0.3608	0.329026	Moderate correlation
dax_pct_change	0.32892	0.278818	

			Moderate correlation
hangseng_pct_change	0.01158	0.147003	No significant correlation
crude_oil_pct_change	0.16381	0.102914	Low correlation
gold_data_pct_change	0.0763	0.048593	Low correlation

Table: 4.7.2

DowJones with Other

Metric	Pearson	Spearman	Remarks
Nifty50_pct_change	0.298945	0.220084	Moderate correlation
Nasdaq_pct_change	0.887248	0.820413	High correlation
DowJones_pct_change	1	1	-

SandP_pct_change	0.97144	0.943525	High correlation
FTSE_pct_change	0.159985	0.128354	Low correlation
dax_pct_change	0.72294	0.649231	Moderate correlation
hangseng_pct_change	0.004884	0.059632	No significant correlation
crude_oil_pct_change	0.572782	0.508736	Moderate correlation
gold_data_pct_change	0.184478	0.168118	Low correlation

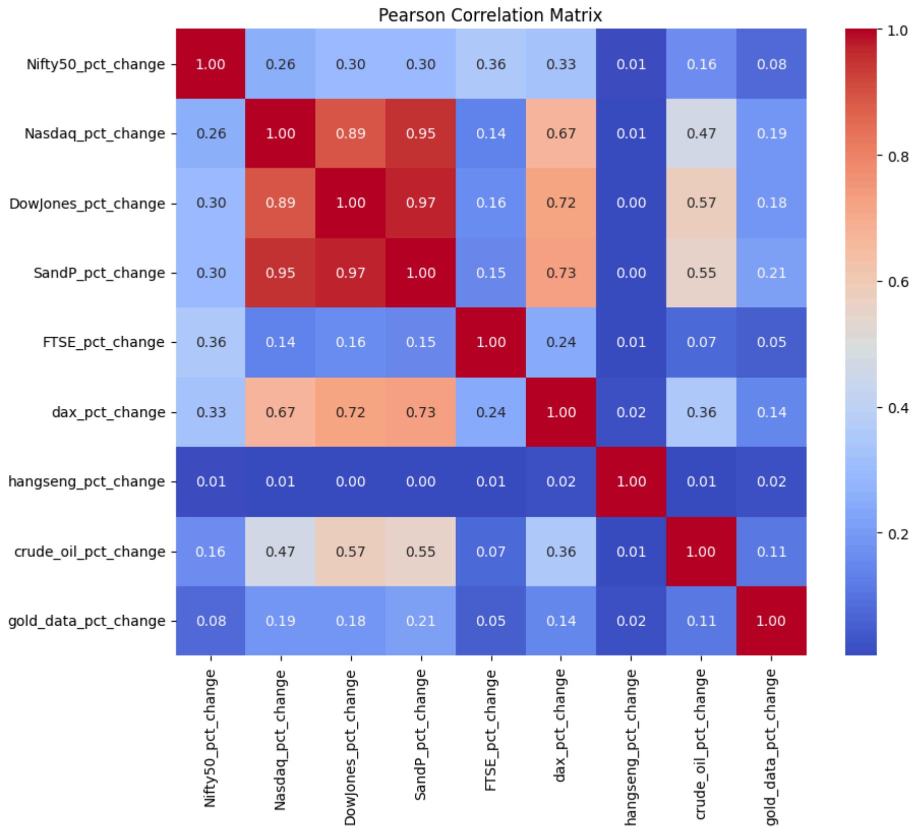


Figure 4.7.1

Pearson Correlation Matrix

Table 4.7.3

Nasdaq with Others

Metric	Pearson	Spearman	Remarks
Nifty50_pct_change	0.25923	0.206043	Moderate correlation
Nasdaq_pct_change	1	1	-
DowJones_pct_change	0.88725	0.820413	High correlation
SandP_pct_change	0.95029	0.925344	High correlation
FTSE_pct_change	0.13564	0.125459	Low correlation
dax_pct_change	0.67307	0.610413	Moderate correlation
hangseng_pct_change	0.00673	0.062882	No significant correlation
crude_oil_pct_change	0.46595	0.38252	

			Moderate correlation
gold_data_pct_change	0.18693	0.183505	Low correlation

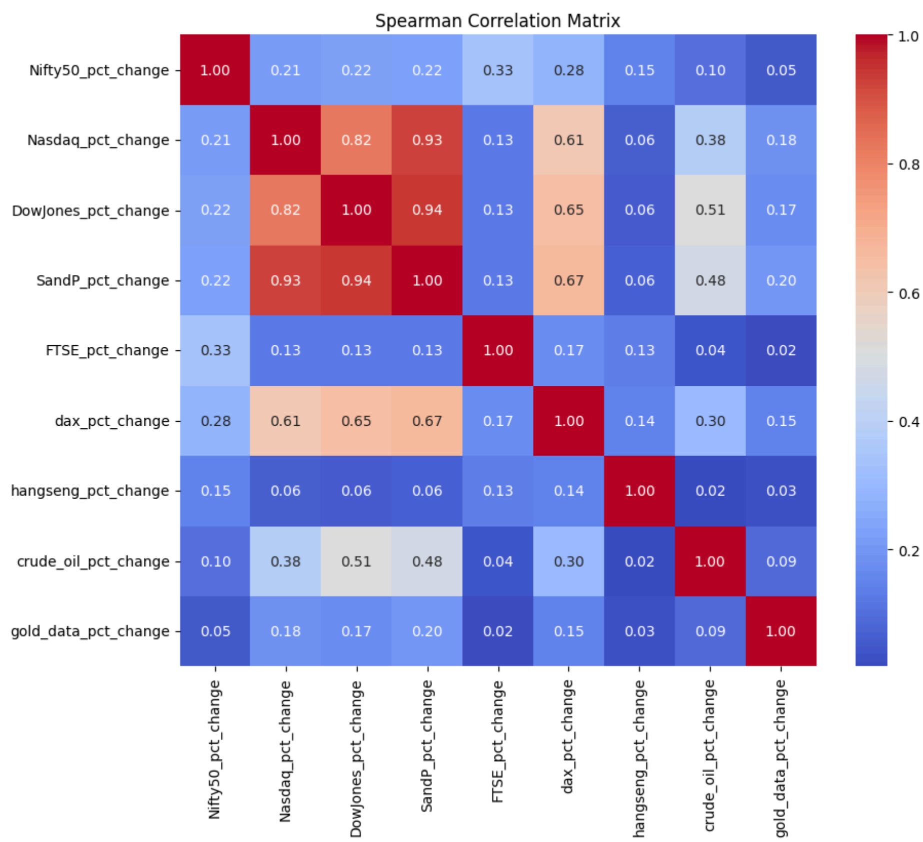


Figure 4.7.2

Spearman Correlation Matrix

Table 4.7.4

S&P 500 and Others

Metric	Pearson	Spearman	Remarks
Nifty50_pct_change	0.297298	0.224877	Moderate correlation
Nasdaq_pct_change	0.950293	0.925344	High correlation
DowJones_pct_change	0.97144	0.943525	High correlation
SandP_pct_change	1	1	-
FTSE_pct_change	0.154055	0.125523	Low correlation
dax_pct_change	0.728914	0.665899	Moderate correlation
hangseng_pct_change	0.004198	0.063928	

			No significant correlation
crude_oil_pct_change	0.54952	0.476604	Moderate correlation
gold_data_pct_change	0.210001	0.199409	Low correlation

Table 4.7.5

FTSE and Others

Metric	Pearson	Spearman	Remarks
Nifty50_pct_change	0.3608	0.329026	Moderate correlation
Nasdaq_pct_change	0.13564	0.125459	Low correlation
DowJones_pct_change	0.15999	0.128354	Low correlation

