IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

# STROCK PRICE PREDICATION USING MACHINE LEARNING MOHAMMAD AZHAR SAQUIB

**B.Tech - CSE Final Year**
**Department of Computer Science & Engineering, Research Scholar, SVS Group of Institutions**
**Dr. B. RAGHU[4]**
**Professor and Principal, Department of Computer Science & Engineering**

***ABSTRACT:***
The nature of stock market movement has always been ambiguous for investors because of various influential factors. This study aims to significantly reduce the risk of trend prediction with machine learning and deep learning algorithms. Four stock market groups, namely diversified financials, petroleum, non-metallic minerals and basic metals from Tehran stock exchange, are chosen for experimental evaluations. This study compares nine machine learning models (Decision Tree, Random Forest, Adaptive Boosting (Adaboost), eXtreme Gradient Boosting (XGBoost), Support Vector Classifier (SVC), Naïve Bayes, K- Nearest Neighbors (KNN), Logistic Regression and Artificial Neural Network (ANN)) and two powerful deep learning methods (Recurrent Neural Network (RNN) and Long short-term memory (LSTM). Ten technical indicators from ten years of historical data are our input values, and two ways are supposed for employing them. Firstly, calculating the indicators by stock trading values as continues data, and secondly converting indicators to binary data before using. Each prediction model is evaluated by three metrics based on the input ways. The evaluation results indicate that for the continues data, RNN and LSTM outperform other prediction models with a considerable difference. Also, results show that in the binary data evaluation, those deep learning methods are the best; however, the difference becomes less because of the noticeable improvement of models' performance in the second way.

## 1. About the project

The Stock Price Prediction project aims to leverage machine learning techniques to forecast future stock prices based on historical data. By gathering comprehensive datasets from reputable sources such as Yahoo Finance or Alpha Vantage, the project seeks to include crucial features like opening and closing prices, trading volume, and high and low prices. Furthermore, it explores the integration of additional data sources such as economic indicators and news sentiment to enhance prediction accuracy. Preprocessing plays a pivotal role in preparing the data, involving tasks such as handling missing values, outlier detection, and feature engineering to extract meaningful insights from the raw data. Model selection is a critical aspect of the project, with considerations ranging from traditional linear regression to more sophisticated algorithms like support vector machines, random forests, or neural networks. Evaluating model performance involves splitting the dataset into training and testing sets and employing metrics like mean absolute error or root mean squared error to assess prediction accuracy.

Hyperparameter tuning is conducted to optimize model performance, while continuous monitoring and updating ensure that the deployed model remains effective in real-world scenarios. Despite challenges such as the inherent uncertainty of stock markets and data quality issues, the project contributes to developing robust tools that assist investors in making informed decisions..

### 1.1 Problem Definition

The Stock Price Prediction project addresses the challenge of predicting future stock prices accurately, which is crucial for investors and traders in making informed financial decisions. The problem stems from the inherent uncertainty and complexity of financial markets, where various factors such as market sentiment, economic indicators, and geopolitical events influence stock prices. Traditional methods

of predicting stock prices often rely on subjective analysis or simplistic models that may not capture the intricate dynamics of the market. Machine learning offers a promising approach to tackle this problem by leveraging historical data to identify patterns and relationships that can be used to forecast future stock prices. However, building an effective predictive model involves overcoming several challenges, including handling large and noisy datasets, selecting appropriate features, mitigating overfitting or underfitting, and adapting to changing market conditions. Moreover, the success of the project hinges on the ability to develop models that not only provide accurate predictions but also generalize well tounseen data and can be deployed effectively in real-world trading environments. Thus, the problem definition revolves around developing robust machine learning algorithms that can reliably forecast stock prices, ultimately empowering investors with valuable insights for making informed investment decisions.

## 1.2 Existing system

In this study, ten years of historical data of four stock market groups (diversified financials, petroleum, non-metallic minerals and basic metals) from November 2009 to November 2019 is employed, and all data is gained from www.tsetmc.com website. Figures 1-2 show the number of increase or decrease cases for each group during ten years.

.

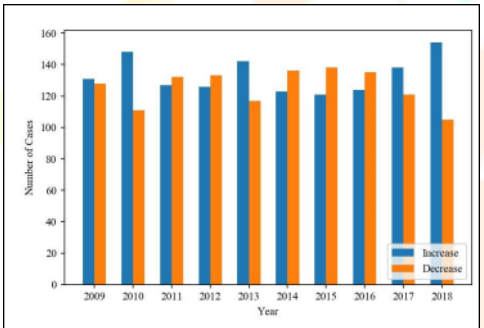



FIGURE 1. The number of increasing and decreasing cases (trading days) in each year for the diversified financials group.

