JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

MACHINE LEARNING-BASED STOCK PREDICTION SYSTEM: A TECHNICAL APPROACH FOR PREDICTING STOCK PRICES

[1]Kavita Shirsat, [2]Prem Patel, [3]Harsh Chheda

[1]Professor, [2]Student, [3]Student
[1]Computer Science,
[1]Vidyalankar Institute of Technology, Mumbai, India

Abstract: This project aims to create a website that users a comprehensive solution for their stock market information and analysis needs. The website is built using data science and machine learning algorithms, providing users with real-time stock prices, historical data, and advanced analysis tools. By utilizing data visualization techniques, complex information becomes easily understandable, and users can explore and manipulate data using various tools

I. INTRODUCTION

The stock market can be an intricate system, posing challenges even for the most seasoned investors. Nonetheless, the advent of data science and machine learning has birthed an array of tools and techniques. These resources aid investors in making well-informed decisions and gaining a better understanding of the market's complexities. This venture showcases a website that leverages data science and machine learning algorithms to furnish users with powerful tools for comprehensive stock market analysis and comprehension. By assimilating and scrutinizing copious amounts of market data, this website offers real-time insights and forecasts on stock prices, trends, and other pivotal indicators. Designed to cater to investors with varying experience levels, the website boasts an intuitive interface and a diverse range of features and tools. These aids enable users to make wiser investment decisions and navigate the intricacies of the stock market effectively. From straightforward charts and graphs to cutting-edge machine learning models, the website provides an extensive array of resources. Users can capitalize on these tools to enhance their stock market understanding and bolster their investment judgments. In the forthcoming sections, we will provide an in-depth overview of the techniques and methodologies employed in developing this website. Additionally, we will present a thorough analysis of the findings and results obtained. We firmly believe that this website holds immense potential as a valuable resource for investors of all levels. We hope that this project will foster a deeper understanding of how data science and machine learning can aid in analysing and comprehending the stock market.

Key Aspects of the Project

1] **Analysis of Historical Data**: Examining the historical data related to stock prices and trading volume is a fundamental step in predicting stock data. Insights into potential future trends and patterns can be gained by studying past information.

2] Data Preprocessing: It is crucial to clean and preprocess the data. This involves addressing missing data, normalizing the information, and managing outliers to ensure the quality of input data for predictive models.

3] Selection of Features: Identifying the most significant indicators or features that impact stock prices is of utmost importance. Common features include price-to-earnings ratios, moving averages, trading volumes, and the sentiment of news.

4] Analysis of Time Series: Stock data often presents itself in the form of time series data. Techniques for time series analysis can prove beneficial in capturing temporal dependencies.

5] Utilization of Machine Learning Models: Employing machine learning techniques such as linear regression, decision trees, random forests, support vector machines, or neural networks is necessary for constructing predictive models.

In recent years, information technology has undergone significant advancements, which have greatly influenced the operations of businesses. Among these groundbreaking innovations, financial markets hold immense sway over the economy of nations. As per a report published by the World Bank in 2018, the global stock market capitalization has exceeded an impressive $68.654 trillion. This surge in stock trading activity can be primarily attributed to the progress made in technology. Investors now seek tools and techniques that can maximize profits and mitigate risks. However, predicting stock market trends, also known as stock market forecasting (SMP), is a complex endeavour due to its non-linear, dynamic, and stochastic nature, rendering it unreliable. SMP is an essential aspect of time prediction, as it swiftly analyses historical data to forecast future trends. The task of predicting financial markets has long been a subject of interest for analysts belonging to various disciplines such as economics, mathematics, information science, and computer science. Accurate prediction of the stock market is crucial for generating profits. There are numerous parameters on which the stock market relies, including share market value, performance of companies, government policies, gross domestic product (GDP) of a country, inflation rate, natural disasters, and more. With the advent of technology, researchers have demonstrated that it is possible to predict stock market prices to a certain extent by analysing historical market

data and inferring changes in the economic and business sectors. The effectiveness of stock market prediction systems largely depends on the quality of the features they employ. While some strategies have been employed by researchers to improve explicit features of these systems, more attention needs to be dedicated to mechanisms for extracting and selecting features.

