iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** V **Month of publication:** May 2022

DOI: https://doi.org/10.22214/ijraset.2022.42747

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue V May 2022- Available at www.ijraset.com

Stock Price Prediction and Analysis using Machine Learning Techniques

Rucha Deshpande[1], Divyank Lunkad[2], Madhura Kunjir[3], Samiksha Ingle[4]
[1, 2, 3, 4]*Department of Computer Engineering, Pune Institute of Computer Technology, Pune, India*

Abstract: ***Stock market is one of the most sought-after investment options today. Investing in the stock market has the potential to multiply people's money in a short period of time, but it also carries many risks. To be successful in the stock market, it is necessary to conduct extensive research about the market and the forces influencing it. This examination involves a review of historical stock prices as well as real-time market developments. When deciding the position in the market, investors need to make an informed choice.***

Market-related information may be obtained from a variety of sources, including news articles, stock research websites, apps, and television channels. However, stock movement prediction by taking into consideration the elements influencing it would undoubtedly assist investors in making sound decisions. The research area of stock movement prediction has seen numerous advancements throughout the years. In the domain of machine learning, various methods of stock price prediction have been introduced.

In this paper, we have discussed different approaches to stock movement prediction based on diverse factors and technical indicators. We also propose a method for detecting market trends, which is a critical aspect in maintaining long-term market positions.

Keywords: Machine Learning, Stock Market, Prediction, Classification, Technical Indicators, Trend prediction

I. INTRODUCTION

Financial Research requires extensive data exploration and data analysis. The data that they have to explore is usually unstructured and not always available at one source. The data becomes redundant when collected from various sources and the information is often ambiguous and details are insignificant.

A financial research analyst needs to have access to all the data that is insightful as well as relevant to them to make decisions promptly. Various factors that influence stock performance need to be thoroughly investigated. Stock market prediction is difficult because of the volatile nature of the market and several factors that directly or indirectly influence the value of stocks. In this paper, we try to explore the three strategies that have used different machine learning algorithms to find the same output i.e. if the stock price will rise or fall.

To get a clear picture of the current stock market situation and make a conscientious decision, one has to do a thorough research of the company's history. It includes fundamental analysis and technical analysis of the company's stock. This procedure is complicated and time-consuming. Furthermore, it is extremely difficult for a human to consider all of the factors that influence stock movements. Investors require a tool that allows them to quickly and easily access the necessary and relevant information without extracting it manually. All the three approaches described in this paper consider appropriate factors w.r.t. The strategy used and try to predict the state of stocks in the near future. Presenting the results of all the strategies to the investors can help them make a deterministic decision.

II. LITERATURE SURVEY

A. Stock Movement Prediction with Financial News using Contextualized Embedding from BERT

BERT also called Bidirectional Encoder Representations from Transformers has used contextualized vector representations of the news headlines to predict the stock prices. The model has obtained the state-of-the-art result on this stock movement prediction task. It has proposed a Fine-Tuned Contextualized-Embedding Recurrent Neural Network (FT-CE-RNN) to predict the stock price movement based on the headlines. The FT-CE-RNN model is trained using Bloomberg News dataset and it generates contextualized embeddings with domain-specific knowledge using all the useful hidden vectors from the BERT model. The accuracy obtained using the FT-CE-RNN model is 74.5%.

B. Stock Market Analysis: A Review and Taxonomy of Prediction Techniques

A concise review of stock markets and taxonomy of stock market prediction methods have been provided in this paper. The paper has focused on some of the research achievements in stock analysis and prediction. It has discussed technical, fundamental, short- and long-term approaches used for stock analysis. The paper has looked upon the research in the past, but mainly focused on modern techniques, highlighting some of the main challenges the field poses and recent achievements for stock analysis and prediction.

C. Stock Prediction Using Twitter Sentiment Analysis

The paper has applied sentiment analysis and machine learning principles to find the correlation between ”public sentiment” and ”market sentiment”. It has used Twitter data to predict public mood and has hence used the predicted mood and previous days’ DJIA values to predict the stock market movements. In order to test the results, it has proposed a new cross-validation method for financial data and obtained 75.56% accuracy using Self Organizing Fuzzy Neural Networks (SOFNN) on Twitter feeds and DJIA. It has also implemented a naive portfolio management strategy based on the predicted values. This work is based on Bollen et al’s famous paper which predicted the same with 87% accuracy.

