ISSN No. 0976-5697

VOLUME 13 SPECIAL ISSUE 1, JUNE 2022

International Journal of Advanced Research in Computer Science

RESEARCH PAPER

Available Online at www.ijarcs.info

FUTURE MARKET TRENDS PREDICTION WITH PYTHON AND MACHINE LEARNING

Lakshyaraj Singh Dalawat
Department of Computer Engineering
Geetanjali Institute of Technical Studies
Udaipur, Rajasthan 313001, India
laksh.dalawat@gmail.com

Divya Soni
Department of Computer Engineering
Geetanjali Institute of Technical Studies
Udaipur, Rajasthan 313001
divyarunwar063@gmail.com

Lovisha Jain
Department of Computer Engineering
Geetanjali Institute of Technical Studies
Udaipur, Rajasthan 313001, India
lavishajain010@gmail.com

Bhowmick Vyas
Department of Computer Engineering
Geetanjali Institute of Technical Studies
Udaipur, Rajasthan 313001, India
bhowmickvyas26@gmail.com

Mr. Bhupendra Teli
Assistant Professor, Department of Computer Science and Engineering
Geetanjali Institute of Technical Studies
Udaipur, Rajasthan 313001, India
bhupendra.teli@gits.ac.in

Abstract: Predicting Future Stock Market Trends with Python & Machine Learning with the advent of technological marvels like global digitization, the prediction of the stock market has entered a technologically advanced era, revamping the old model of trading. With the ceaseless increase in market capitalization, stock trading has become a centre of investment for many financial investors. Many analysts and researchers have developed tools and techniques that predict stock price movements and help investors in proper decision-making. Advanced trading models enable researchers to predict the market using non-traditional textual data from social platforms. The application of advanced machine learning approaches such as text data analytics and ensemble methods have greatly increased the prediction accuracies.

Keywords: Future Market Trend, Machine learning model.

I. INTRODUCTION

A advancement in the fundamental aspects of information technology over the last few decades has altered the route of businesses. As one of the most captivating inventions, financial markets have a pointed effect on the nation's economy. The World Bank reported in 2018 that the stock market capitalization worldwide has surpassed 68.654 trillion US$. Over the last few years, stock trading has become a centre of attention, which can largely be attributed to technological advances. Investors search for tools and techniques that would increase profit and reduce the risk. However, Stock Market Prediction (SMP) is not a simple task due to its non-linear, dynamic, stochastic, and unreliable nature.

Market Data: Market data are the temporal historical price-related numerical data of financial markets. Analysts and traders use the data to analyze the historical trend and the latest stock prices in the market. They reflect the information needed for the understanding of market behaviour. The market data is usually free, and can be directly downloaded from the market websites. Various researchers have used this data for the prediction of price movements using machine learning algorithms. The previous studies have focused on two types of predictions. Some studies have used stock index predictions like the Dow Jones Industrial Average (DJIA), Nifty, Standard and Poor's (S&P) 500, National Association of Securities Dealers Automated Quotations (NASDAQ), the Deutscher Aktien Index (DAX) index, and multiple indices.

Technical Analysis: Technical analysis is the study of stock prices to make a profit, or to make better investment decisions.
Technical analysis predicts the direction of the future price movements of stocks based on their historical data, and helps to analyze 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.

Fundamental Analysis: Fundamental analysis calculates a 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 agreeing 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.

II. LITERATURE SURVEY

"What other people think" has always been an important piece of information for most of us during the decision-making process. The Internet and the Web have now (among other things) made it possible to find out about the opinions and experiences of those in the vast pool of people that are neither our personal acquaintances nor well-known professional critics — that is, people we have never heard of. And conversely, more and more people are making their opinions available to strangers via the Internet. The interest that individual users show in online opinions about products and services, and the potential influence such opinions wield, is something that is driving force for this area of interest. And there are many challenges involved in this process which need to be walked all over in order to attain proper outcomes out of them. In this survey we analyzed basic methodology that usually happens in this process and measures that are to be taken to overcome the challenges being faced.

III. PROBLEM STATEMENT

Time Series forecasting & modelling plays an important role in data analysis. Time series analysis is a specialized branch of statistics used extensively in fields such as Econometrics & Operation Research. Time Series is being widely used in analytics & data science. Stock prices are volatile in nature and price depends on various factors. The main aim of this project is to predict stock prices using long short-term memory (LSTM).

IV. OBJECTIVE

The main objective of this paper is to propose an AI based solution for predicting future market trends for any organization, production or sales in this we have taken into account Microsoft data.
Stock market is one of the major fields that investors are dedicated to, thus stock market price trend prediction is always a hot topic for researchers from both financial and technical domains.
As the artificial intelligence techniques evolved in recent years, many proposed solutions attempted to combine machine learning and deep learning techniques based on previous approaches, and then proposed new metrics that serve as training features such as Liu and Wang.
There are three key contributions of our work :

- a new dataset extracted and cleansed
- a comprehensive feature engineering, and
- a customized long short-term memory (LSTM) based learning model.