Fundamental analysis calculates the genuine value of a sector/company and determines the amount that one share of that company should cost. A supposition is made that if given sufficient time, the company will move to a cost agreement with the prediction. If a sector/company is undervalued, then the market value of that company should rise, and conversely, if a company is overvalued, then the market price should fall. The analysis is performed considering various factors, such as yearly fiscal summaries and reports, balance sheets, a future prospectus, and the company's work environment. Fundamental analysis can be used for the consideration of financial ratios to distinguish poor stocks from quality stocks. Fundamental analysis lacks adequate knowledge of the rules governing the workings of the system, and there is non-linearity in the system.

Technical analysis is the study of stock prices to make a profit or to make better investment decisions. The technical analysis predicts the direction of the future price movements of stocks based on their historical data and helps to analyse financial time series data using technical indicators to forecast stock prices. Meanwhile, it is assumed that the price moves in a trend and has momentum. Technical analysis uses price charts and certain formulae, and studies patterns to predict future stock prices; it is mainly used by short-term investors. The price would be considered high, low, or open, or the closing price of the stock, where the time points would be daily, weekly, monthly, or yearly. Dow theory puts forward the main principles for technical analysis, which are that the market price discounts everything, prices move in trends, and historic trends usually repeat the same patterns. The flaws of technical analysis are that expert's opinions define rules in technical analysis, which are fixed and are reluctant to change. Various parameters that affect stock prices are ignored.

The prerequisite is to overcome the deficiencies of fundamental and technical analysis, and the evident advancement in the modelling techniques has motivated various researchers to study new methods for stock price prediction. A new form of collective intelligence has emerged, and new innovative methods are being employed for stock value forecasting. The methodologies incorporate the work of machine learning algorithms for stock market analysis and prediction.

Machine Learning Approach: Because of global digitization, SMP has entered a technological era. Machine learning in stock price prediction is used to discover patterns in data. Usually, a tremendous amount of structured and unstructured heterogeneous data is generated from stock markets. Using machine learning algorithms, it is possible to quickly analyze more complex heterogeneous data and generate more accurate results. Various machine-learning methods have been used for SMP. Machine learning approaches are mainly categorized into supervised and unsupervised approaches. In the supervised learning approach, named input data and the desired output are given to the learning algorithms. Meanwhile, in the unsupervised learning approach, unlabeled input data is provided to the learning algorithm, and the algorithm identifies the patterns and generates the output accordingly. Furthermore, different algorithmic approaches have been used in SMP, such as the Support Vector Machine (SVM), k Nearest Neighbors (kNN), Decision Trees, Fuzzy time series, and Evolutionary Algorithms.

II. RELATED WORK

Related work in stock prediction systems typically refers to previous research, models, or approaches that have been developed in the field of stock market forecasting. It's essential to review related work to understand the current state of the art and build upon existing knowledge. Here are some areas of related work.

1] **Time Series Analysis**: Stock prediction models often rely on time series analysis, with researchers devising various techniques like autoregressive integrated moving averages (ARIMA), GARCH models, and Exponential Smoothing to formulate predictions and models for stock prices.

2] **Machine Learning Models:** A considerable amount of research has been centered on employing machine learning algorithms for stock prediction. This encompasses utilizing models such as support vector machines (SVM), random forests, decision trees, neural networks, and deep learning models like LSTM and GRU for analyzing time series data.

3] **Technical Analysis:** Investigate research pertaining to technical indicators and chart patterns utilized in stock trading. The effectiveness of diverse technical analysis tools in forecasting stock price movements has been scrutinized by researchers.

4] **Fundamental Analysis:** Explore studies that leverage fundamental data, including financial reports, earnings, and economic indicators, to anticipate stock prices. This approach involves comprehending the financial well-being of a company.