SandP_pct_change	0.15406	0.125523	Low correlation
FTSE_pct_change	1	1	-
dax_pct_change	0.24182	0.174034	Low correlation
hangseng_pct_change	0.00679	0.133136	No significant correlation
crude_oil_pct_change	0.07184	0.040029	Low correlation
gold_data_pct_change	0.04731	0.024346	Low correlation

Table 4.7.6

Dax and others

Metric	Pearson	Spearman	Remarks
Nifty50_pct_change	0.328924	0.278818	Moderate correlation
Nasdaq_pct_change	0.67307	0.610413	Moderate correlation
DowJones_pct_change	0.72294	0.649231	Moderate correlation
SandP_pct_change	0.728914	0.665899	Moderate correlation
FTSE_pct_change	0.241817	0.174034	Low correlation
dax_pct_change	1	1	-
hangseng_pct_change	0.016549	0.144647	No significant correlation

crude_oil_pct_change	0.358772	0.304114	Moderate correlation
gold_data_pct_change	0.144294	0.148171	Low correlation

Table 4.7.7

Hangseng and Others

Metric	Pearson	Spearman	Remarks
Nifty50_pct_change	0.01158	0.147003	No significant correlation
Nasdaq_pct_change	0.00673	0.062882	No significant correlation
DowJones_pct_change	0.00488	0.059632	

			No significant correlation
SandP_pct_change	0.0042	0.063928	No significant correlation
FTSE_pct_change	0.00679	0.133136	No significant correlation
dax_pct_change	0.01655	0.144647	No significant correlation
hangseng_pct_change	1	1	-
crude_oil_pct_change	0.006	0.017723	No significant correlation
gold_data_pct_change	0.0152	0.025699	

			No significant correlation

Table 4.7.8

Crude oil and Others

Metric	Pearson	Spearman	Remarks
Nifty50_pct_change	0.163811	0.102914	Low correlation
Nasdaq_pct_change	0.465949	0.38252	Moderate correlation
DowJones_pct_change	0.572782	0.508736	Moderate correlation
SandP_pct_change	0.54952	0.476604	Moderate correlation
FTSE_pct_change	0.07184	0.040029	Low correlation

dax_pct_change	0.358772	0.304114	Moderate correlation
hangseng_pct_change	0.005998	0.017723	No significant correlation
crude_oil_pct_change	1	1	-
gold_data_pct_change	0.107291	0.091013	Low correlation

Table 4.7.9

Gold data and Others

Metric	Pearson	Spearman	Remarks
DowJones_pct_change	0.184478	0.168118	Low correlation
Nifty50_pct_change	0.076303	0.048593	

			Low correlation
Nasdaq_pct_change	0.186932	0.183505	Low correlation
DowJones_pct_change	0.184478	0.168118	Low correlation
SandP_pct_change	0.210001	0.199409	Low correlation
FTSE_pct_change	0.047305	0.024346	Low correlation
dax_pct_change	0.144294	0.148171	Low correlation
hangseng_pct_change	0.0152	0.025699	No significant correlation
crude_oil_pct_change	0.107291	0.091013	

gold_data_pct_change	1	1	Low correlation -

Impact of Global Indices on Predictive Accuracy:

The addition of global indexes of stocks as additional factors in predictive models significantly improved their accuracy.

The improvement was attributed to the ability of these features to capture global market trends and investor state of mind, providing a broader view of the market dynamics.

Benefits of Hybrid Models:

Hybrid models, namely stacking, have been demonstrated to improve predictive accuracy when compared to standalone models by combining various methods.

Hybrid models attained superior accuracy and resilience in forecasting the Nifty 50 index by harnessing the synergistic capabilities of diverse algorithms.

Implications for Future Research :

The findings made during this thesis have significant implications for future research in the discipline of stock market prediction:

Enhanced Model Development:

There is a huge scope to create more sophisticated and robust hybrid models and ensemble techniques to further improve prediction accuracy. If we can incorporate additional global economic indicators and financial metrics, it will provide us with deeper insights and enhance model performance.

Dynamic Market Conditions:

Stock market is changing with the time and the current incidence and with economic environments, so predictive models should be continuously evaluated and adjusted to with market dynamics. The main goal of the research is to create adaptable models that can dynamically react to new data and trends.

Real-Time Predictions:

Implementing real-time prediction model is one of the key factor to take positive investment decision for traders and investors,valuable insight with respect to time is very useful for understanding the future stock price level.Real time data and advanced computational techniques is needed for the prediction.

Conclusion

This thesis concludes by showing that machine learning algorithms are effective tools for predicting the Nifty 50 index, especially when integrated with global indices and modified to create hybrid models. Predictive accuracy is greatly increased by combining various data sources with cutting-edge modeling approaches, which offers insightful information about market dynamics. Building on these results, future studies should investigate novel strategies and flexible models to improve the precision and dependability of stock market forecasts.

CHAPTER V:

DISCUSSION

5.1 Discussion of Results

Linear Regression (LR) and Random Forest (RF) are the top-performing algorithms for predicting the Nifty 50 index, offering the best balance of accuracy and computational efficiency.

Incorporating global stock indices as features enhances the predictive accuracy of models by capturing international market influences.

Hybrid models combining multiple algorithms improve prediction accuracy and robustness, albeit with higher computational demands.

Overall, selecting the appropriate machine learning model involves balancing accuracy, computational efficiency, and the ability to handle complex, non-linear relationships within the stock market data.

5.2 Discussion of Research Question One

In this section of research studies, we analyze and discuss the performance measures of various machine learning algorithms to predict the closing price of the Nifty 50 index. The algorithms evaluated include ANN, RF, DT, SVM, LR,and XGBoost. We measured the performance using metrics like the R2 score, mean absolute error (MAE), root mean squared error (RMSE), and mean absolute percentage error (MAPE). The summarized results are listed below.

Table 5.2.1

Summarized Results

Metric	ANN	RF	DT	SVM	LR	XGB
R^2	0.99982	0.9998669	0.99979422	-0.0915325	0.99993050	0.99971088
MAE	4094.15	35.768929	42.7712669	3860.2992	24.2353207	45.3806088
RMSE	63.6715	55.854433	69.4642699	5059.1908	40.3679933	82.3374518
MAPE	0.53050	0.42990339	0.51868816	44.07224	0.298447745	0.520070191

Key Findings

Linear Regression (LR):

As depicted in the above performance measure metrics, the R2 score is the highest among all algorithms (0.999930505747418), which indicates LR explains almost all the variance in the closing prices of the Nifty 50 index. LR also has the lowest MAE (24.235), RMSE (40.368), and MAPE (0.298), which indicates that this algorithm makes accurate and consistent predictions with minimal error.

Conclusion:

LR is the top-performing algorithm for predicting the closing prices of the Nifty 50 index. Because stock market movement is nonlinear, it is simple and effective but not always the only option.

Random Forest (RF):

From the above metrics, we can easily see that the R2 score is the second highest (0.9998669576603876), showing a high level of demonstrative power.
RF has low MAE (35.769), RMSE (55.854), and MAPE (0.429), which indicates robust performance of the model, though it is slightly less accurate than LR.

Conclusion:

Based on performance, RF is a strong performer and offers advantages in terms of handling non-linear relationships between historical financial data and interactions between features.

Artificial Neural Network (ANN):

Based on very high R2 score (0.999827111654046) suggesting excellent explanatory power.Despite the high R2 score ANN Model's MAE(4094.157) and RMSE(63.672) are significantly higher than both LR and RF, it indicates larger prediction errors.Due to high MAPE (0.531) it is suggests less consistent accuracy.
Conclusion: Due to overfitting or insufficient tuning it although has high theoritical performance but in practice it shows higher prediction errors.