V. METHODOLOGY

Proposed systems: The prediction methods can be roughly divided into two categories, statistical methods and artificial intelligence methods. Statistical methods include logistic regression model, ARCH model, etc. Artificial intelligence methods include multi-layer perception, convolution neural network, naive Bayes network, back propagation network, single-layer LSTM, support vector machine, recurrent neural network, etc. We used a Random Forest Classifier.

Random Forest Classifier: The **random forest** is a classification algorithm consisting of many decision trees. It uses bagging and features randomness when building each individual tree to try to create an uncorrelated forest of trees whose prediction by committee is more accurate than that of any individual tree.

Working of Random Forest: A random forest algorithm consists of many decision trees. The 'forest' generated by the random forest algorithm is trained through bagging or bootstrap aggregating. Bagging is an ensemble meta-algorithm that improves the accuracy of machine learning algorithms.
The (random forest) algorithm establishes the outcome based on the predictions of the decision trees. It predicts by taking the average or mean of the output from various trees. Increasing the number of trees increases the precision of the outcome.

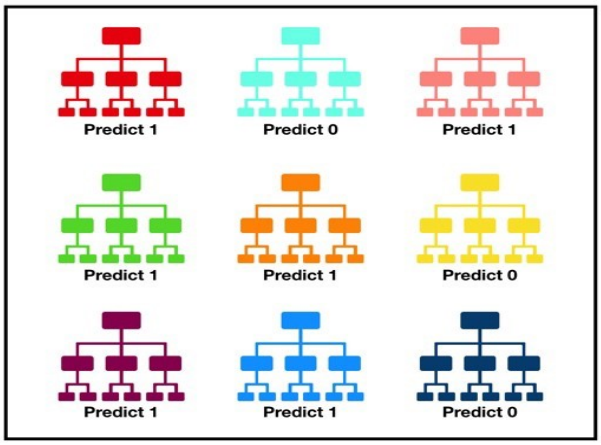

Fig1. Tally: Six 1s and three 0s (prediction1)

Features of a Random Forest Algorithm :-It's more accurate than the decision tree algorithm.

- It provides an effective way of handling missing data.
- It can produce a reasonable prediction without hyper-parameter tuning.
- It solves the issue of over fitting in decision trees.
- In every random forest tree, a subset of features is selected randomly at the node's splitting point.

How random forest algorithm works :

- Understanding decision trees- Decision trees are the building blocks of a random forest algorithm. A decision tree is a decision support technique that forms a tree-like structure. An overview of decision trees will help us understand how random forest algorithms work.
- A decision tree consists of three components: decision nodes, leaf nodes, and a root node.

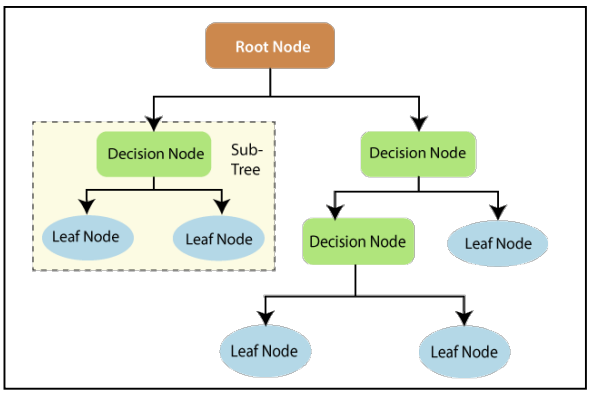

Fig2. Decision Tree

Applying decision trees in random forest- The main difference between the decision tree algorithm and the random forest algorithm is that establishing root nodes and segregating nodes is done randomly in the latter. The random forest employs the bagging method to generate the required prediction.

- Bagging involves using different samples of data (training data) rather than just one sample. A training dataset comprises observations and features that are used for making predictions. The decision trees produce different outputs, depending on the training data fed to the random forest algorithm. These outputs will be ranked, and the highest will be selected as the final output.

- Classification in random forests- Classification in random forests employs an ensemble methodology to attain the outcome. The training data is fed to train various decision trees. This dataset consists of observations and features that will be selected randomly during the splitting of nodes.