D. Deep Learning for Event-Driven Stock Prediction

This paper has proposed a deep learning method for event-driven stock market prediction. First, events have been extracted from news text and represented as dense vectors, trained using a novel neural tensor network. Second, a deep convolutional neural network has been used to model both short-term and long-term influences of events on stock price movements. Experimental results show that the model can achieve nearly 6 percent improvements in S&P 500 index prediction and individual stock prediction, respectively, compared to state-of-the-art baseline methods. In addition, market simulation results show that the system is more capable of making profits than previously reported systems trained on S&P 500 stock historical data.

E. A LSTM model using Risk Estimation loss function for stock trades in market

The LSTM model using the Risk Estimation loss function has used the effective representations that are extracted from historical data. Then a neural network based on LSTM has been constructed to learn useful knowledge to direct the trading behaviors. A loss function has elaborately been designed to ensure the network optimizes the profit and minimizes the risk. Finally, according to the predictions of this neural network, buying and selling plans are carried out.The model has used the past 30-day data to train itself and then predict the next day's closing values. The feature window is of 30 days and such 30 days feature vectors are being used to train the model. The accuracy obtained using the above model is 54%.

III. METHODOLOGIES

Stock prices keep on continuously changing and predicting them is a complex task. We have proposed three methodologies, each dependent on the stock’s closing price for predicting the future value. Calculating the exact price of the next day is never possible so we have tried to predict whether a stock will go high or low on the next day through these strategies.

A. One-Day Ahead Strategy

Stock market price trend prediction is always a popular issue for academics from both financial and technological areas since the stock market is one of the primary topics in which investors are interested. Our goal in this study is to develop a cutting-edge price trend prediction model that focuses on short-term price trend prediction. The primary aim of this binary model is to predict the movement of stocks on the next day. Every stock can be categorized in a sector such as IT, Banks, etc. For this model, we need the returns, sentiment score and volume of the stock and the respective sector each day. Every day, a web crawler is triggered, which gathers all the news of a stock, and then a sentiment model classifies it as positive or negative news. The mathematical model is as follows:

$R_t = \log(P_{t+1} / P_t)$

$S_t = (\text{Positive News}_t - \text{Negative News}_t) / (\text{Positive News}_t + \text{Negative News}_t)$

Feature vector = $[x_{t-2}, x_{t-1}, x_t]$

x_t is vector of : $[R_t, R_{t0}, S_t, S_{t0}, V_t, V_{t0}]$

where,
t = day
R_t = return of a company for day 't'
R_{t0} = overall return for day 't' (average of returns of all stocks of the sector)
S_t = sentiment of a company for day 't'
S_{t0} = overall sentiment for day 't' (average of the sentiment of all stocks of the sector)
V_t = volume of a company for day 't'
V_{t0} = overall volume for day 't' (sum of volumes of all stocks of the sector)
P_t = closing price of day 't'

Label: sign(output)
Output: high or low
High represents that the price will rise the next day.
Low represents that the price will fall the next day.

B. Five-Day Ahead Strategy

The previous strategy does not help in predicting the movement of a stock in the immediate week period. The previous model works on numerical data as well as sentiment data. But there is another important feature, which every trader uses - MarketIndicators. Market indicators are quantitative in nature and aim to anticipate market movements by interpreting stock or financial index data. Market indicators are a type of technical indicator that is made up of formulae and ratios. They help investors make better investing and trading decisions. This model uses indicators to predict for a longer duration. To calculate the indicators, we have used an open-source library named Ta-Lib. This is a ternary classification model, with the possible outputs being 1,0 and -1. 1 indicates that the closing price of the stocks is greater than today's closing price in at least 4 out of the next 5 days. -1 indicates that the closing price of the stocks is lower than today's closing price in at least 4 out of the next 5 days. Otherwise, it is 0.