5] **News Sentiment Analysis:** Another crucial aspect to examine is the impact of news sentiment and social media sentiment on stock prices. Studies have analyzed how natural language processing (NLP) techniques can be employed to measure sentiment and its influence on trading decisions.

6] **Hybrid Models:** Some researchers have proposed hybrid models that amalgamate multiple approaches, combining technical and fundamental analysis with machine learning techniques to enhance prediction accuracy.

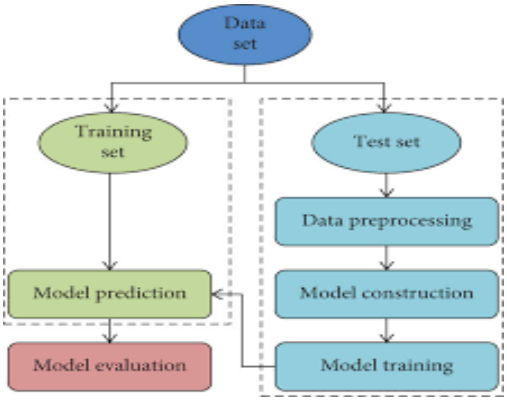
7] **Ensemble Methods:** Ensemble methods such as bagging and boosting have been utilized to merge predictions from several models, leading to improved accuracy and reliability in stock prediction.

When reviewing related work, it's essential to critically assess the strengths and weaknesses of existing models and methodologies. This assessment can help you identify gaps in the literature and areas where your research can contribute to improving stock prediction systems. By examining the related work in the field of copyright management, this paper establishes itself within the context of ongoing research and development efforts. It underscores the significance of addressing existing limitations and exploring the transformative potential of stock market technology in the protection and management of intellectual property rights.

III. DESIGN

1] Generic Scheme for SMP: The subsequent images portray the overall procedure involved in SMP. The process commences with data collection, followed by the pre-processing of said data in preparation for utilization by a machine learning model. Typically, prediction models rely on market data, which will be explored in the upcoming section. Afterward, the subsequent section classifies previous studies based on

the type of data employed, while the subsequent section offers a survey of previous studies based on the diverse data-preprocessing approaches implemented.



2] Types of Data: SMP systems can be categorized according to the specific type of data utilized as input. The majority of studies have hitherto used market data for their analytical purposes. More recently, certain studies have incorporated textual information derived from online sources. This section endeavors to classify studies based on the type of data employed for prediction objectives.

2. Market Data: Market data refers to the temporal historical numerical data associated with financial markets, typically encompassing price-related information. This data is utilized by analysts and traders to analyze historical trends and the latest stock prices, thereby offering insights into market behavior. Market data is readily available and can be downloaded directly from market websites. Several researchers have harnessed this data to predict price movements utilizing machine learning algorithms. A range of studies has concentrated on individual stock prediction considering specific corporations such as Apple or Google, in addition to groups of companies. Moreover, these studies have focused on time-specific predictions, such as intraday, daily, weekly, monthly, etc. Furthermore, a considerable portion of previous research has centered on categorical prediction, which classifies predictions into discrete categories like up, down, positive, or negative. SMP has extensively utilized technical indicators due to their comprehensive representation of trends in time series data. Various types of technical indicators, including trend, momentum, volatility, and volume indicators, have been explored in different studies. Additionally, multiple investigations have employed a combination of various technical indicators for SMP.

