STOCK MARKET PREDICTION

by Devanga Paul

Submission date: 12-Apr-2024 07:05PM (UTC+0530)
Submission ID: 2347563271
File name: Report_stockMarketPrediction.doc (222.97K)
Word count: 3583
Character count: 22456

A PROJECT REPORT

on

“STOCK MARKET PREDICTION”

Submitted to

KIIT Deemed to be University

In Partial Fulfillment of the Requirement for the Award of

BACHELOR’S DEGREE IN
COMPUTER SCIENCE ENGINEERING

BY

DEVANGA PAUL	21051049
SAURABH BANDYOPADHYAY	21051059
ADITYA MONDAL	21051277
ARITRA CHATTERJEE	21051290
KAUSIK KAR	21052989

UNDER THE GUIDANCE OF
DR. SAMPRITI SOOR

KALINGA INSTITUTE OF INDUSTRIAL TECHNOLOGY
BHUBANESWAR, ODISHA - 751024

KIIT Deemed to be University

School of Computer Engineering

Bhubaneswar, ODISHA 751024

CERTIFICATE

This is certify that the project entitled

"STOCK MARKET PREDICTION USING MACHINE LEARNING"

submitted by

DEVANGA PAUL	21051049
SAURABH BANDYOPADHYAY	21051059
ADITYA MONDAL	21051277
ARITRA CHATTERJEE	21051290
KAUSIK KAR	21052989

is a record of bonafide work carried out by them, in the partial fulfilment of the requirement for the award of Degree of Bachelor of Engineering (Computer Science & Engineering OR Information Technology) at KIIT Deemed to be university, Bhubaneswar. This work is done during year 2022-2023, under our guidance.

Date: / /

DR. SAMPRITI SOOR
Project Guide

Acknowledgements

We are profoundly grateful to DR. SAMPRITI SOOR of SCHOOL OF COMPUTER ENGINEERING for his expert guidance and continuous encouragement throughout to see that this project rights its target since its commencement to its completion. ....

DEVANGA
SAURABH
ADITYA
ARITRA
KAUSIK

ABSTRACT

This study delves into the effectiveness of four distinct machine learning algorithms—Random Forest, XGBoost, Support Vector Classifier (SVC), and Multilayer Perceptron (MLP)—in forecasting stock market trends. Utilizing a dataset encompassing historical stock prices, trading volumes, technical indicators, and macroeconomic factors, preprocessing techniques are applied to refine the dataset, thereby augmenting predictive accuracy.

Through rigorous training and validation via cross-validation methods, the models undergo evaluation, with performance metrics including accuracy, precision, recall, and F1-score facilitating comparative analysis.

Findings underscore the unique strengths of each algorithm. Random Forest and XGBoost, leveraging ensemble learning methodologies, demonstrate robust predictive capabilities, adept at capturing intricate data relationships. SVC exhibits proficiency in volatile market conditions, leveraging its capacity to segregate classes in high-dimensional feature spaces. Conversely, MLP showcases adeptness in discerning complex patterns, particularly in scenarios characterized by non-linear relationships.

This research furnishes valuable insights into the suitability of diverse machine learning algorithms for stock market prediction, furnishing investors and financial analysts with informed decision-making tools to navigate dynamic market landscapes.

Keywords: Random Forest, XGBoost, Support Vector Classifier (SVC), Multilayer Perceptron (MLP)

Contents

Catalog

Chapter 1

Introduction

The domain of stock market prediction remains a focal point for researchers and practitioners alike, driven by the perpetual quest for reliable methodologies to anticipate market trends and fluctuations. Within this landscape, machine learning algorithms have emerged as promising tools, offering avenues for enhanced predictive capabilities. Notably, Random Forest, XGBoost, Support Vector Classifier (SVC), and Multilayer Perceptron (MLP) have garnered attention for their distinct methodologies and strengths in stock market prediction.

13

Random Forest, known for its ensemble learning approach, combines multiple decision trees to mitigate overfitting and improve prediction accuracy. Its adaptability and robustness make it well-suited for handling large datasets and capturing complex interdependencies among financial variables.