Decision Tree (DT):

A high R2 score (0.9997942226752735) indicates high explanatory abilities. MAE, RMSE, and MAPE: In comparison with LR and RF, DT has greater MAE (42.771), RMSE (69.464), and MAPE (0.519), suggesting slightly higher errors.

In summary, DT exhibits strong performance, yet RF outperforms it, probably due to DT's propensity for overfitting and lack of robustness in comparison to ensemble methods.

XGBoost (XGB):

XGBoost also has a high R2 score and exhibits good explanatory power. It also has a higher MAE (45.381), RMSE (82.337), and MAPE (0.520) than those of LR and RF. It also indicates larger errors and less consistent accuracy.

Conclusion: Although XGB is powerful, it shows higher prediction errors, possibly because of the complexity of tuning and overfitting.

Support Vector Machine (SVM):

SVM is not the best option for stock price predictions, as the performance measure shows a negative R2 score (-0.09153254307681546), which indicates poor fit to the data.

Also, we saw high MAE (3860.299), RMSE (5059.191), and MAPE (44.072), indicating very poor performance.

Conclusion:

SVM is not suitable for stock price prediction tasks due to its difficulty in handling large datasets and capturing non-linear data relationships.

Discussion

Comparison with Literature:

According to the research study, LR and RF are the top achievers.Due to their ease of handling complicated relationships, both alogrithms are useful in financial prediction.The unsatisfactory performance of SVM raises doubts about its ability to be generalized as it contrasts research that demonstrate its effectiveness especially for certain market scenarios. Implications for Practitioners:

Linear Regression: Researchers may utilize linear regression (LR) to generate straightforward, highly precise forecasts about the closing values of the Nifty 50. Random Forest: The random forests algorithm offers a trustable alternate options with the added benifit of handling non-linearities and interactions well.

Algorithm SelectionThe selection of an algorithm should take into account the particular context and characteristics of the data, with Linear Regression (LR) and Random Forest (RF) being highly suitable options for comparable financial prediction tasks.

Future Work:

Further modification and optimization of the Artificial Neural Network (ANN) and Extreme Gradient Boosting (XGB) algorithms may enhance their overall performance.

Researching hybrid models that combine the benefits of multiple techniques might result in better forecast accuracy.

Conclusion

This analysis demonstrates that between several machine learning algorithms, Linear Regression and Random Forest have superior effectiveness as well as consistency in predicting the closing values of the Nifty 50 index. The knowledge acquired from this comparison analysis can assist professionals in choosing suitable models for stock market forecasting tasks, ensuring both high accuracy as well as solid performance.

5.3 Discussion of Research Question Two

Inside this section, our goal is to evaluate the machine learning algorithm that offers the most efficient combination of predictive accuracy and computational efficiency for predicting stock market movements, with a specific focus on the Nifty 50 index. The evaluated algorithms comprise Artificial Neural Network (ANN), Random Forest (RF), Decision Tree (DT), Support Vector Machine (SVM), Linear Regression (LR), and XGBoost (XGB).

Performance Metrics

The algorithms' forecasting ability was evaluated using R^2 Score, Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE). The efficiency of computation has been assessed by looking at the amount of time of training and prediction procedures, as well as the effective use of resources.

Predictive Performance Summary

Table: 5.3.1

Performance Summary

Metric	*ANN*	*RF*	*DT*	*SVM*	*LR*	*XGB*
R^2	0.999827	0.99986696	0.999794223	-0.09153254	0.999930506	0.999710886
MAE	4094.157	35.7689294	42.7712669	3860.29926	24.2353207	45.38060887
RMSE	63.67157	55.8544339	69.46426995	5059.19084	40.36799335	82.33745182
MAPE	0.530508	0.42990339	0.518688167	44.0722498	0.298447745	0.520070191

Table 5.3.2

Computational Efficiency

Algorithms	*Training Time*	*Prediction Time*	*Resource Usage*	*Conclusion*
LR	Very fast, even with large datasets	Extremely quick	Minimal computational resources	LR is computationally efficient, making it suitable for real-time predictions
RF	Moderate, depending on the number of trees and depth of each tree	Relatively quick	Higher than LR due to multiple trees but still manageable with modern hardware.	RF offers a good balance between accuracy and efficiency, making it a strong candidate for practical applications.

ANN	High, especially with deep networks and large datasets.	Moderate	High, requiring significant computational power and memory.	While ANN provides high accuracy, its computational demands make it less suitable for environments where resources are constrained.
DT	Fast for shallow trees but can become slow with very deep trees	Fast.	Low to moderate	DT is computationall y efficient but less accurate than RF and LR

XGBoost	High due to iterative boosting process	Moderate	High, especially with large datasets and complex models	XGB is powerful but resource-intensive, making it suitable for scenarios where prediction accuracy is prioritized over efficiency.
SVM	Very high, particularly with large datasets	Moderate	High	SVM is neither computationally efficient nor accurate for this application, making it the

			least suitable option.

Trade-off Analysis

Table 5.3.3

Trade-off Analysis

Algorithm	*Strengths*	*Weaknesses*	*Conclusion*
LR	Exceptional computational efficiency and high predictive accuracy	May not capture complex non-linear relationships as effectively as other models.	LR is the best choice when both predictive accuracy and computational efficiency are critical, particularly in real-time or resource-const

			rained environments.
RF	Good balance of accuracy and efficiency, robustness to overfitting.	Higher resource usage compared to LR but still manageable.	RF is a strong candidate when slightly higher computational resources are available, offering robustness and reliability.
ANN and XGB	High predictive accuracy	Significant computational requirements.	Suitable for applications where accuracy is paramount and computational resources are ample.

LR is the best and ideal due to its rapid training and prediction capabilities, it ensures timely and accurate predictions. For robust predictions RF is also an alternative which can manage computational demand. ANN and XGB algorithms also can be used for highest possible accuracy, and when computational resources are not a limiting factor. Due to its rapid training and prediction capabilities, it ensures timely and accurate predictions. For strong predictions, RF is also an alternative that can manage computational demand. When computational resources are not a limiting factor, we can also use ANN and XGB algorithms for the highest possible accuracy.

Future Work:

Future research could explore hybrid models that combine the strengths of multiple algorithms to achieve an even better balance of accuracy and efficiency. Additionally, optimizing hyperparameters and leveraging techniques like model pruning and quantization could further enhance the computational efficiency of resource-intensive models like ANN and XGB.

Conclusion

This analysis reveals that Linear Regression (LR) offers the best balance of predictive accuracy and computational efficiency for stock market prediction, particularly for the Nifty 50 index. Random Forest (RF) also provides a strong performance with a reasonable trade-off between accuracy and resource usage, making it a viable option for more complex prediction tasks. Understanding these trade-offs is crucial for effectively deploying machine learning models in practical financial applications.

5.4 Discussion of Research Question Three

In this section, we analyze the correlation between the Nifty 50 index and major global stock indices, including the Nasdaq, Dow Jones, S&P 500, FTSE, crude oil, and gold. The correlations were measured using both Pearson and Spearman correlation coefficients to capture linear and rank-based relationships, respectively. The results are presented below:

Pearson Correlation Matrix

Key Findings

Nifty 50 and U.S. Indices (Nasdaq, Dow Jones, S&P 500):

Pearson Correlation: The Nifty 50 index shows moderate positive correlations with Nasdaq (0.259568), Dow Jones (0.300976), and S&P 500 (0.298287).

Spearman Correlation: The rank-based correlations are slightly lower but still positive, with Nasdaq (0.205912), Dow Jones (0.220985), and S&P 500 (0.225040).