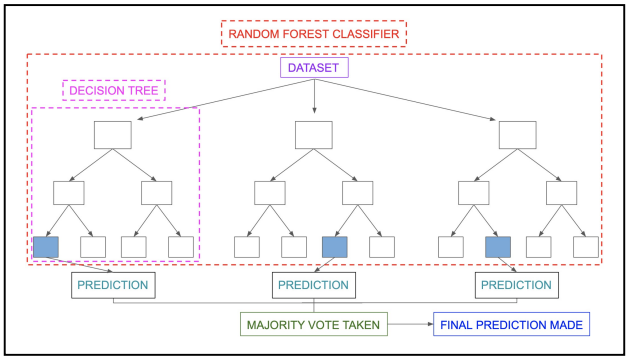

Fig3. Random Forest Classifier

Random Forest works in two-phase first is to create the random forest by combining N decision tree, and second is to make predictions for each tree created in the first phase.
The Working process can be explained in the below steps and diagram

- **Step-1:** Select random K data points from the training set.
- **Step-2:** Build the decision trees associated with the selected data points (Subsets).
- **Step-3:** Choose the number N for decision trees that you want to build.
- **Step-4:** Repeat Step 1 & 2.
- **Step-5:** For new data points, find the predictions of each decision tree, and assign the new data points to the category that wins the majority votes.

VI. IMPLEMENTATION

Let's say we want to make money by buying stocks. Since we want to make money, we only want to buy stock on days when the price will go up (we're against shorting the stock). We'll create a machine learning algorithm to predict if the stock price will increase tomorrow. If the algorithm says that the price will increase, we'll buy stock. If the algorithm says that the price will go down, we won't do anything.
We want to maximize our true positives - days when the algorithm predicts that the price will go up, and it actually goes up. Therefore, we'll be using precision as our error metric for our algorithm which is true positives / (false positives + true positives). This will ensure that we minimize how much money we lose with false positives (days when we buy the stock, but the price actually goes down).
This means that we will have to accept a lot of false negatives - days when we predict that the price will go down, but it actually goes up. This is okay, since we'd rather minimize our potential losses than maximize our potential gains.

Method: Before we get to machine learning, we need to do a lot of work to acquire and clean up the data. Here are the steps we'll follow:

- Download historical stock prices from Yahoo finance
- Explore the data
- Setup the dataset to predict future prices using historical prices
- Test a machine learning model
- Setup a back testing engine
- Improve the accuracy of the model

Downloading the data: First, we'll download the data from Yahoo Finance. We'll save the data after we download it, so we don't have to re-download it every time (this could cause our IP to get blocked).
We'll use data for a single stock (Microsoft) from when it started trading to the present.
As we can see, we have one row of data for each day that Microsoft stock was traded. Here are the columns:

- Open - the price the stock opened at.
- High - the highest price during the day
- Low - the lowest price during the day
- Close - the closing price on the trading day
- Volume - how many shares were traded

	Open	High	Low	Close	Volume	Dividends	Stock Splits
1986-03-13	0.055898	0.064119	0.055898	0.061378	1031788800	0.0	0.0
1986-03-14	0.061378	0.064667	0.061378	0.063570	308160000	0.0	0.0
1986-03-17	0.063570	0.065215	0.063570	0.064667	133171200	0.0	0.0
1986-03-18	0.064667	0.065215	0.062474	0.063022	67766400	0.0	0.0
1986-03-19	0.063022	0.063570	0.061378	0.061926	47894400	0.0	0.0

Fig4. Data Table

Preparing the data:

First, we'll identify a target that we're trying to predict. Our target will be if the price goes up or down tomorrow. If the price went up, the target will be 1 and if it went down, the target will be 0. Next, we'll

shift the data from previous days "forward" one day, so we can use it to predict the target price. This ensures that we don't accidentally use data from the same day to make predictions! Then, we'll combine both so we have our training data.

	Actual_Close	Target	Close	Volume	Open	High	Low
1986-03-14	0.063570	1.0	0.061378	1.031789e+09	0.055898	0.064119	0.055898
1986-03-17	0.064667	1.0	0.063570	3.081600e+08	0.061378	0.064667	0.061378
1986-03-18	0.063022	0.0	0.064667	1.331712e+08	0.063570	0.065215	0.063570
1986-03-19	0.061926	0.0	0.063022	6.776640e+07	0.064667	0.065215	0.062474
1986-03-20	0.060282	0.0	0.061926	4.789440e+07	0.063022	0.063570	0.061378

Fig5. Data Table (prepared)

Creating a machine learning model:

Because we're dealing with time series data, we can't just use cross-validation to create predictions for the whole dataset. This will cause leakage where data from the future will be used to predict past prices. This doesn't match with the real world, and will make us think that our algorithm is much better than it actually is.

Instead, we'll split the data sequentially. We'll start off by predicting just the last 100 rows using the other rows.

We'll use a random forest classifier to generate our predictions. This is a good "default" model for a lot of applications, because it can pick up nonlinear relationships in the data, and is somewhat robust to over fitting with the right parameters.