Similarly, XGBoost, an extension of gradient boosting techniques, has gained prominence for its iterative refinement of predictions by minimizing loss functions. Its scalability and efficiency make it a powerful tool for integrating diverse features and time-series data, thereby enhancing predictive accuracy in stock market forecasting.

Support Vector Classifier (SVC) operates by identifying the hyperplane that best separates different classes, making it part22ularly effective for discerning market trends and predicting price movements. Its ability to handle both linear and non-linear relationships within financial data renders it invaluable across diverse market conditions.

Conversely, Multilayer Perceptron (MLP) represents a cornerstone of deep learning methodologies, capable of capturing intricate patterns and non-linear relationships inherent in financial datasets. Through multiple layers of interconnected neurons, MLP excels in extracting nuanced features from historical market data, offering potential for enhanced prediction accuracy in stock market analysis.

23

This study endeavors to explore and compare the efficacy of these machine learning algorithms—Random Forest, XGBoost, SVC, and MLP—in the context of stock market prediction. By examining their methodologies, strengths, and limitations, this research aims to provide insights that can inform decision-making processes for investors and financial analysts navigating the complexities of financial markets.

Chapter 2

Basic Concepts

Random Forest: Random Forest stands as a pivotal ensemble learning technique in stock market prediction, leveraging multiple decision trees to navigate intricate relationships within financial datasets. Noteworthy attributes of Random Forest include its adeptness at managing extensive data volumes, conducting feature importance analyses to discern influential variables, and its resilience against overfitting. Researchers often delve into strategies to optimize parameters such as the number of trees and maximum depth, aiming to refine predictive accuracy. Furthermore, the parallelization capabilities inherent to Random Forest facilitate efficient processing of vast financial datasets, contributing significantly to timely and precise predictions within stock market analysis.

XGBoost: Renowned as eXtreme Gradient Boosting, XGBoost has emerged as a potent tool in machine learning for stock market prediction. Rooted in gradient boosting principles, XGBoost iteratively hones predictive prowess by minimizing loss functions. In stock market prediction scenarios, XGBoost offers advantages including scalability, computational efficiency, and versatility in handling both numeric and categorical features. Researchers harness XGBoost's flexibility to integrate diverse financial indicators and time-series data, thereby enriching predictive accuracy. Additionally, XGBoost's interpretability features empower analysts to glean insights into the underlying drivers of market trends, facilitating informed decision-making processes.

Support Vector Classifier (SVC): Support Vector Classifier, a prevalent binary classification algorithm, finds utility in stock market prediction by elucidating trends and patterns within financial data. SVC operates by discerning the hyperplane that best separates different classes, thus proving adept at identifying market trends and forecasting future price movements. Researchers explore an array of kernel functions and regularization techniques to optimize SVC performance in stock market prediction endeavors. Moreover, SVC's adaptability enables it to address both linear and non-linear relationships within financial data, broadening its applicability across diverse market conditions.

Multilayer Perceptron (MLP): Serving as a foundational element in deep learning architectures, Multilayer Perceptron holds promise in stock market prediction owing to its capacity to capture intricate patterns and non-linear relationships inherent in financial datasets. Comprising multiple layers of interconnected neurons, MLP excels in extracting nuanced features from historical market data. Researchers delve into architectural nuances, activation functions, and optimization algorithms to refine MLP's performance in stock market forecasting tasks. Furthermore, advancements in deep learning methodologies, including recurrent and convolutional neural networks, further augment MLP's capabilities in capturing temporal and spatial dependencies within financial time series data, thereby paving the way for more precise and reliable predictions within stock market analysis.

Chapter 3

Related Works:

Stock market prediction presents a challenge to the Efficient Market Hypothesis (EMH) proposed by Fama and Malkiel (1970), which asserts that stock prices reflect all available information. However, this theory is contentious, as noted by Malkiel (2003), prompting researchers to explore prediction algorithms capable of capturing the complex dynamics of financial systems.

A variety of algorithms, including Support Vector Machines (SVM), Neural Networks, Linear Discriminant Analysis, Linear Regression, K-Nearest Neighbors (KNN), and Naive Bayesian Classifier, have been applied in stock prediction. SVM has notably received significant attention, as evidenced by Li, Li, and Yang (2014), who investigated sensitivity to external conditions in stock prices.