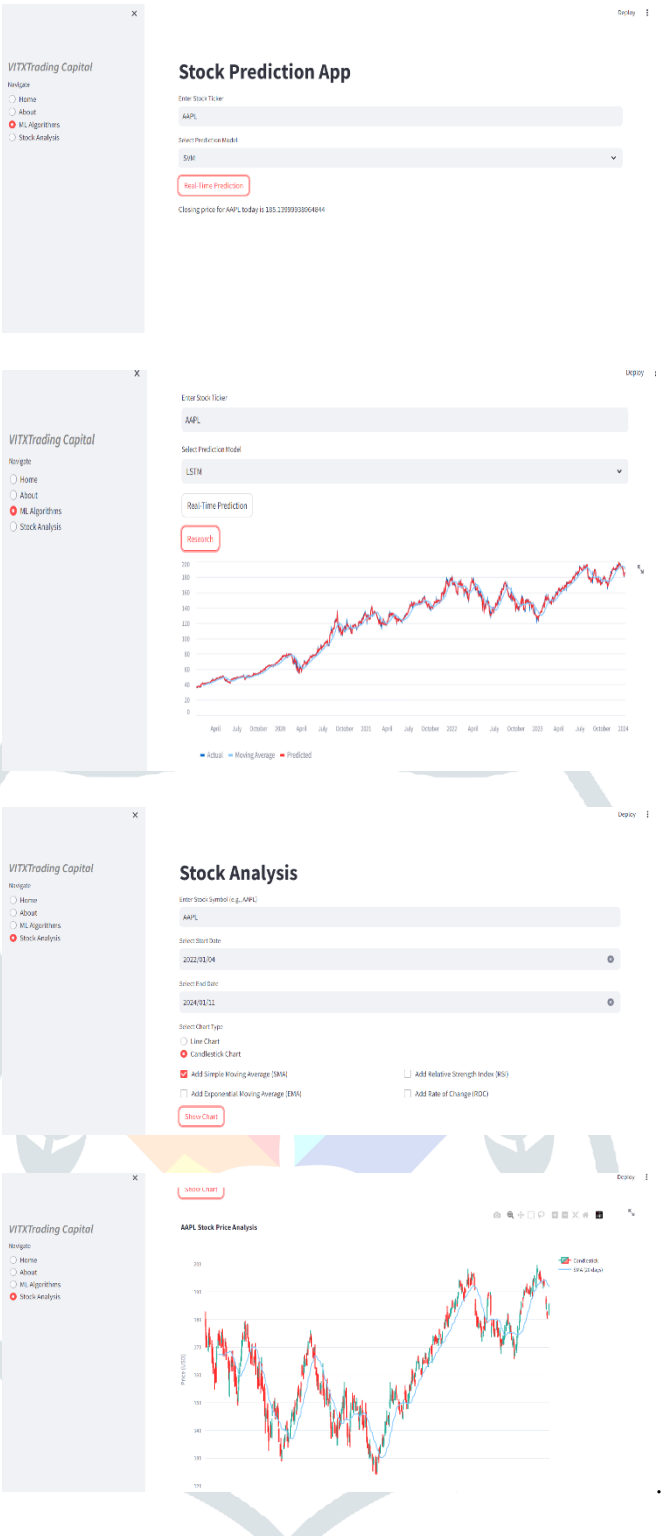
3] Data Pre-Processing: Once the data is available, it needs some pre-processing so that it can be fed to a machine-learning model. The significance of the output depends on the pre-processing of the data. The textual data must be transformed into a structured format that can be used in a machine-learning model. Previous studies revealed that there are three significant pre-processing steps, i.e., feature selection, order reduction, and the representation of features.

Order Reduction- The feature selection process for the market data leads to an increase in the number of features. High-dimensional features are extremely difficult to process and lead to the poor efficiency of most of the learning algorithms. This phenomenon is known as the curse of dimensionality. Lower numbers of features will decrease the training complexity of the algorithms. A well-known form of multi-variant analysis, Principal Component Analysis (PCA), is used to select the most relevant features, reducing the dimensionality of the features.

Feature Representation - Feature representation is one of the important factors for the efficient training of machine learning algorithms. Once the number of required features is determined, the input data is converted to a numeric representation so that machine learning algorithms can readily process it. Boolean representation is one of the most basic techniques of feature representation, in which the presence and absence of the feature (word) are represented by 1 and 0, respectively.

