WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

A Review on Analysis of Stock Market Prediction Using Supervised Machine Learning Algorithms

I. Indu

B. Tech Student, Department of Information Technology, GMR Institute of Technology, Rajam-532127, Andhra Pradesh, India.
E-mail: 21341A1250@gmrit.edu.in

ABSTRACT:

Stock market has a huge impact on the world economically. Greater economic development can be done in a nation when more investors attracted to a nations overall stock market which gives an overall confidence for an economical investment. So, predicting stock market has become a crucial and important task. In the past decades companies used traditional methods for prediction. Later on, with the increase in technology of Machine Learning a variety of additional approaches and algorithms are designed to predict the stock market. However, there are many factors like social media, sources from which data collected, false advertisements and many other have a huge impact on the predictions. Various parameters like Accuracy, Precision, F-score, Recall, Average error value are considered for the prediction. Various supervised machine learning algorithms like linear regression, logistic regression, Support vector Machine, Decision tree, Random Forest are anticipated. This study analyses the previous researches and addresses the efficient algorithms and which suits better for predicting the stock market.

Keywords: Stock Market, Stock Prediction, Machine Learning, Neural Networks, Accuracy

1. Introduction:

Stock prediction and analysis is one of the major factors that can affect the world's economy. A variety of supervised machine learning algorithms, such as Support Vector Machines, Random Forests, Neural Networks, and Sentimental analysis, are thoroughly examined for their effectiveness in constructing models and predicting changes in stock prices. The task of predicting stock market trends has always been a challenging problem for both finance experts and statisticians. Traditionally, stock market prediction relied on fundamental analysis, which considers a company's technique and fundamental information, or technical analysis, which focuses on historical stock prices and patterns. However, with the advancement of learning techniques and the emergence of data science, machine learning and deep learning methods have been increasingly employed to improve the accuracy and performance of stock market prediction models. The research work evaluates how these algorithms integrate historical stock data, market indicators, and macroeconomic variables to produce predictive models. In addition to discussing the various methodologies, the analysis also addresses the challenges faced by researchers in this domain, such as data quality, overfitting, and the efficient incorporation of external factors. The findings contribute to a deeper understanding of the strengths and limitations of various algorithms, aiding both academics and practitioners in making informed decisions when selecting and implementing predictive models in real-world financial scenarios.

2. Literature Survey:

The objective of the paper is to predict stock market movements by analyzing various sources of information, including social mood, historical stock data, search engine queries, Twitter, and web news.

Various categories of algorithms are used for stock market prediction. These categories include:

1. Traditional Machine Learning Algorithms are well-established and widely used algorithms that have been proven to be effective for a variety of tasks. They are typically based on statistical principles and are relatively simple to implement.

- Linear Regression: A simple algorithm that predicts a continuous target variable based on a linear relationship with one or more input variables.
- Logistic Regression: A similar algorithm that predicts a binary target variable (e.g., yes or no) based on a linear relationship with one or more input variables.
- Decision Trees: A tree-like structure that makes a series of decisions to classify or predict a target variable.
- Random Forests: An ensemble method that combines multiple decision trees to improve accuracy and reduce overfitting.

- Support Vector Machines (SVMs): A powerful algorithm that finds a hyperplane that best separates the data into two classes.

2. Sentiment Analysis is the process of extracting and classifying subjective information from text. It is commonly used to understand public opinion about a product, service, or brand.

- Binary sentiment analysis: Classifies text as positive or negative.
- Multi-class sentiment analysis: Classifies text as positive, negative, or neutral.
- Fine-grained sentiment analysis: Classifies text as expressing specific emotions, such as joy, sadness, anger, or fear.

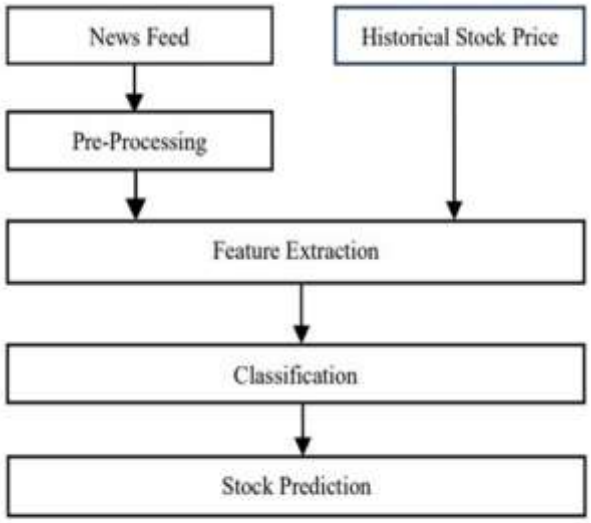
3. Predictive Analysis is the process of using historical data to predict future events or trends

- Sales forecasting: Predicting future sales of a product or service.
- Fraud detection: Identifying fraudulent transactions.
- Customer churn: Predicting which customers are likely to churn.
- Risk assessment: Assessing the risk of events such as loan defaults or medical complications.

3. Methodology:

Data Description: Data is collected from various sources, including social media data, web news data, news articles, historical stock prices, and textual information historical and real-time trading data, technical indicators, news analytics, and stock price time series data in various studies and models.

Architecture:



Data Preprocessing: The data is pre-processed for analysis through several methods, including data cleaning, data transformation, data reduction, feature selection, natural language processing (NLP), and dictionary-based approaches.