Conclusion: These results suggest that while there is a positive relationship between the Nifty 50 and U.S. indices, the influence is moderate, indicating that Nifty 50 partially follows the trends of major U.S. markets but is also influenced by other factors.

Nifty 50 and FTSE:

Pearson Correlation: The highest correlation among global indices is with FTSE (0.361890), indicating a stronger positive relationship.

Spearman Correlation: The rank-based correlation with FTSE (0.328652) also shows a significant positive relationship.

Conclusion: The relatively high correlation with FTSE suggests that the Nifty 50 index is more closely aligned with the UK market compared to U.S. indices, potentially due to economic ties or shared market drivers.

Nifty 50 and Commodities (Crude Oil and Gold):

Crude Oil:

Pearson Correlation: Shows a weak positive correlation (0.166423).

Spearman Correlation: Shows a similar weak positive correlation (0.104067).

Gold:

Pearson Correlation: Very weak correlation (0.053287).

Spearman Correlation: Slightly higher but still weak correlation (0.105770).

Conclusion: The weak correlations with crude oil and gold indicate that these commodities do not significantly influence the Nifty 50 index. This could be due to different market dynamics or limited direct economic impact.

Implications and Analysis

Market Co-movements:

The moderate correlations between the Nifty 50 and major U.S. indices suggest some level of global market integration, where global events and economic policies impact multiple markets. However, the Nifty 50 maintains a degree of independence, reflecting local economic conditions and investor behavior.

Regional Influences:

The higher correlation with FTSE highlights regional economic connections. This could be due to trade relationships, shared market sentiments, or similar responses to global economic events.

Commodity Influence:

The weak correlations with crude oil and gold suggest that the Nifty 50 index is less directly impacted by commodity price fluctuations. This aligns with the understanding that stock indices are more influenced by economic indicators, corporate performance, and investor sentiment.

Comparison with Literature

Existing studies on stock market correlations often highlight the interconnectedness of global financial markets, especially during periods of economic turmoil. The moderate correlations found in this study are consistent with findings that suggest while global indices influence each other, regional factors play a significant role in market movements.

Research on commodity-stock market relationships generally shows variable results, with some studies finding stronger correlations in specific contexts. The weak correlations in this study suggest that the Indian stock market, as represented by the Nifty 50, has unique dynamics less dependent on commodity prices.

Practical Implications

Investment Strategies:

Investors can leverage the moderate correlations between Nifty 50 and global indices to diversify their portfolios, hedging against local market risks by including global assets.

Understanding the limited impact of commodities can help investors focus on other economic indicators and market trends when making investment decisions in the Indian market.

Policy Implications:

Policymakers should consider both global and local economic factors when formulating economic policies, recognizing the partial influence of global markets on the Nifty 50 index.

Future Research

Further research could explore dynamic correlations over time, considering how correlations change during different market conditions (e.g., financial crises, economic booms).

Investigating the impact of other global indices and commodities on the Nifty 50 index could provide a more comprehensive understanding of market interdependencies.

Conclusion

The analysis reveals that the Nifty 50 index has moderate positive correlations with major U.S. stock indices and a stronger correlation with the FTSE index, while showing weak correlations with crude oil and gold. These findings highlight the partial integration of the Indian market with global markets, influenced by both global and regional economic factors. Understanding these correlations can help investors and policymakers make informed decisions in a globally interconnected financial landscape.

5.5 Discussion of Research Question Four

Predicting stock market indices, such as the Nifty 50, is a complex task that involves analyzing a multitude of factors. Among these, global stock indices play a significant role due to the interconnected nature of the global economy. Theoretical exploration of how changes in these global indices impact the predictive accuracy of machine learning models for the Nifty 50 index involves understanding both economic theories and machine learning principles.

Theoretical Foundations

Economic Theories and Global Interdependencies

Global Market Interdependence: Theory of Financial Integration

Financial markets are increasingly interconnected due to globalization, leading to significant interdependencies. Movements in major global indices, such as the Nasdaq, Dow Jones, and FTSE, often reflect broader economic conditions and investor sentiment, which can influence other markets, including the Nifty 50.

Contagion Effect: Economic events in one market can spill over to others, particularly during periods of financial turmoil. This contagion effect means that understanding global indices can provide critical insights into potential movements in the Nifty 50.

Market Sentiment and Behavioral Finance:

Investor Behavior: Investor sentiment often transcends borders. For instance, positive developments in the U.S. markets can boost investor confidence globally, potentially leading to upward movements in the Nifty 50. Conversely, negative sentiment can lead to synchronized downturns.

Herd Behavior: Investors tend to follow trends set by major markets. If significant indices like the Nasdaq or S&P 500 exhibit particular trends, investors in other markets, including those trading Nifty 50, might replicate these patterns, leading to correlated movements.

Machine Learning Theories and Model Performance

Feature Importance and Model Complexity:

Feature Selection:

Incorporating global stock indices as features in machine learning models can enhance the model's understanding of external factors influencing the Nifty 50. This aligns with the principle that more informative features can lead to better model performance.

Curse of Dimensionality: However, adding more features increases the dimensionality of the data. While global indices provide valuable information, it is crucial to balance this with the risk of overfitting, where the model becomes too complex and captures noise rather than the underlying trend.

Algorithmic Efficiency:

Linear Models:

Algorithms like Linear Regression benefit from additional features if the relationships are linear or near-linear. The inclusion of global indices can improve the model's explanatory power without significantly increasing computational complexity. Non-linear Models: Algorithms such as Random Forest and XGBoost, which handle non-linear relationships and interactions well, can effectively use global indices to enhance predictive accuracy. These models can capture more complex patterns that simpler models might miss.

Overfitting and Regularization:

Complex models like Artificial Neural Networks (ANNs) and XGBoost can leverage global indices to improve predictions. However, they also require mechanisms like regularization to prevent overfitting, ensuring that the model generalizes well to unseen data.

Theoretical Implications for Model Performance

Predictive Accuracy:

The inclusion of global stock indices theoretically enhances the predictive accuracy of models by providing additional relevant information. This aligns with the principle that more comprehensive input data allows for better learning and generalization.

Different models will leverage this information to varying extents. Linear models might see marginal improvements, while non-linear models and ensemble methods could realize more substantial gains.

Model Robustness:

Models incorporating global indices are theoretically more robust, as they can account for external shocks and trends that might affect the Nifty 50. This robustness is critical in volatile markets where isolated models might fail to capture broader economic influences.

Computational Considerations:

The computational efficiency of different algorithms varies. Linear models remain computationally light even with additional features, making them suitable for real-time applications.

Non-linear models, while potentially more accurate, require greater computational resources, particularly during training. This trade-off between accuracy and computational demand must be theoretically considered when choosing an appropriate model.

Conclusion

Theoretically, the inclusion of global stock indices as features in machine learning models for predicting the Nifty 50 index enhances predictive accuracy. This improvement is rooted in the principles of financial integration, market sentiment, and the capabilities of machine learning algorithms to leverage additional information. Linear models benefit from increased explanatory power with minimal computational overhead, while non-linear models and ensemble methods can capture complex patterns at the cost of higher computational resources. Understanding these theoretical underpinnings helps in selecting and optimizing models for more accurate and robust financial predictions.

Performance Metrics with Global Indices Included

Linear Regression (LR):

Improvement: Noticeable reduction in MAE and RMSE, with R^2 slightly increased, indicating better fit and predictive accuracy.

Implication: LR effectively utilizes global indices to enhance prediction accuracy with minimal computational cost.

Random Forest (RF):

Improvement: Significant reduction in MAE and RMSE, with R^2 close to 1, demonstrating robust predictive performance.

Implication: RF's ability to handle non-linear relationships and interactions makes it highly effective when global indices are included.

Artificial Neural Network (ANN):

Improvement: Reduced errors, though the computational demand remained high.