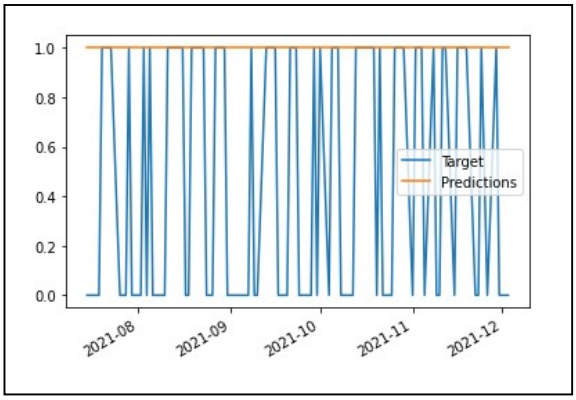

Fig6. Graph (prediction)

Improving accuracy:

The model isn't very accurate, but at least now we can make predictions across the entire history of the stock. For this model to be useful, we have to get it to predict more accurately.

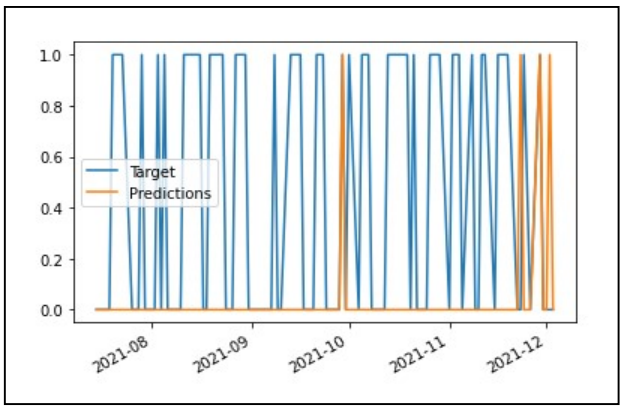

Fig7. Graph (after improving accuracy)

VII. CONCLUSION

This solution is to build an Artificial Intelligence based Model which can predict Microsoft trends by analyzing the past year's data (/ trends). Microsoft trends can be predicted and thus it helps to grow the business. Also, through this we find out This is an ideal algorithm for developers because it solves the problem of overfitting of datasets. It's a very resourceful tool for making accurate predictions needed in strategic decision making in organizations.

VIII. REFERENCES

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

[2] Upadhyay, A.; Bandyopadhyay, G. Forecasting Stock Performance in Indian Market using Multinomial Logistic Regression. J. Bus. Stud. Q. 2012, 3, 16–39.

[3] Ali Khan, J. Predicting Trend in Stock Market Exchange Using Machine Learning Classifiers. Sci. Int. 2016, 28, 1363–1367.

[4] Dutta, A. Prediction of Stock Performance in the Indian Stock Market Using Logistic Regression. Intern. J. Bus. Inf. 2012, 7, 105–136.

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

[6] Yolcu, O.C.; Lam, H.-K. A combined robust fuzzy time series method for prediction of time series. Neurocomputing 2017, 247, 87–101. [CrossRef]

[7] Vivek Rajput and Sarika Bobde - Stock market prediction using hybrid approach.

[8] Anurag Sinha Department of computer science, Student, Amity University Jharkhand Ranchi, Jharkhand (India), 834001 - Stock Market Prediction Using Machine Learning.

[9] S. Selvin, R. Vinayakumar, E. A. Gopalkrishnan, V. K. Menon and K. P. Soman, "Stock price prediction using LSTM, RNN and CNN-sliding window model," in International Conference on Advances in Computing, Communications and Informatics, 2017.

[10] Murtaza Roondiwala, Harshal Patel, Shraddha Varma, "Predicting Stock Prices Using LSTM" in Undergraduate Engineering Students, Department of Information Technology, Mumbai University, 2015.

[11] Xiongwen Pang, Yanqiang Zhou, Pan Wang, Weiwei Lin, "An innovative neural network approach for stock market prediction", 2018.

[12] Ishita Parmar, Navanshu Agarwal, Sheirsh Saxena, Ridam Arora, Shikhin Gupta, Himanshu Dhiman, Lokesh Chouhan Department of Computer Science and Engineering National Institute of Technology, Hamirpur – 177005, INDIA - Stock Market Prediction Using Machine Learning.

[13] Pranav Bhat Electronics and Telecommunication Department, Maharashtra Institute of Technology, Pune. Savitribai Phule Pune University - A Machine Learning Model for Stock Market Prediction.

[14] Anurag Sinha Department of computer science, Student, Amity University Jharkhand Ranchi, Jharkhand (India), 834001 - Stock Market Prediction Using Machine Learning.

[15] V Kranthi Sai Reddy Student, ECM, Sreenidhi Institute of Science and Technology, Hyderabad, India - Stock Market Prediction Using Machine Learning.