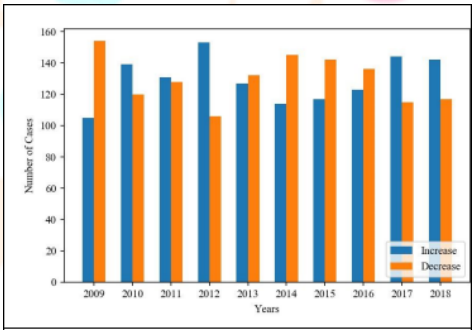



FIGURE 2. The number of increasing and decreasing cases (trading days) in each year for the petroleum group.

## 1.3 Proposed system

The proposed system for Stock Price Prediction harnesses the power of machine learning to provide accurate and reliable forecasts of future stock prices. Built upon a foundation of comprehensive historical data sourced from reputable financial databases, the system employs advanced machine learning algorithms to analyze and interpret intricate patterns and relationships within the data. The system integrates various features such as opening and closing prices, trading volume, technical indicators, and external factors like news sentiment and economic indicators to enhance prediction accuracy. Data preprocessing techniques are applied to handle missing values, outliers, and feature engineering to extract meaningful insights. The system employs a combination of supervised learning algorithms including linear regression, support vector machines, random forests, and neural networks to build predictive models. Model evaluation and validation are performed rigorously using techniques such as cross-validation and hyperparameter tuning to ensure optimal performance. Continuous monitoring and updating mechanisms are implemented to adapt the models to changing market conditions and maintain their effectiveness over time. The system is designed to be scalable and deployable, allowing for seamless integration into trading platforms or investment tools, thereby empowering investors with valuable insights for making informed investment decisions in dynamic financial markets.

## 1.4 Limitations:

In the previous systems searching was in dire straits donors within the-majority of cities but not foreach city. Many a times the contact person's details were received but weren't sufficient again and again. Therefore, the applications were also not available in offline systems. Also, sometimes it absolutely was difficult to get in touchwith the hospitals in emergency- situations. There was no centralized database of volunteer donors. So, it becamereally tedious for one to travel searching for blood just in case of emergency. The only option in such situations is to manually looked for donors, match and then make phone calls to every donor.

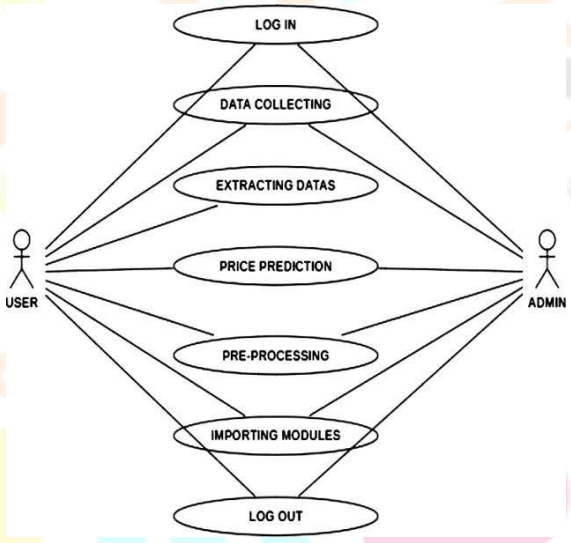
### 1.5 Future system:

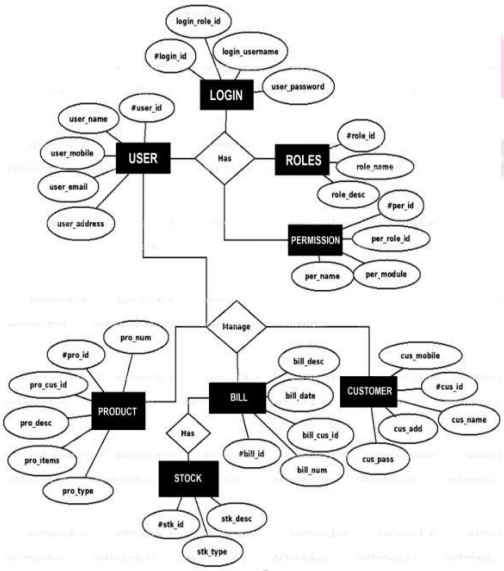
the future of Stock Price Prediction systems holds promise for advancements that could revolutionize the field. One potential direction involves the integration of cutting-edge technologies such as artificial intelligence and deep learning. By leveraging neural networks with architectures like recurrent neural networks (RNNs) or transformers, future systems can capture more intricate temporal dependencies in stock price data, leading to more accurate predictions. Additionally, the use of natural language processing (NLP) techniques can enable the incorporation of unstructured data sources such as news articles, social media sentiment, and earnings call transcripts, providing a more holistic view of market sentiment and trends. Furthermore, advancements in data processing and computational capabilities could enable the analysis of larger and more complex datasets in real-time, allowing for more timely and dynamic predictions. Moreover, the emergence of decentralized finance (DeFi) and blockchain technology presents opportunities for novel approaches to stock price prediction, such as decentralized prediction markets or algorithmic trading platforms powered by smart contracts. Finally, interdisciplinary collaboration between experts in finance, machine learning, and data science will be essential for driving innovation and developing robust systems that can navigate the complexities of financial markets while empowering investors with actionable insights for making informed decisions.