Feature vector = [
Simple_moving_average (SMA),
Exponential_moving_average (EMA),
Relative_strength_index (RSI),
Standard_deviation (SD),
Average_directional_index (ADI),
Stochastic_oscillator (SO),
Moving_average_convergence_divergence (MACD),
Bollinger_bands (BB)
]

Output = 1,0 or -1

C. Trade Max Strategy

An attempt to predict the next day's prices using the data of the past 30 days has been made in this model called Trade Max. The author has designed a solution that uses the historical data of the stock and converts it into useful features for the model. These features are then trained on the Random Forest Model to predict for the next day. Finally buying and selling of the stocks could be carried out accordingly.

The features for the model include 12 different variables namely open prices, high prices, low prices, close prices, the volume of the stock, MACD, RSI, BOLL, MA, VMA, Cross price, Price Volume. All the above 12 variables are stored in a feature array and passed on to the model. This data was collected id of the past 30 days. The open, high, low, and close prices and volume of the stock are being crawled from the National Stock Exchange (NSE). For the other variables, the 'Talib' library is used. The Talib library uses the closing prices array to calculate the MACD data. Similarly, the other feature variables are calculated by using the Talib library which in turn uses the open/high/low/close/volume data that has been extracted from NSE.

Feature vector = [
Open_price (OROCP),
High_price (HROCP),
Low_price (LROCP),
Close_price (ROCP),
Volume,
Moving_average_convergence_divergence (MACD),
Relative_strength_index (RSI),
BOLL,
Moving_average (MA),
Volume_moving_average (VMA),
Cross_price,
Price_volume
]

Output : 0 and 1

The label for the model is calculated using the closing prices. For a 30 day data as feature input the 29th, 30th and 31st day's closing prices are used for calculating the label. A decay variable is used which decays the value of the label for every iteration. The label value is a real number and it is then converted into binary data. A binary classifier function is used to generate the 0 and 1 combination of the label. If the label value is greater than or equal to 0 then the label is classified as '1' meaning the value tomorrow will increase and if the value of the label is less than 0 then it is classified as '0' meaning the value tomorrow will decrease.
The Random Forest Model is then fed with the features array and the binary classified label value for 30 days and the model predicts for 31st day. The output of the model is a binary value. If the output is 1 meaning the value tomorrow will increase and if it is 0 meaning the value tomorrow will decrease.

D. Trend Prediction

When analyzing the stock market, it is crucial to detect market regimes. Investing entails buying and selling stocks for long-term gains, which necessitates understanding the state of the market over a longer period of time. Existing solutions take various approaches, such as identifying market regimes from equity data using a Hidden Markov Model, or using a correlation matrix and clustering windows on asset returns and factor returns. However, in order to include the effects of trend in the methodologies outlined above, we attempted to forecast the trend as a variable. In this section, we have discussed a method to detect the current trend in the market by using a 'trend variable'.

The formula to calculate the Trend Variable for predicting the state of the market is as follows:

$$T_t^d = \frac{abs(P_t - P_{t-1})}{\sum_{n=1}^{d} abs(P_t - P_{t-1})}$$

Where,
$P_t \rightarrow$ Closing price on day 't'
$T_t^d \rightarrow$ Trend variable on day 't', based on the behaviour on previous 'd' days
d = {2, 3, 5, 10}
0 → pure sideways
1 → pure trend

We predicted the trend variable for the next 5 days using the ARMA model in time series. The training data has trend variable values of the previous 40 days. Based on these values, the trend variables for the next 5 days are predicted.