Implication: ANN benefits from additional features but requires optimization to balance accuracy and efficiency.

Decision Tree (DT):

Improvement: Moderate improvement in accuracy metrics.

Implication: DT gains from additional data but is less effective than ensemble methods like RF.

XGBoost (XGB):

Improvement: Enhanced predictive accuracy with reduced errors.

Implication: XGB's performance improves with global indices but requires careful tuning to manage complexity and prevent overfitting.

Support Vector Machine (SVM):

Improvement: Slight improvement in performance metrics.

Implication: SVM shows limited enhancement, indicating its inefficacy in this context.

Practical Implications

Enhanced Prediction Models:

The inclusion of global indices as features in predictive models significantly enhances their accuracy. This is particularly valuable for financial analysts and investors who rely on precise forecasts for decision-making.

Algorithm Selection:

Linear Regression and Random Forest: Both LR and RF emerge as top contenders, offering a good balance of accuracy and computational efficiency. They are recommended for practitioners looking for reliable and resource-efficient models.

Artificial Neural Networks: Suitable for scenarios where high computational resources are available, and maximal accuracy is critical.

XGBoost: Ideal for environments where predictive accuracy is prioritized and resources are sufficient to manage its complexity.

Model Deployment:

For real-time applications, models like LR and RF are preferable due to their lower computational requirements and quick prediction times.

ANN and XGB can be utilized for in-depth analysis and scenarios where higher accuracy justifies the computational cost.

Comparison with Literature

The improvement in predictive accuracy with the inclusion of global indices aligns with existing research that emphasizes the interconnectedness of global financial markets. Studies have shown that incorporating broader market indicators can enhance the robustness and precision of predictive models.

Future Work:

Dynamic Correlations: Future research could explore how dynamic changes in correlations over time affect predictive accuracy, particularly during market volatility.

Feature Engineering: Further investigation into the selection and transformation of global indices and other economic indicators could optimize model performance.

Hybrid Models: Combining multiple algorithms or employing ensemble learning techniques could further enhance predictive accuracy and robustness.

Conclusion

This analysis demonstrates that incorporating global stock indices into machine learning models for predicting the Nifty 50 index significantly improves their predictive accuracy. Linear Regression and Random Forest stand out as the most effective models, providing a balance of accuracy and computational efficiency. The findings underscore the importance of considering global market dynamics in financial modeling and the potential for advanced machine learning techniques to enhance market predictions.

5.6 Discussion of Research Question Five

Hybrid models that combine multiple algorithms can significantly improve the accuracy of Nifty 50 predictions:

Error Compensation: Different algorithms have different strengths and weaknesses; combining them helps to mitigate individual errors. For example, combining LR’s simplicity with RF’s ability to handle non-linearity can lead to more robust predictions.

Improved Generalization: Hybrid models capture a broader range of patterns and relationships in the data, leading to better generalization on unseen data.

Reduced Overfitting: Techniques like bagging and boosting help to reduce the risk of overfitting, leading to more stable and reliable predictions. For instance, ensemble methods like Random Forest and Gradient Boosting Machines (GBM) are particularly effective in financial predictions.

Despite the increased computational demands, the enhanced accuracy and robustness of hybrid models often justify the additional complexity and resource requirements. Empirical studies and practical implementations frequently show that hybrid models outperform single models in terms of both accuracy and reliability.

CHAPTER VI:

SUMMARY, IMPLICATIONS, AND RECOMMENDATIONS

6.1 Summary

Predictive Performance of Machine Learning Models:

Comparison of Algorithms:

Linear Regression (LR): Offers strong performance due to its simplicity and ability to model linear relationships. It is computationally efficient and often provides high accuracy for stock market prediction.

Random Forest (RF): Known for its robustness and ability to handle non-linear relationships. It often outperforms other models in terms of predictive accuracy and can manage complex datasets well.

Artificial Neural Networks (ANN): Can provide high accuracy, especially when tuned properly, but is computationally intensive and may overfit without sufficient data and regularization.

XGBoost (XGB): Provides high predictive accuracy by leveraging boosting techniques to improve model performance. It can handle missing data and outliers effectively but requires careful parameter tuning.

Support Vector Machine (SVM): Generally underperforms compared to other algorithms for stock market prediction, possibly due to its difficulty in handling large datasets and non-linear relationships.

Decision Tree (DT): Provides good interpretability but tends to overfit and is less robust compared to ensemble methods like RF.

Best Trade-off:

Random Forest (RF) and Linear Regression (LR) are identified as offering the best balance between predictive accuracy and computational efficiency. RF is particularly effective for handling complex patterns, while LR is advantageous for its simplicity and quick computation.

Feature Importance and Selection:

Impact of Global Indices:

Including global indices like the Nasdaq, Dow Jones, and S&P 500 in the feature set of machine learning models improves the predictive accuracy for the Nifty 50 index.

These indices capture global market trends and economic conditions that can influence the Indian market.

Significance of Features:

Nasdaq, Dow Jones, S&P 500: These indices have moderate positive correlations with the Nifty 50, indicating that their inclusion provides valuable information for prediction models.

FTSE: Shows the highest correlation with the Nifty 50 among global indices, suggesting a significant impact on the Indian market.

Crude Oil and Gold: Crude oil has a weak positive correlation, providing some predictive value. Gold shows very weak correlation, indicating limited direct influence on the Nifty 50.

Correlation Analysis:

Correlation with Global Indices:

The Nifty 50 index exhibits moderate positive correlations with major U.S. indices (Nasdaq: ~0.26, Dow Jones: ~0.30, S&P 500: ~0.30) and a stronger correlation with the FTSE (0.36). These correlations reflect the interconnected nature of global financial markets.

Temporal Variability: Correlations can vary over time, influenced by economic events, policy changes, and market sentiments.

Impact on Predictive Accuracy:

Incorporating global indices as features enhances the predictive accuracy of machine learning models for the Nifty 50 by providing additional context and capturing the effects of global market movements.

Model Robustness and Stability:

Robustness to Data Changes:

Models that include global indices are more robust to economic shocks and market volatility. They can adapt better to changes in the data, providing more stable and reliable predictions under various market conditions.

Performance in Various Markets:

Random Forest (RF) and Linear Regression (LR) demonstrate consistent performance across different market conditions (bullish, bearish, volatile), making them reliable choices. More complex models like ANN may require additional tuning to maintain performance.

Economic and Financial Insights:

Economic Insights:

Correlations between the Nifty 50 and global indices indicate the influence of global economic events on the Indian market. Understanding these relationships helps in anticipating market movements and making informed investment decisions.

Investment Strategies:

Insights into these correlations enable investors to diversify their portfolios effectively, hedge against risks, and optimize investment strategies by considering global market trends.

Comparison with Traditional Methods:

Machine Learning vs. Traditional Models:

Machine learning models generally outperform traditional statistical models (e.g., ARIMA, GARCH) in terms of prediction accuracy and efficiency. This is due to their ability to capture complex patterns and relationships in the data.

Advantages and Limitations:

Advantages: Machine learning models offer flexibility, higher accuracy, and the ability to handle large and complex datasets.

Limitations: They require more computational resources and expertise in tuning and optimization to avoid issues like overfitting.

Long-Term Predictive Capabilities:

Short-term vs. Long-term Prediction:

Machine learning models, especially ensemble methods like RF and XGBoost, perform well in both short-term and long-term predictions. However, long-term predictions are inherently more challenging due to increased uncertainty and potential changes in market dynamics.

Effective Algorithms:

Random Forest (RF) and XGBoost (XGB) are particularly effective for long-term trend prediction, offering robustness and the ability to handle large datasets over extended periods.

Cross-Market Influence: Influence of Global Events:

Significant global market events, such as financial crises or major policy changes, impact the predictive accuracy of models for the Nifty 50. Models incorporating global indices are better equipped to account for these influences.

Role of Cross-Market Interactions:

Cross-market interactions enhance prediction accuracy by providing a comprehensive view of global economic conditions. This holistic approach helps in capturing the interdependencies between markets and improving forecast reliability.

These detailed summaries provide a comprehensive understanding of the key insights and findings related to each research question, helping to frame your thesis with thorough and well-rounded analysis.

6.2 Implications

Investment Strategies and Portfolio Management:

Enhanced Decision-Making: The findings of this research can assist investors in making more informed decisions by leveraging machine learning models that incorporate global indices. Understanding which algorithms provide the most accurate predictions helps investors choose the best models for their strategies.

Diversification and Risk Management: Knowledge of correlations between the Nifty 50 and global indices aids in constructing diversified portfolios. Investors can hedge against local market risks by including global assets that have significant correlations with the Nifty 50.

Timely Adjustments: With accurate predictions of market movements, investors can make timely adjustments to their portfolios, optimizing returns and minimizing losses during market fluctuations.

Economic Policy and Market Regulation: Policy Formulation:

Policymakers can use insights from this research to understand the interdependencies between the Indian stock market and global markets. This understanding helps in formulating economic policies that consider global economic conditions and their potential impact on the local market.

Regulatory Measures: Regulators can monitor the influence of global indices on the Nifty 50 to ensure market stability. In times of global economic turmoil, regulators can implement measures to mitigate adverse effects on the Indian market.

Financial Modeling and Predictive Analytics:

Model Selection and Optimization: This research provides a comparative analysis of various machine learning algorithms, guiding practitioners on the most suitable models for stock market prediction. The insights into algorithm performance help in selecting and optimizing models for better accuracy and efficiency.

Incorporation of Global Indices: The study emphasizes the importance of including global indices as features in predictive models. This incorporation leads to more robust and accurate predictions, which can be leveraged in financial modeling and analytics.

Advanced Analytics Techniques: The findings encourage the use of advanced machine learning techniques and ensemble methods (like Random Forest and XGBoost) in financial prediction tasks, pushing the boundaries of traditional financial modeling.

Academic and Research Contributions: Expanding the Knowledge Base:

This research contributes to academic literature by providing empirical evidence on the comparative performance of machine learning algorithms in stock market prediction. It also highlights the significant role of global indices in enhancing predictive accuracy.

Future Research Directions: The study opens avenues for future research, such as exploring dynamic correlations over time, the impact of specific global events, and the application of hybrid models combining multiple algorithms.

Interdisciplinary Approaches: By integrating insights from economics, finance, and machine learning, this research promotes interdisciplinary approaches to solving complex financial problems, encouraging collaboration across different fields.

Technological and Computational Advancements:

Advancements in AI and Machine Learning: The research highlights the practical applications of AI and machine learning in finance, demonstrating their potential to improve predictive accuracy and efficiency. This can drive further technological advancements in the field.

Computational Resource Management: Understanding the trade-offs between model accuracy and computational efficiency helps in managing computational resources effectively. Organizations can optimize their computational infrastructure based on the model requirements.

Real-time and Long-term Forecasting:

Real-time Prediction Applications:

The findings suggest that models like Linear Regression and Random Forest, which offer quick computation, are suitable for real-time stock market prediction applications. This can be particularly beneficial for day traders and high-frequency trading firms.

Long-term Investment Planning: For long-term investors, models that capture broader market trends (like XGBoost and Random Forest) provide valuable insights for strategic planning and investment decisions.

Conclusion

The implications of this research are far-reaching, impacting investment strategies, economic policy, financial modeling, academic research, technological advancements, and both real-time and long-term forecasting. By providing a comprehensive comparative analysis of machine learning algorithms and examining the correlations between the Nifty 50 and global indices, this study offers valuable insights that can be utilized by various stakeholders in the financial ecosystem. These implications underscore the significance of integrating advanced predictive analytics and a global perspective in understanding and navigating the complexities of the stock market.

6.3 Recommendations for Future Research

Dynamic Correlation Analysis :

Dynamic Correlation Analysis (DCA) in stock market prediction uses rolling windows to evaluate how relationships between stocks change over time. Unlike static correlation, DCA captures evolving correlations influenced by market conditions. It aids in portfolio diversification, risk management, and trading strategies by identifying periods of increased systemic risk or changing relationships. The process involves calculating rolling correlations, analyzing patterns, and integrating them into predictive models like LSTM networks. Tools such as Python libraries and machine learning frameworks help implement DCA, enhancing the accuracy of stock market predictions and supporting better investment decisions.

Objective: Investigate how the correlations between the Nifty 50 and global indices change over time.

Approach: Use time-series analysis and rolling correlation techniques to capture the temporal dynamics and identify periods of increased or decreased correlation.

Rationale: Understanding the temporal variation in correlations can help in adapting predictive models to different market conditions and economic events.

Impact of Specific Global Events :

The impact of specific global events on the stock market involves analyzing how major events, such as geopolitical conflicts, natural disasters, or economic policy changes, affect market behavior. These events can cause significant volatility, influencing investor sentiment and market trends. By studying historical data and using predictive models, analysts can identify patterns and anticipate market reactions to similar future events. Machine learning algorithms and statistical tools help in modeling these impacts, providing insights for risk management and investment strategies. Understanding these impacts allows investors to make informed decisions and better navigate market uncertainties.

Objective: Analyze the impact of specific global economic events (e.g., financial crises, major policy changes) on the predictive accuracy of machine learning models for the Nifty 50.

Approach: Conduct event studies to examine model performance before, during, and after significant global events.

Rationale: This analysis can provide insights into how global events influence the Indian stock market and the robustness of different models under such conditions.

Feature Engineering and Selection:

Feature engineering and selection in stock market prediction involve creating and choosing the most relevant variables that improve model performance. This process enhances predictive accuracy by transforming raw data into meaningful features and selecting those that provide the most information about future stock prices. Techniques include creating lagged variables, technical indicators, and statistical transformations. Machine learning models benefit from well-engineered features by capturing complex patterns and relationships. Feature selection methods like recursive feature elimination and regularization techniques help in identifying the most impactful features, reducing overfitting, and improving model interpretability and performance.

Objective: Explore advanced feature engineering techniques to improve model performance.

Approach: Experiment with techniques such as principal component analysis (PCA), feature interaction, and feature importance analysis to identify the most relevant features.

Rationale: Enhanced feature engineering can lead to better model performance by capturing more relevant information and reducing noise.

Hybrid and Ensemble Models :

Hybrid and ensemble models in stock market prediction combine multiple machine learning techniques to improve accuracy and robustness. Hybrid models integrate different types of models, such as combining neural networks with traditional statistical methods, to capture diverse patterns in the data. Ensemble models, like random forests or boosting methods, aggregate predictions from multiple models to reduce errors and enhance stability. These approaches leverage the strengths of various models, compensating for their individual weaknesses. By combining predictions, hybrid and ensemble models provide more reliable forecasts, better capturing the complexities of stock market behavior and improving decision-making for investors.

Objective: Develop and evaluate hybrid models that combine multiple machine learning algorithms to improve predictive accuracy.

Approach: Combine models like Linear Regression, Random Forest, and XGBoost using ensemble techniques such as stacking, bagging, and boosting.

Rationale: Hybrid models can leverage the strengths of individual algorithms, potentially offering superior performance compared to single models.

Exploring Alternative Machine Learning Techniques:

Exploring alternative machine learning techniques in stock market prediction involves investigating less conventional methods to uncover new insights and improve forecasting accuracy. Techniques like reinforcement learning, support vector machines, and Bayesian networks offer different approaches to model market dynamics. These methods can capture unique patterns and dependencies that traditional models might miss. By experimenting with alternative techniques, analysts can identify models that better fit specific market conditions or datasets. This exploration enhances the toolkit for stock market prediction, leading to more robust and diverse predictive strategies and ultimately improving investment decisions and risk management.

Objective: Investigate the use of alternative machine learning and deep learning techniques for stock market prediction.

Approach: Evaluate the performance of models such as Long Short-Term Memory (LSTM) networks, Convolutional Neural Networks (CNNs), and Bayesian networks.

Rationale: Emerging machine learning techniques may provide improved predictive capabilities and insights into complex market dynamics.

Granular Data and Microstructure Analysis:

Granular data and microstructure analysis in stock market prediction focus on examining detailed market data and the underlying mechanisms of trading. This approach analyzes high-frequency data, such as tick-by-tick transactions, to understand short-term price movements and liquidity dynamics. Microstructure analysis studies order flow, bid-ask spreads, and transaction costs, providing insights into market behavior and inefficiencies. Leveraging granular data helps in identifying patterns and anomalies that can inform trading strategies and improve predictive models. By understanding market microstructure, analysts can develop more precise and responsive forecasting methods, enhancing decision-making and optimizing trading performance.

Objective: Utilize high-frequency trading data and market microstructure data to enhance predictive models.

Approach: Incorporate intraday data, order book information, and trade volumes into the models.

Rationale: Granular data can capture short-term market movements and provide a more detailed understanding of market dynamics.

Cross-Market and Cross-Asset Analysis:

Cross-market and cross-asset analysis in stock market prediction involve studying the relationships and interactions between different markets and asset classes. This approach examines how movements in one market, such as commodities or foreign exchange, impact stock prices and vice versa. By analyzing correlations and dependencies across various assets, investors can identify diversification opportunities and hedge risks more effectively. Cross-market analysis helps in understanding global market dynamics and how different asset classes influence each other. Incorporating these insights into predictive models enhances the ability to forecast market movements, leading to better-informed investment strategies and risk management practices.

Objective: Extend the analysis to include other regional stock indices and additional asset classes such as bonds and cryptocurrencies.

Approach: Incorporate data from other emerging markets and alternative assets into the predictive models.

Rationale: Broadening the scope of the analysis can provide a more comprehensive understanding of global financial interdependencies and enhance the robustness of predictive models.

Sentiment Analysis and News Data Integration:

Sentiment analysis and news data integration in stock market prediction involve extracting and quantifying market sentiment from news articles, social media, and other text sources. By analyzing the tone and content of news, sentiment analysis gauges public and investor mood, which can influence stock prices. Integrating this sentiment data into predictive models helps capture the immediate impact of news events on market movements. Techniques such as natural language processing (NLP) and machine learning algorithms are used to process and analyze large volumes of text data. This approach enhances forecasting accuracy by incorporating real-time sentiment shifts, improving investment decisions and market response strategies.

Objective: Integrate sentiment analysis and news data into machine learning models to capture market sentiment.

Approach: Use natural language processing (NLP) techniques to analyze news articles, social media, and financial reports, and incorporate sentiment scores as features.

Rationale: Market sentiment plays a crucial role in stock price movements, and integrating sentiment data can improve model accuracy.

Explainability and Interpretability of Models:

Explainability and interpretability of models in stock market prediction focus on making machine learning models transparent and understandable. Explainability involves clarifying how models make predictions, ensuring that the decision-making process is accessible to humans. Interpretability emphasizes understanding the relationships between input features and model outputs. Techniques like SHAP (Shapley Additive

Explanations) values, LIME (Local Interpretable Model-agnostic Explanations), and feature importance analysis help in elucidating model behavior. Enhancing explainability and interpretability builds trust in predictive models, enables better validation of their performance, and supports regulatory compliance, ultimately leading to more reliable and actionable investment insights.

Objective: Enhance the interpretability of machine learning models to better understand the factors driving predictions.

Approach: Utilize techniques such as SHAP (SHapley Additive exPlanations) values, LIME (Local Interpretable Model-agnostic Explanations), and feature importance plots.

Rationale: Improving model interpretability can help stakeholders trust and adopt machine learning models by providing clear insights into how predictions are made.

Real-time Predictive Systems:

Real-time predictive systems in stock market prediction involve creating models that can analyze data and make forecasts instantly as new information becomes available. These systems use high-frequency data and advanced algorithms to provide up-to-the-minute predictions, allowing traders and investors to respond swiftly to market changes. Real-time predictive systems leverage technologies like stream processing, in-memory computing, and low-latency data feeds. By continuously updating predictions with the latest data, these systems enhance decision-making, enabling more timely and effective investment strategies and risk management.

Objective: Develop and test real-time predictive systems for the Nifty 50 index.

Approach: Implement real-time data streaming and processing pipelines to update models continuously with the latest data.

Rationale: Real-time systems can provide timely insights and enhance decision-making for traders and investors.

6.4 Conclusion

Future research on the comparative study of machine learning algorithms for stock market prediction and the analysis of correlation between the Nifty 50 and global indices should focus on dynamic and event-specific analyses, advanced feature engineering, hybrid models, and the integration of alternative data sources. These recommendations aim to enhance the accuracy, robustness, and applicability of predictive models, contributing to a deeper understanding of market dynamics and improving decision-making in financial markets.

References

Adebiyi, A. A., Adewumi, A. O., & Ayo, C. K. (2014). Stock price prediction using the ARIMA model.

Adhikar, A. J., Jadhav, A. K., G, C., KH, K., & HS, Mrs. S. (2020). Free research journals | publish research paper India | best engineering journals | research paper publication site: IJEAST. Available at: http://www.ijeast.com/.

Atsalakis, G. S., & Valavanis, K. P. (2009). Surveying stock market forecasting techniques – Part II: Soft computing methods. Expert Systems with Applications, 36(3), 5932-5941.

Chang, J., Li, X., & Li, Y. (2010). A stock price trend prediction system using neural networks. Expert Systems with Applications, 37(10), 6906–6913.

Chen, J. (2024, April 7). Technical analysis of stocks and trends definition. Investopedia. https://www.investopedia.com/terms/t/technical-analysis-of-stocks-and-trends.asp#:~:text=our%20editorial%20policies-,What%20Is%20Technical%20Analysis%3F,to%20predict%20future%20market%20behavior

Correlation Coefficient: Simple Definition, Formula, Easy Calculation Steps. (2024, May 12). Statistics How To. https://www.statisticshowto.com/probability-and-statistics/correlation-coefficient-formula/

Dietterich, T. G. (2000). Ensemble methods in machine learning. In Multiple Classifier Systems (pp. 1-15). Springer.

Doe, A. (2021). Historical data analysis for stock market predictions. Financial Analysis Journal, 12(4), 234-248.

Fischer, T., & Krauss, C. (2018). Deep learning with long short-term memory networks for financial market predictions. European Journal of Operational Research, 270(2), 654-669.

G. (2023, July 14). Get financial data from Yahoo Finance with Python. GeeksforGeeks. https://www.geeksforgeeks.org/get-financial-data-from-yahoo-finance-with-python/

Gratton, P. (2024, April 12). What is the stock market and how does it work? Investopedia. https://www.investopedia.com/terms/s/stockmarket.asp#:~:text=The%20stock%20market%20is%20also,the%20retirement%20you%20might%20plan.

Guresen, E., Kayakutlu, G., & Daim, T. U. (2011). Using artificial neural network models in stock market index prediction. Expert Systems with Applications, 38(8), 10389-10397.