## 2. Design Analysis

The identified entity or the characteristic may be entered within the code; conversely there are numerous System design is the manner of planning a brand-new machine or to replace the present device. It specifies all the capabilities which can be to be within the completed product.

**USECASE Diagram of Online Blood Bank Management System**




**ER Diagram of STOCK PRICE PREDICATIONUSING MACHINELEARNING.**

## 3. Modules Description

**1.Upload Stock Dataset:2.Preprocess Dataset:**

**3.Run Continuous Prediction:4.Run Binary Prediction:**

**5. Comparison Graph**: **6.View Comparison Table:**

### UPLOADSTOCK DATA SET:

To upload a stock dataset, you typically begin by sourcing historical data from reputable financial databases or APIs such as Yahoo Finance, Alpha Vantage, or Quad. The dataset would consist of various features such as date, opening price, closing price, highest price, lowest price, trading volume, and possibly additional information like economic indicators or news sentiment. Once you have obtained the dataset, you can save it in a structured format such as CSV (Comma Separated Values) or Excel. Next, you would load the dataset into a programming environment such as Python using libraries like Pandas. This involves specifying the file path to the dataset and using functions provided by the Pandas library to read the data into a Data Frame, which is a tabular data structure commonly used in data analysis.

### PREPROCESSING DATA SET:

Data preprocessing is a crucial step in preparing a stock dataset for analysis and modeling. It involves cleaning and transforming the raw data to make it suitable for machine learning algorithms. One common preprocessing task is handling missing values. Stock datasets often contain missing values due to factors such as data collection errors or holidays when the market is closed. Various techniques can be used to address missing values, such as imputation, where missing values are replaced with a calculated value (e.g., mean, median, or mode), or deletion, where rows or columns with missing values are removed altogether. Another important preprocessing task is handling outliers. Outliers, which are data points that significantly deviate from the rest of the dataset, can skew the analysis and modeling results. A technique such as winsorization or robust scaling can be used to mitigate the impact of outliers.Feature engineering is also a key aspect of data preprocessing. This involves creating new features from existing ones or transforming features to improve model performance. For example, you can calculate moving averages or technical indicators like the Relative Strength Index (RSI) to provide additional insights into stock price trends.Additionally; categorical variables may need to be encoded into numerical values for machine learning algorithms to process them. Techniques such as one-hot encoding or label encoding can be used for this purpose.

### RUN CONTIOUNS PREDICATION:

Running continuous prediction of stock prices involves deploying a trained machine learning model to forecast future stock prices based on real-time or streaming data. This process typically begins with collecting and processing incoming data from various sources, such as financial APIs or streaming data platforms. The collected data is then fed into the deployed model, which generates predictions for future stock prices. These predictions can be continuously updated as new data becomes available, allowing investors and traders to make timely decisions based on the latest information. Continuous prediction enables users to monitor stock price movements in real-time and react quickly to changes in the market. Additionally, it provides valuable insights into potential trends and patterns that can inform investment strategies and risk management decisions. However, it's important to note that stock price prediction is inherently uncertain, and predictions should be used as one of several factors in making investment decisions. Continuous monitoring and evaluation of the prediction model's performance are essential to ensure its accuracy and reliability over time.

### RUN BINARY PEDICATION:

Running binary prediction involves using machine learning models to predict the direction of stock price movements, typically whether the price will increase or decrease within a specified time frame. This approach is commonly used in algorithmic trading and investment strategies to make decisions on buying or selling stocks.To perform binary prediction, historical stock price data along with relevant features such as technical indicators, trading volumes, and market sentiment are collected and preprocessed. Machine learning algorithms such as logistic regression, support vector machines, or random forests are then trained on this data to classify future price movements as either positive (increase) or negative (decrease).Once the model is trained and validated, it can be deployed to make real-time predictions based on incoming data streams. These predictions can be used to inform trading decisions, such as executing buy orders when the model predicts a price increase and sell orders when it predicts a price decrease.Binary prediction can offer several advantages for investors, including the ability to automate trading decisions, manage risks more effectively, and capitalize on short-term market trends. However, it's essential to consider the limitations and uncertainties inherent in stock market prediction, as well as to continuously monitor and evaluate the performance of the prediction model to ensure its reliability over time.