Dai and Zhang (2013) employed Logistic Regression, Quadratic Discriminant Analysis, and SVM to predict stock prices, with SVM demonstrating superior accuracy in long-term predictions. Xinjie (2014) utilized an ensemble method with technical indicators and SVM, while Devi, Bhaskaran, and Kumar (2015) proposed a hybrid model combining Cuckoo Search with SVM. Additionally, Giacomel, Galante, and Pareira (2015) introduced a neural network ensemble for predicting stock market movements.

Boonpeng and Jeatrakul (2016) applied One vs All and One vs One neural networks to historical data from the Stock Exchange of Thailand, achieving notable accuracy.

Despite these advancements, the potential of ensemble learning algorithms like Random Forest in stock market prediction remains underexplored. Random Forest, comprising multiple decision trees, holds promise for improving predictive accuracy.

This paper aims to utilize Random Forest to develop a predictive model for stock market trends. Subsequent sections will cover data collection, preprocessing, feature extraction, model training, and comparative analysis, showcasing the efficacy of the proposed algorithm.

Chapter 4

Implementation

In this section, present the implementation done by us during the project development.

4.1 Methodology OR Proposal

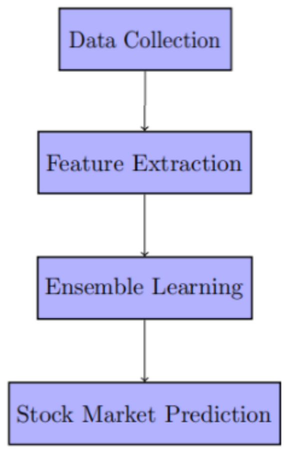

Fig 1: Proposed Methodology

The learning algorithms used in our paper is random forest,SVc,Xboost, Multi-Layer Perceptron. The time series data is acquired,smoothed and technical indicators are extracted. Technical indicators are parameters which provide insights to the expected stock price behavior in future. These technical indicators are then used to train the random forest. The details of each step will be discussed in this section.

4.2 Data Collection

The process of data collection for stock market prediction using machine learning involves sourcing datasets from reliable platforms like Yahoo Finance, encompassing a diverse selection of stocks such as Tesla, Apple, Microsoft, and Google. These stocks represent key players in the financial markets, offering valuable insights into market trends and behavior.

From Yahoo Finance, historical data for each stock is collected, spanning a significant timeframe to capture various market conditions and fluctuations. The dataset typically includes daily stock metrics like open, high, low, and close prices, along with trading volume. Additionally, fundamental data such as earnings reports, dividends, and corporate events may also be incorporated to enrich the dataset.

Furthermore, feature extraction techniques like Volume Weighted Average Price (VWAP) and signal correction are applied to enhance the predictive power of machine learning models. VWAP provides a measure of average price weighted by trading volume, offering valuable information about price trends and market sentiment. Signal correction techniques refine the predictive model to account for biases or inaccuracies in the data, ensuring more accurate predictions.

By leveraging datasets derived from Yahoo Finance and incorporating feature extraction methods like VWAP and signal correction, machine learning algorithms can effectively analyze historical stock data and predict future price movements. This enables investors and financial analysts to make well-informed decisions in navigating the complexities of the stock market.

4.3 Feature extraction

Volume Weighted Average Price (VWAP) is a widely recognized metric in financial markets, calculated by dividing the total value of trades by the total volume traded over a specified time period. VWAP serves as a crucial benchmark for institutional traders to evaluate trade execution relative to the market average.

In the realm of stock market prediction using machine learning, VWAP emerges as a significant feature for predictive models. By integrating VWAP data with other pertinent features such as historical prices, trading volumes, technical indicators, and market sentiment, machine learning algorithms can discern patterns and correlations to forecast future stock prices or market trends effectively.

VWAP_5: This represents the Volume Weighted Average Price calculated over the last 5 periods. It gives traders an average price for the security over the past 5 trading sessions, weighted by the volume traded during each session.