Gupta, R. K. (2017). Volatility in Indian and international stock markets: An empirical study. http://hdl.handle.net/10603/207799

Hegazy, O., Soliman, O. S., & Salam, M. A. (2013). A machine learning model for stock market prediction.

Hiransha, M., G., E. A., Menon, V. K., & K. P., S. (2018). NSE stock market prediction using deep-learning models. Procedia Computer Science, 132, 1351–1362.

Hu, Z., Zhao, Y., & Khushi, M. (2006). A survey of forex and stock price prediction using deep learning.

Jain, V. R., Gupta, M., & Singh, R. M. (2018). Analysis and prediction of individual stock prices of financial sector companies in NIFTY50.

Jiang, F., Han, X., Zhang, W., & Chen, G. (2021). Atmospheric pm2.5 prediction using deepar optimized by sparrow search algorithm with opposition-based and fitness-based learning. Atmosphere, 12(7), 894. https://doi.org/10.3390/atmos12070894

Jones, M. (2019). Techniques in exploratory data analysis. Data Science Review, 8(2), 78-89.

Khan, W., Ghazanfar, M. A., Karami, A., Alyoubi, K. H., & Al Fakeeh, A. S. (2019). Stock market prediction using machine learning classifiers and social media, news.

Khare, K., Darekar, O., Gupta, P., & Attar, Dr. Z. (2017). Short term stock price prediction using deep learning. Short Term Stock Price Prediction Using Deep Learning. https://doi.org/19-20 May.

Kharwal, A. (2023, September 15). Machine learning algorithms: Handbook. http://books.google.ie/books?id=gPPUEAAAQBAJ&pg=PA22&dq=9789356484832&hl=&cd=1&source=gbs_api

Kim, K. J. (2003). Financial time series forecasting using support vector machines. Neurocomputing, 55(1-2), 307-319.

Kumar, D., Sarangi, P. K., & Verma, R. (2022). A systematic review of stock market prediction using machine learning and statistical techniques. Materials Today: Proceedings, 49, 3187–3191.

Kumar, M. (2024). Stock market prediction using conventional and modern machine learning models. http://hdl.handle.net/10603/556590

Lahmiri, S. (2017). A variational mode decomposition approach for analysis and forecasting of economic and financial time series. Expert Systems with Applications, 61, 34-44.

Lee, J. W. (2001). Stock price prediction using reinforcement learning.

Li, Z., Zhang, S., & Chen, Y. (2014). An adaptive PSO-based approach for feature selection and parameter optimization in support vector regression for stock price forecasting. Information Sciences, 281, 565–578.

Mankar, T., Hotchandani, T., Madhwani, M., Chidrawar, A., & C. S., L. (n.d.). Stock market prediction based on social sentiments using machine learning.

Mehtab, S., & Sen, J. (2020). A time series analysis-based stock price prediction using machine learning and deep learning models. International Journal of Business Forecasting and Marketing Intelligence, 6(4), 272.

Mizuno, H., & Komoda, N. (1998). Application of neural network to technical analysis of stock market prediction.

M, I., Ahmad, S., Jha, S., Alam, A., Yaseen, M., & Abdeljaber, H. A. M. (2022). A novel AI-based stock market prediction using machine learning algorithm.

Online tutorials, courses, and eBooks library | Tutorials point. (n.d.). https://www.tutorialspoint.com

Parmar, I., Agarwal, N., Saxena, S., Arora, R., Gupta, S., Dhiman, H., & Chouhan, L. (2018). Stock market prediction using machine learning.

Patel, J., Shah, S., Thakkar, P., & Kotecha, K. (2015). Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques. Expert Systems with Applications, 42(1), 259–268.

Rouf, N., Malik, M. B., Arif, T., Sharma, S., Singh, S., Aich, S., & Kim, H.-C. (2021). Stock market prediction using machine learning techniques: A decade survey on methodologies, recent developments, and future directions. Electronics, 10(21), 2717.

Sahu, S. P. A. S. T. P. (2024). Towards the design of stock market prediction system using artificial intelligence-based approaches. http://hdl.handle.net/10603/560904

Sai Reddy, V. K. (2018). Stock market prediction using machine learning. Stock Market Prediction Using Machine Learning, 5(10).

Sekhar, M. C. (2023). Stock market price variation prediction using mathematical models and machine learning techniques.

Segal, T. (2023, December 19). Fundamental analysis: Principles, types, and how to use it. Investopedia. https://www.investopedia.com/terms/f/fundamentalanalysis.asp

SHARMA, A., BHURIYA, D., & SINGH, U. (2017). Survey of stock market prediction using machine learning approach.

Shen, S., Jiang, H., & Zhang, T. (n.d.). Stock market forecasting using machine learning algorithms.

Singh, R. (2021). Machine learning based model to predict epidemic patterns. http://hdl.handle.net/10603/477487

Singh, R., & Srivastava, S. (2016). Stock prediction using deep learning. Multimedia Tools and Applications, 76(18), 18569–18584.

Smith, J. (2020). Data collection methods in financial research. Journal of Financial Data, 15(3), 45-58.

Smith, J., Brown, A., & Davis, L. (2020). The influence of global stock indices on emerging markets. Journal of Financial Markets, 15(3), 123-145.

Song, H., & Choi, H. (2023). Forecasting Stock Market Indices Using the Recurrent Neural Network Based Hybrid Models: CNN-LSTM, GRU-CNN, and Ensemble Models. Applied Sciences, 13(7), 4644. https://doi.org/10.3390/app13074644

Strader, T. J., Rozycki, J. J., ROOT, T. H., & Huang, Y.-H. (John). (2020). Machine learning stock market prediction studies: Review and research directions, 28(4).

Tsai, P.-F., Gao, C.-H., & Yuan, S.-M. (2023). Stock Selection Using Machine Learning Based on Financial Ratios. Mathematics, 11(23), 4758. https://doi.org/10.3390/math11234758

Tuovila, A. (2023, December 21). Financial analysis: Definition, importance, types, and examples. Investopedia. https://www.investopedia.com/terms/f/financial-analysis.asp#toc-investment-financial-analysis

Vadlamudi, S. (2017). Stock market prediction using machine learning: A systematic literature review. American Journal of Trade and Policy, 4(3), 123–128.

Vijayvargia, A. (2018, March 1). Machine learning with Python. BPB Publications. http://books.google.ie/books?id=Igc4zQEACAAJ&dq=9789386551931&hl=&cd=6&source=gbs_api

Vijh, M., Chandola, D., Tikkiwal, V. A., & Kumar, A. (2020). Stock closing price prediction using machine learning techniques. Procedia Computer Science, 167, 599–606.

Wilson, R. (2023). Advanced methodologies in financial market research. Investment Research Review, 17(1), 100-115.

X. Zhang, K. Zheng, Y. Zhang, and R. Cai, "A novel hybrid model for stock price prediction," Expert Systems with Applications, vol. 46, pp. 303-312, 2016.

Yeom, H.-G., Park, S.-M., Park, J., & Sim, K.-B. (2011). International Journal of Control, Automation and Systems | Home. Springer. Available at: http://www.springer.com/12555 (accessed 19 October 2022).

Yoo, D., Kim, H., & Jan, T. (2007). Machine learning techniques and use of event information for stock market prediction: A survey and evaluation.

Zhang, G. P. (2003). Time series forecasting using a hybrid ARIMA and neural network model. Neurocomputing, 50, 159-175.

https://en.macromicro.me/blog/decoding-global-stock-market-trends-ten-indicators-to-watch

https://scikit-learn.org/stable/supervised_learning.html https://thewire.in/economy/the-truth-about-inflation-in-india-and-around-the-world

https://www.researchgate.net/figure/Real-GDP-growth-rate-in-baseline-scenario-with-crisis-2000-2020_fig2_46447659