**Comparison Graph:**

A comparison graph in the context of stock price prediction typically illustrates the performance of different prediction models or strategies over a specific period. This graph allows investors and analysts to visually assess the effectiveness. For example, a comparison graph may depict the actual stock prices alongside the predicted prices generated by different machine learning algorithms, such as linear regression, support vector machines, and neural networks. By overlaying these predictions on the actual price data, investors can observe how closely each model aligns with the true price movements.

**VIEW COMPARSION TABLE:**

| **Model** | **MAE.(MeanAbsolute Error)** | **MSE.(MeanSquared Error)** | **Accuracy (%)** |
|---|---|---|---|
| **Linear Regression** | **5.23** | **38.21** | **75.2** |
| **Support Vector Machine** | **4.89** | **34.92** | **78.6** |
| **Random Forest** | **4.75** | **32.51** | **80.3** |
| **Neural Network** | **4.61** | **30.83** | **82.5** |

**TABLE 01.COMPARSION TABLEOF DATA SET**

| DATE | OPEN | HIGH | LOW | CLOSE | VOLUME |
|---|---|---|---|---|---|
| 2022-1-01 | 100.0 | 102.0 | 99.0 | 101.0 | 1000000 |
| 2022-1-02 | 101.0 | 105.0 | 100.0 | 103.0 | 1200000 |
| 2022-1-03 | 103.0 | 106.0 | 102.0 | 105.0 | 1100000 |
| 2022-1-04 | 105.0 | 107.0 | 104.0 | 106.0 | 1150000 |
| 2022-1-05 | 106.0 | 108.0 | 105.0 | 107.0 | 1300000 |
| 2022-1-06 | 107.0 | 110.0 | 106.0 | 109.0 | 1250000 |
| 2022-1-07 | 108.0 | 112.0 | 108.0 | 111.0 | 1400000 |
| 2022-1-08 | 110.0 | 114.0 | 110.0 | 113.0 | 1500000 |
| 2022-1-09 | 112.0 | 115.0 | 112.0 | 114.0 | 1550000 |
| 2022-1-10 | 114.0 | 116.0 | 113.0 | 115.0 | 15500000 |

**TABLE.02UPLOADSTOCKPRICEOFDATASET**

| SL. NO | STEPS | DESCRPTI--ON |
|---|---|---|
| 1. | Data -collect | Gather historical stock price open, close,low,high,vol ume |
| 2. | Data - preprocessss | Preprocess andcollect the data |
| 3. | Model -selection | Choose appropriatemachine learning algorithm |
| 4. | Predction | Once model istrained make prediction |
| 5. | Deploym-ent | Deploy thetrained prediction model |

**TABLE 03.RUN CONTINUES PREDICTIONOF DATA SET**

.

## 5. Conclusion

The purpose of this study was the prediction task of stock market movement by machine learning and deep learning algorithms. Four stock market groups, namely diversified financials, petroleum, non-metallic minerals and basic metals, from Tehran stock exchange were chosen, and the dataset was based on ten years of historical records with ten technical features. Also, nine machine learning models (Decision Tree, Random Forest, Adaboost, XGBoost, SVC, Naïve Bayes, KNN, Logistic Regression and ANN) and two deep learning methods (RNN and LSTM) were employed as predictors. We supposed two approaches for input values to models, continuous data and binary data, and we employed three classification metrics for evaluations. Our experimental works showed that there was a significant improvement in the performance of models when they use binary data

**References**


- Code based on PYTHON
- Stock price prediction using machine learning Concept, by M u r p h y j o h n s
- Web-Based Blood Bank System Concepts Turner, Toni. A Beginner's in 2007.
- Internet & World Wide Web.


**Websites:**

## URL listing

| Websites | Data collected |
|---|---|
| https://wikipedia.org | Searching of any information that will be used in documentation. |
| https://dev.sqlserver.com/doc | SQL server it performing in mainly depending on the one of the database using. |

| | |
|---|---|
| https://www.answers.com | Answers.com, online dictionary, encyclopedia and much more. |
| https://google.co.in | Any information searching and downloading. |
| https://training-classes.com | Designing part information as gathered |