The following graph shows 5-day trend variable over a period of 2 years

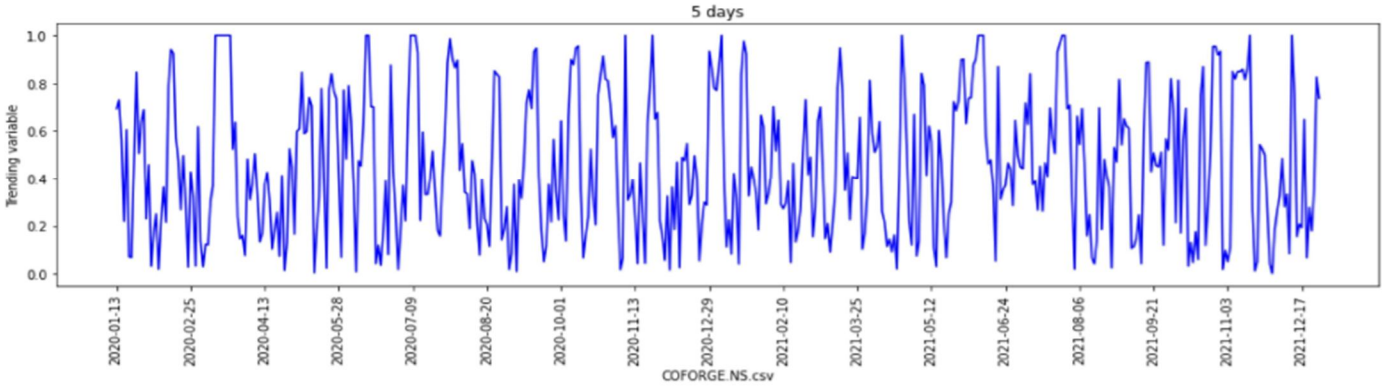


Fig. 1 Distribution of Trend Variable

The following graph shows 5-day trend prediction using past 40-day data

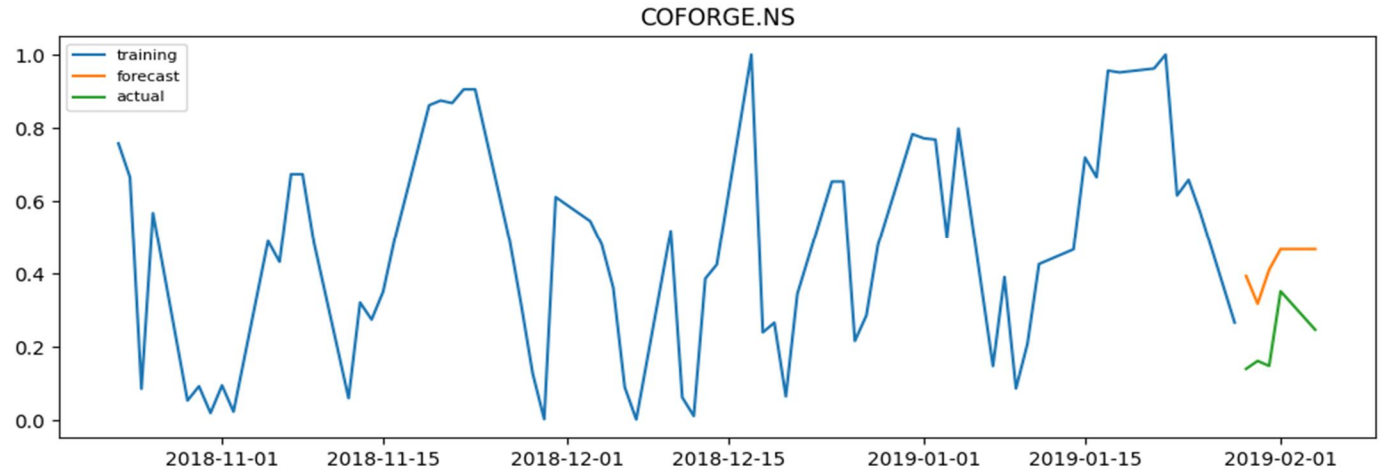


Fig. 2 Trend Prediction results for COFORGE.NS

IV. RESULTS

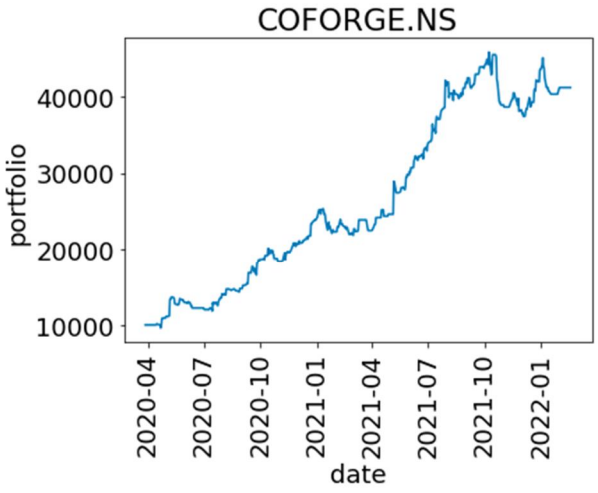
The one-day ahead, five-day ahead, and Trade Max models all aid in the prediction of stock values based on their various methodologies. Predictions derived from the approaches described above can be assessed in a variety of ways. The one-day ahead forecast provides investors with a quick overview of predicted stock movement on the next day, whilst the five-day ahead prediction provides an overall perspective of projected stock movement over a longer period of time. The trade max prediction is a more extensive study that employs technical indicators to make forecasts, but it requires more data than the other two methods.

Below is the accuracy table of the 3 strategies:

TABLE I. Results

Sr. No.	Model	Accuracy
1.	One-Day Ahead Prediction	58%
2.	Five-Day Ahead Prediction	55%
3.	Trade Max Prediction	68%

We experimented with the one-day ahead model by investing virtual money in every stock and used the prediction to buy or sell the stocks. The profit and loss (P&L) results are as followed:



	ticker_id	investment	p&l
0	COFORGE.NS	10000	31172.196599
1	HCLTECH.NS	10000	6288.698059
2	INFY.NS	10000	7566.690477
3	MINDTREE.NS	10000	29687.912426
4	TCS.NS	10000	8976.170485
5	TECHM.NS	10000	12970.413924
6	WIPRO.NS	10000	8554.484108

```
Total p&l:  105216.56607771208