IV IMPLEMENTATION

Machine learning algorithms have demonstrated their tremendous efficacy in forecasting stock prices. Our website leverages the capabilities of these algorithms to deliver timely predictions by continuously analysing real-time data. By training our models on half a decade's worth of historical stock data, we empower our website to generate accurate predictions for the future. To construct robust prediction models, we utilize extensive historical stock data spanning five years. This ample dataset enables our models to effectively identify recurring patterns, trends, and correlations. To build accurate prediction models, we deploy machine learning algorithms such as support vector machines (SVM), random forests, and long short-term memory (LSTM) networks. We offer users a comprehensive range of indicators to gain valuable insights into stock performance. These indicators encompass moving averages, relative strength index (RSI), and moving average convergence divergence (MACD). By considering these indicators in conjunction with our machine learning predictions, users can make more informed investment decisions. In conclusion, our stock prediction website integrates the power of machine learning algorithms, indicators, and data visualization tools to provide users with a comprehensive platform for analysing and predicting stock market trends.



Unique Feature: At the core of my stock prediction website lies a powerful integration of machine learning and data science algorithms. This integration allows my website to provide accurate and reliable predictions by analysing the stock's past data for up to 5 years. By extracting and utilizing this historical data for training and testing, my website can make insightful predictions that can help investors make informed decisions. Machine learning and data science algorithms are the backbone of my website's prediction feature. These algorithms are designed to analyse vast amounts of historical data to identify patterns, trends, and relationships. By leveraging these algorithms, my website can provide accurate stock predictions that can help investors navigate the complex world of stock trading.

Comparison to Other Websites: Traditional stock analysis websites often rely on fundamental analysis or technical analysis to predict stock movements. While these approaches have their merits, they may not always account for the dynamic nature of the stock market. In contrast, my website's integration of machine learning and data science algorithms allows it to analyse vast amounts of data and identify patterns that may not be evident to human analysts. This results in more accurate and reliable predictions, giving investors a competitive edge.

V FUTURE SCOPE

The future scope for stock prediction systems is vast and evolving as technology, data availability, and research advances. Here are some key areas of future potential and development in stock prediction systems:

Advanced Machine Learning Techniques -As machine learning and deep learning continue to progress, it is expected that more advanced

models will emerge. Researchers will explore methods such as reinforcement learning, generative adversarial networks (GANs), and transformers t to enhance stock prediction accuracy.

Interdisciplinary approach - Collaborations among data scientists, financial experts, and domain specialists will become increasingly common. Combining machine learning with domain expertise can result in more robust and precise models.

Big Data and Real-Time Data - With the growing availability of big data and real-time data sources, stock prediction models will be capable of rapidly processing and analyzing vast amounts of information. This will lead to quicker and well-informed trading decisions.

Alternative Data Sources - The utilization of non-traditional data sources such as satellite imagery, social media sentiment, and geospatial data will expand. These alternative sources can provide unique insights into market trends and company performance.

Ethical and Regulatory Considerations - As stock prediction models gain prominence in financial markets, ethical and regulatory concerns will become more significant. Addressing issues related to bias, fairness, and market manipulation will be crucial for researchers and practitioners.

Global Markets and Cross-Asset Prediction - Stock prediction systems will increasingly need to consider global markets and the interconnections with other asset classes such as commodities, forex, and cryptocurrencies.

The future of stock prediction systems will be shaped by ongoing technological advancements, the availability of data, and the ability of researchers and practitioners to address the challenges and opportunities in the financial markets. Additionally, regulatory changes and evolving market dynamics will play a crucial role in defining the future landscape of stock prediction systems.