Evaluation Metrics: The performance measures used are:

- Precision
- Recall
- F1 score
- Accuracy

Precision is a measure of how many of the positive predictions made are correct (true positives).

Recall is a measure of how many of the actual positive instances correctly identified by the model (true positives).

F-score is a measure that combines precision and recall.

Accuracy is a measure of how many predictions are correct overall.

4. Results

The results and discussions related to stock prediction using machine learning techniques. Some of the key findings and discussions include:

- **Accuracy of Prediction Models:** The paper presents the accuracy of various prediction models, such as SVM, ANN, and LSTM. For example, the SVM model achieved an accuracy of 76.65%. The authors also discuss the need to improve the accuracy of prediction models, especially when dealing with textual data.

- **Comparison of Approaches:** The paper compares the performance of different approaches, such as statistical and machine learning techniques, for stock prediction. It suggests that a combination of these techniques is likely to be more useful for accurate stock prediction.

- **Limitations and Future Improvements:** The paper highlights the limitations of stock prediction models, such as difficulties in implementing natural language processing and challenges in transmitting data to the network. It also discusses the need for future improvements in prediction accuracy and model efficiency.

- **Sentiment Analysis and Stock Prediction:** The paper discusses the use of sentiment analysis from sources like Twitter and news articles for stock prediction. It presents the results of sentiment analysis and its impact on stock price prediction.

These results and discussions provide insights into the performance, challenges, and potential improvements in stock prediction using machine learning techniques.

	Normal Dataset				Leaked Dataset			
Classifier	**Precision**	**Recall**	**F Score**	**Accuracy**	**Precision**	**Recall**	**F Score**	**Accuracy**
Random Fores	0.72	0.53	0.61	0.54	0.77	0.94	0.85	0.79
Bagging	0.71	0.43	0.54	0.5	0.79	0.94	0.86	0.82
AdaBoost	0.76	0.33	0.46	0.48	0.69	0.87	0.77	0.68
Decision Tree	0.66	0.49	0.56	0.49	0.78	0.82	0.8	0.75
SVM	0.67	1	0.8	0.67	0.61	1	0.76	0.61
K-NN	0.68	0.61	0.64	0.54	0.67	0.77	0.71	0.62
ANN	0.66	0.52	0.58	0.5	0.63	0.88	0.73	0.61

5. Conclusion

This paper considered the analysis of various supervised learning algorithms for stock market prediction and confirmed that the SVM algorithm consistently outperforms other algorithms in terms of accuracy and reliability. While deep learning algorithms, such as recurrent neural networks (RNNs) and long short-term memory (LSTM) networks, have shown promising results in certain cases, their overall performance still lags behind that of SVM. This is likely due to the inherent complexity of deep learning models, which can make them difficult to train and optimize for stock market prediction tasks. Despite the dominance of SVM, it is important to note that no single algorithm is universally best for all stock market prediction scenarios. The choice of algorithm should be based on a careful consideration of the specific data set and prediction objectives. In some cases, hybrid approaches that combine multiple algorithms may be more effective. Overall, our findings suggest that SVM remains a powerful tool for stock market prediction, and its performance is comparable or superior to that of more recent deep learning algorithms. However, further research is needed to develop more robust and generalizable deep learning models for stock market prediction.

References:

List all the material used from various sources for making this project proposal

Research Papers:

1. Sirimevan, N., Mamalgaha, I. G. U. H., Jayasekara, C., Mayuran, Y. S., & Jayawardena, C. (2019, December). Stock market prediction using machine learning techniques. In 2019 International Conference on Advancements in Computing (ICAC) (pp. 192-197). IEEE.
2. Subasi, A., Amir, F., Bagedo, K., Shams, A., & Sarirete, A. (2021). Stock market prediction using machine learning. Procedia Computer Science, 194, 173-179.
3. Patel, R., Choudhary, V., Saxena, D., & Singh, A. K. (2021, June). Review of stock prediction using machine learning techniques. In 2021 5th International Conference on Trends in Electronics and Informatics (ICOEI) (pp. 840-846). IEEE.
4. Soni, P., Tewari, Y., & Krishnan, D. (2022). Machine Learning approaches in stock price prediction: A systematic review. In Journal of Physics: Conference Series (Vol. 2161, No. 1, p. 012065). IOP Publishing.

5. Vachhani, H., Obiadat, M. S., Thakkar, A., Shah, V., Sojitra, R., Bhatia, J., & Tanwar, S. (2020). Machine learning based stock market analysis: A short survey. In Innovative Data Communication Technologies and Application: ICIDCA 2019 (pp. 12-26). Springer International Publishing.

6. Peng, Z. (2019, January). Stocks analysis and prediction using big data analytics. In 2019 international conference on intelligent transportation, big data & smart City (ICITBS) (pp. 309-312). IEEE.

7. Soujanya, R., Goud, P. A., Bhandarkar, A., & Kumar, G. A. (2020). Evaluating future stock value asset using machine learning. Materials Today: Proceedings, 33, 4808-4813.

8. 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.

9. Pathak, A., & Shetty, N. P. (2019). Indian stock market prediction using machine learning and sentiment analysis. In Computational Intelligence in Data Mining: Proceedings of the International Conference on CIDM 2017 (pp. 595-603). Springer Singapore.

10. Soujanya, R., Goud, P. A., Bhandwalkar, A., & Kumar, G. A. (2020). Evaluating future stock value asset using machine learning. Materials Today: Proceedings, 33, 4808-4813.