```

V. CONCLUSION

In this paper, we have discussed three strategies for predicting stock price movement for the next trading day using data from the previous days. We have also provided a method for detecting existing market trends and anticipated changes in order to determine market position. These strategies can further be implemented to maintain a user portfolio and provide real-time suggestions for investing. Unlike many other ways of stock price prediction that study the historical data, our methods take into account the sentiment analysis of real-time events for the company to enhance predictions. This can help a user to make informed and intelligent decisions. The real-time events analyzed in this study are limited to news articles and tweets about the company. The methods of sentiment calculation are out of the scope of this paper.

VI. ACKNOWLEDGMENT

We would like to thank Dr. Aniruddha Pant, CEO of AlgoAnalytics Pvt. Ltd., for his constant guidance and enthusiastic support during our research.

We would also like to thank Mrs. A. G. Phakatkar, Assistant Professor at PICT, and our project guide, for her timely advice and detailed research insights.

REFERENCES

[1] Qinkai Chen, “Stock Movement Prediction with Financial News using Contextualized Embedding from BERT” arXiv:2107.08721, 19 Jul 2021.

[2] Xiao Ding, Yue Zhang, Ting Liu and Junwen Duan,“Using Structured Events to Predict Stock Price Movement: An Empirical Investigation, Research Center for Social Computing and Information Retrieval Harbin Institute of Technology”, Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, pp. 1415-1425, January 2014.