VWAP_10: Similarly, VWAP_10 represents the Volume Weighted Average Price calculated over the last 10 periods. It provides an average price for the security over the past 10 trading sessions, again weighted by the volume traded during each session.

VWAP_20: VWAP_20 represents the Volume Weighted Average Price calculated over the last 20 periods. It gives traders an average price for the security over the past 20 trading sessions, weighted by the volume traded during each session.

signal correction augments the utility of VWAP-based predictions. Signal correction entails refining the predictive model to address any biases or inaccuracies in the input data, thereby ensuring more precise predictions. This refinement process may involve outlier detection, data normalization, or adjusting for market anomalies.

In essence, the utilization of VWAP data and the implementation of signal correction techniques within machine learning models contribute to the development of robust and accurate predictions in stock market forecasting. Such advancements aid traders and investors in making well-informed decisions in dynamic financial markets.

UPSAMPLING : Class imbalance occurs when one class (e.g., "up" movements) is significantly more prevalent than the other class (e.g., "down" movements) in the dataset. This can lead to biased models that favor the majority class and perform poorly in predicting the minority class.

To mitigate this issue, upsampling involves increasing the number of instances in the minority class by randomly duplicating existing instances or generating synthetic data points. By balancing the class distribution, upsampling helps machine learning models learn more effectively from both classes, leading to improved prediction performance.

In the context of stock market prediction, upsampling ensures that the model is trained on a balanced representation of both positive and negative price movements, allowing it to make more accurate predictions across different market conditions.

4.4 **Prediction Models**:

Random Forest : It's implementation in stock market prediction can incorporate features like Volume Weighted Average Price (VWAP) and signal correction to enhance predictive accuracy. Firstly, historical stock data containing VWAP alongside other relevant metrics is collected and preprocessed. Signal correction techniques are applied to refine the data and mitigate biases.

Subsequently, the Random Forest model is trained using this augmented dataset, where VWAP serves as a crucial feature alongside traditional indicators. During training, each decision tree within the Random Forest ensemble learns to utilize VWAP and other features to make predictions.

During prediction, the ensemble of decision trees aggregates individual predictions, leveraging the collective intelligence of the forest. By incorporating VWAP and signal correction, Random Forest can better capture the nuances of stock market behavior, leading to more accurate predictions of stock prices and market trends. This comprehensive approach aids investors and analysts in making informed decisions in dynamic financial markets.

Support Vector Classifier (SVC) implementation in stock market prediction involves integrating features like Volume Weighted Average Price (VWAP) and signal correction to enhance predictive accuracy. Initially, historical stock market data, including VWAP and other relevant metrics, is collected and preprocessed. Signal correction techniques are then applied to refine the data and mitigate biases.

Subsequently, the SVC model is trained using this augmented dataset, where VWAP serves as a crucial feature alongside traditional indicators. During training, SVC identifies the hyperplane that best separates different market trends and price movements, leveraging VWAP and other features.

During prediction, SVC utilizes the learned hyperplane to classify future market conditions, considering VWAP and other features to make accurate predictions. By incorporating VWAP and signal correction, SVC can better discern trends and patterns within financial data, leading to more reliable predictions of stock prices and market trends.

XGBoost implementation in stock market prediction involves incorporating features like Volume Weighted Average Price (VWAP) and signal correction to enhance predictive accuracy. Initially, historical stock data containing VWAP alongside other relevant metrics is collected and preprocessed. Signal correction techniques are then applied to refine the data and mitigate biases.

Subsequently, the XGBoost model is trained using this augmented dataset, where VWAP serves as a crucial feature alongside traditional indicators. During training, XGBoost iteratively improves predictions by minimizing loss functions, leveraging VWAP and other features to capture complex market dynamics.

During prediction, XGBoost utilizes the learned patterns to forecast future stock prices and market trends, considering VWAP and other features to make accurate predictions. By incorporating VWAP and signal correction, XGBoost can better capture the nuances of stock market behavior, leading to more reliable predictions and assisting investors and analysts in making informed decisions.

Multilayer Perceptron (MLP) implementation in stock market prediction involves incorporating features like Volume Weighted Average Price (VWAP) and signal correction to enhance predictive accuracy. Initially, historical stock data containing VWAP alongside other relevant metrics is collected and preprocessed. Signal correction techniques are then applied to refine the data and mitigate biases.

Subsequently, the MLP model is trained using this augmented dataset, where VWAP serves as a crucial feature alongside traditional indicators. During training, MLP learns complex patterns and relationships within the data through multiple layers of interconnected neurons, leveraging VWAP and other features to make predictions.

During prediction, MLP utilizes the learned patterns to forecast future stock prices and market trends, considering VWAP and other features to make accurate predictions. By incorporating VWAP and signal correction, MLP can better capture the nuances of stock market behavior, leading to more reliable predictions and assisting investors and analysts in making informed decisions.

Chapter 5

Results and Outputs

		precision	recall	f1-score	support
RandomForest	NO	0.91	0.93	0.92	103
	YES	0.3	0.23	0.26	13
	accuracy			0.85	116
	macro avg	0.6	0.58	0.59	116
	weighted avg	0.84	0.85	0.84	116

		precision	recall	fl-score	support
XGBoost	NO	0.92	0.89	0.91	103
	YES	0.31	0.38	0.34	13
	accuracy			0.84	116
	macro avg	0.62	0.64	0.63	116
	weighted avg	0.85	0.84	0.84	116

		precision	recall	fl-score	support
SVC	NO	0.96	0.43	0.59	103
	YES	0.16	0.85	0.27	13
	accuracy			0.47	116
	macro avg	0.56	0.64	0.43	116
	weighted avg	0.87	0.47	0.55	116

Multi Layer Perceptron		precision	recall	fl-score	support

NO	0.92	0.59	0.72	103
YES	0.16	0.62	0.25	13
accuracy			0.59	116
macro avg	0.54	0.68	0.49	116
weighted avg	0.84	0.59	0.67	116

Chapter 6

Conclusion and Future Scope

6.1 **Conclusion :**

Stock market prediction using machine learning (ML) techniques has made significant strides in recent years, offering investors and traders valuable insights into market trends and opportunities. Through the application of various ML algorithms, such as regression, classification, and deep learning, researchers and practitioners have been able to develop models capable of analyzing vast amounts of financial data and making predictions with a certain degree of accuracy.

One key conclusion drawn from the advancements in stock market prediction is that ML models can provide valuable decision support tools for investors. These models can analyze historical price data, technical indicators, market sentiment, and other relevant factors to generate forecasts about future price movements. While no model can predict the market with absolute certainty, ML-based approaches have demonstrated the potential to identify patterns and trends that human analysts might overlook.

Moreover, the integration of alternative data sources, such as social media sentiment, news articles, and satellite imagery, into stock market prediction models has expanded the scope of analysis and enhanced prediction accuracy. By incorporating a wide range of data inputs, ML models can capture nuanced market dynamics and adapt to changing market conditions more effectively.

6.2 Future Scope

The future scope of stock market prediction using machine learning involves advanced algorithms leveraging big data analytics, sentiment analysis, and deep learning models to enhance accuracy and reliability. Integration of real-time market data and evolving ML techniques promises to revolutionize forecasting, enabling better decision-making for investors and financial institutions.

References

[1]S. M. Metev and V. P. Veiko, Laser Assisted Microtechnology, 2nd ed., R. M. Osgood, Jr., Ed. Berlin, Germany: Springer-Verlag, 1998.

[2] Breckling, Ed., The Analysis of Directional Time Series: Applications to Wind Speed and Direction, ser. Lecture Notes in Statistics. Berlin, Germany: Springer, 1989, vol. 61.

[3] S. Zhang, C. Zhu, J. K. O. Sin, and P. K. T. Mok, “A novel ultrathin elevated channel low-temperature poly-----

[4] M. Wegmuller, J. P. von der Weid, P. Oberson, and N. Gisin, “High resolution fiber distributed measurements with coherent OFDR,” in Proc. ECOC’00, 2000, paper 11.3.4, p. 109.

[5] R. E. Sorace, V. S. Reinhardt, and S. A. Vaughn, “High-speed digital-to-RF converter,” U.S. Patent 5 668 842, Sept. 16, 1997.

[6] (2002) The IEEE website. [Online]. Available: http://www.ieee.org/

[7] M. Shell. (2002) IEEEtran homepage on CTAN. [Online]. Available: http://www.ctan.org/tex-archive/macros/latex/contrib/supported/IEEEtran/

SAMPLE INDIVIDUAL CONTRIBUTION REPORT:

STOCK MARKET PREDICTION

Devanga Paul	21051049
Saurabh Bandyopadhyay	21051059
Aditya Mondal	21051277
Aritra Chatterjee	21051290
Kausik Kar	21052989

Devanga Paul (21051049): Responsible for data collection and studying research papers relevant to the project. His contributions likely involved gathering the necessary datasets, ensuring data quality, and conducting literature reviews to understand existing methodologies and techniques in stock market prediction using machine learning.

Saurabh Bandyopadhyay (21051059): Took charge of structuring the report and also studied research papers pertinent to the project. His role likely involved organizing the report into coherent sections, defining the chapters, and outlining the content flow to ensure clarity and coherence throughout the document.

Aditya Mondal (21051277): Focused on feature extraction, a critical aspect of machine learning model development. He likely worked on identifying relevant features from the dataset that could serve as predictors for stock price movements. Feature extraction involves selecting, transforming, and engineering features to improve model performance.

Aritra Chatterjee (21051290): Aritra was involved in the prediction work using two specific models: XGBoost and Multi-layer Perceptron (MLP). His contributions likely included model implementation, parameter tuning, model evaluation, and interpreting results obtained from these models in the context of stock market prediction.

Kausik Kar (21052989): Kausik was responsible for the prediction work using two other models: Random Forest and Support Vector Classifier (SVC). His role likely encompassed similar tasks as Aritra's, including model implementation, experimentation, performance evaluation, and comparing the results with other models used in the project.

Full Signature of Supervisor: Full signature of the student:

………………………………. ……………………………….

TURNITIN PLAGIARISM REPORT

(This report is mandatory for all the projects and plagiarism must be below 25%)

STOCK MARKET PREDICTION

ORIGINALITY REPORT

20%	16%	4%	14%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Banaras Hindu University Student Paper	6%
2	financedocbox.com Internet Source	3%
3	jcdv.org Internet Source	2%
4	www.coursehero.com Internet Source	1%
5	repository.sustech.edu Internet Source	1%
6	Submitted to Liverpool John Moores University Student Paper	1%
7	Submitted to International Islamic University Chittagong Student Paper	1%
8	Submitted to Sim University Student Paper	1%
9	Submitted to University of Bristol	

Student Paper

<1%

10 Submitted to Aston University
Student Paper

<1%

11 Submitted to University of Glasgow
Student Paper

<1%

12 bigquant.com
Internet Source

<1%

13 Osman İsa Çelik, Gürcan Büyüksalih, Cem Gazioğlu. "Improving the Accuracy of Satellite-Derived Bathymetry Using Multi-Layer Perceptron and Random Forest Regression Methods: A Case Study of Tavşan Island", Journal of Marine Science and Engineering, 2023
Publication

<1%

14 Submitted to Engineers Australia
Student Paper

<1%

15 Submitted to Universite Saint Joseph
Student Paper

<1%

16 www.smallstock.info
Internet Source

<1%

17 mdpi-res.com
Internet Source

<1%

18 5dok.org
Internet Source

<1 %

19	discovery.researcher.life Internet Source	<1 %
20	tudr.thapar.edu:8080 Internet Source	<1 %
21	www.mdpi.com Internet Source	<1 %
22	assets.researchsquare.com Internet Source	<1 %
23	www.ijraset.com Internet Source	<1 %
24	Suryoday Basak, Saibal Kar, Snehanshu Saha, Luckyson Khaidem, Sudeepa Roy Dey. "Predicting the direction of stock market prices using tree-based classifiers", The North American Journal of Economics and Finance, 2018 Publication	<1 %

Exclude quotes	Off	Exclude matches	Off
Exclude bibliography	Off		