VI CONCLUSIONS

Financial markets offer an outstanding opportunity for traders and investors, who can engage in trading activities using any internet-connected device. Stock trading has gained increasing popularity in recent years. Technology has also impacted the stock market, transformed processes and enabled investment growth. Online trading has altered the way individuals buy and sell stocks, while advancements in the financial markets have facilitated the creation of a connected global marketplace. Such progress opens the door to new possibilities. Unlike traditional frameworks, stock market prediction (SMP) now utilizes machine learning, big data analytics, and deep learning, leading to more optimized decision-making. Presently, stock markets are vulnerable to social media sentiments and cyber-attacks. It is crucial for researchers to develop robust and secure SMP frameworks to excel in these areas.

1] Multifaceted Nature: Stock prediction encompasses multiple methodologies, including technical analysis, fundamental analysis, machine learning, and sentiment analysis. These strategies aim to forecast future stock prices and market trends.

2] Data-Driven Transformation: The availability of vast amounts of financial and alternative data combined with advanced machine learning techniques has revolutionized stock prediction. Predictive models can now process extensive datasets and extract valuable insights.

3] Collaboration across Disciplines: The development of robust and accurate stock prediction models requires collaboration between data scientists, financial experts, and domain specialists. Drawing expertise from various fields enhances the quality of predictions.

4] Market Dynamics and Challenges**:** Numerous factors, such as economic conditions, geopolitical events, investor sentiment, and company-specific news, influence stock markets. Grasping the complexity of these dynamics is essential for successful prediction.

5] Long-Term Perspective**:** Successful stock prediction necessitates a long-term outlook. Even the most advanced models cannot guarantee short-term success, and investors must be prepared to navigate market volatility.

VII REFERENCES

[1] Ali, A. 2001.Macroeconomic variables as common pervasive risk factors and the empirical content of the Arbitrage Pricing Theory. Journal of Empirical finance, 5(3): 221–240.

[2] Basu, S. 1997. The Investment Performance of Common Stocks in Relation to their Price to Earnings Ratio: A Test of the Efficient Markets Hypothesis. Journal of Finance, 33(3): 663-682.

[3] Bhatti, U. and Hanif. M. 2010. Validity of Capital Assets Pricing Model.Evidence from KSE-Pakistan.European Journal of Economics, Finance and Administrative Science, 3 (20).

[4] Krishna, V. ScienceDirect ScienceDirect NSE Stock Stock Market Market Prediction Prediction Using Using Deep-Learning Deep-Learning Models Models. Procedia Comput. Sci. 2018, 132, 1351–1362. [Google Scholar]

[5] Market Capitalization of Listed Domestic Companies (Current US$) Data. Available online: https://data.worldbank.org/indicator/CM.MKT.LCAP.CD (accessed on 19 May 2021).

[6] Market Capitalization of Listed Domestic Companies (Current US$) Data. Available online: https://data.worldbank.org/indicator/CM.MKT.LCAP.CD (accessed on 19 May 2021).

[7] Tan, T.Z.; Quek, C.; Ng, G.S. Biological Brain-Inspired Genetic Complementary Learning for Stock Market and Bank Failure Prediction. Comput. Intell. 2007, 23, 236–261. [Google Scholar] [CrossRef]

[8] Fama, E.F. Random walks in stock-market prices. Financ. Anal. J. 1995, 51, 75–80. [Google Scholar] [CrossRef][Green Version]

[9] Bujari, A.; Furini, M.; Laina, N. On using hashtags to predict companies' stock trends. In Proceedings of the 2017 14th IEEE Annual Consumer Communications & Networking Conference (CCNC), Las Vegas, NV, USA, 8–11 January 2017; pp. 25–28. [Google Scholar]

[10] Inthachot, M.; Boonjing, V.; Intakosum, S. Artificial Neural Network and Genetic Algorithm Hybrid Intelligence for Predicting Thai Stock Price Index Trend. Comput. Intell. Neurosci. 2016, 2016, 3045254. [Google Scholar] [CrossRef][Green Version]

[11] Park, C.-H.; Irwin, S.H. What Do We Know about the Profitability of Technical Analysis? J. Econ. Surv. 2007, 21, 786–826. [Google Scholar] [CrossRef]