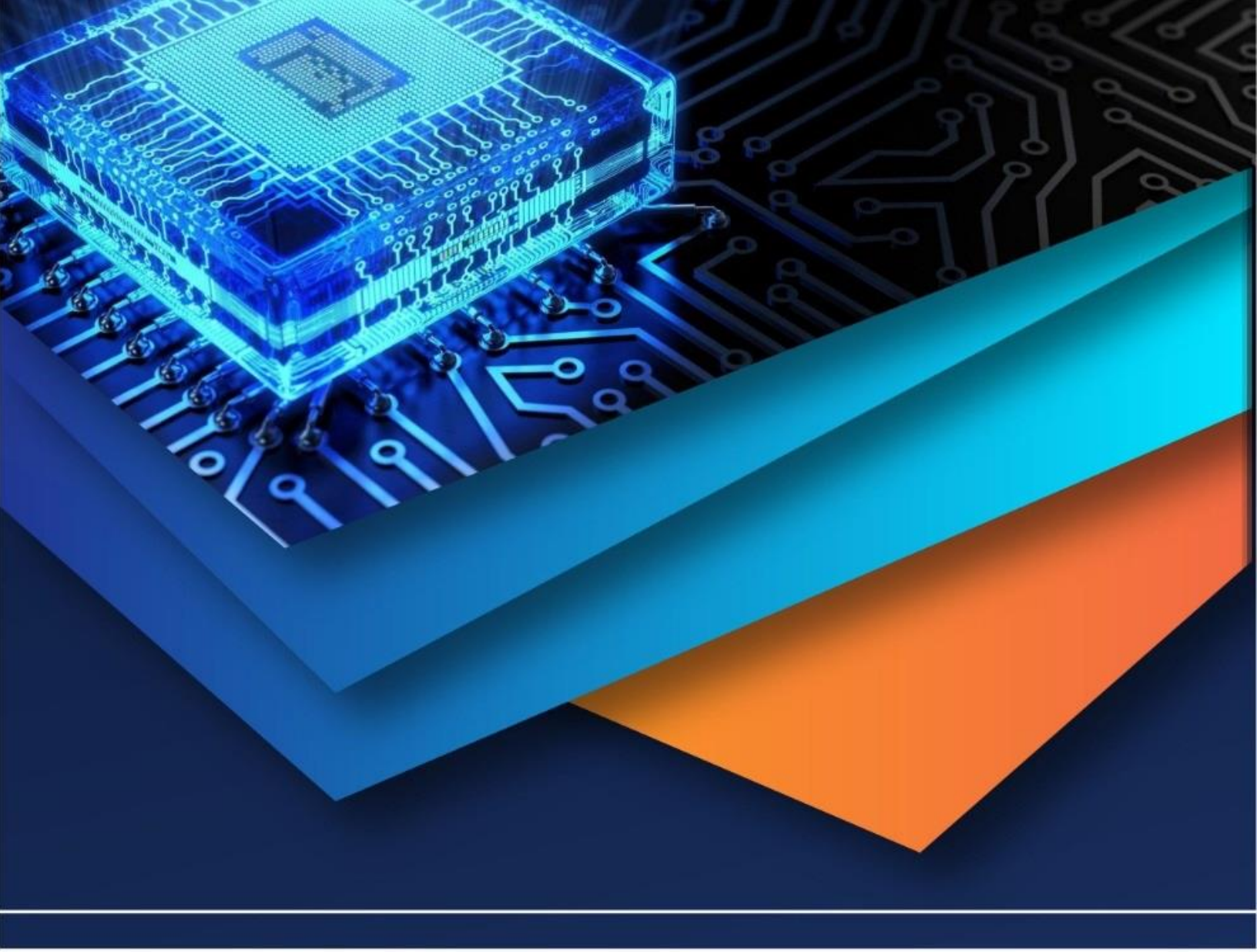
[3] Anshu Mittal and Arpit Goel, “Stock Prediction Using Twitter Sentiment Analysis", Standford University, CS229, vol. 15, 2012.

[4] Dev Shah, Haruna Isah and Farhana Zulkernine,"Stock Market Analysis: A Review and Taxonomy of Prediction Techniques", International Journal of Financial Studies, 7(2), 2019.

[5] Adebiyi, A.A., Ayo, C.K., Adebiyi, M.O. and Otokiti, S.O., "Stock price prediction using neural network with hybridized market indicators", Journal of Emerging Trends in Computing and Information Sciences, 2012, 3, 1–9.

[6] Ariyo, A.A., Adewumi, A.O. and Ayo, C.K., "Stock price prediction using the ARIMA model", In Proceedings of the 2014 UKSim-AMSS 16th International Conference on Computer Modelling and Simulation, pp. 106–112, 2014.

[7] Ding, X., Zhang, Y., Liu, T. and Duan, J., "Using structured events to predict stock price movement: An empirical investigation", In Proceedings of the Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp.1415–1425, 2014.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY