Stock Market Analyzing And Prediction Using Machine Learning Techniques.

[1] Dr. Vandana Navale, [2] Prathamesh Kadale, [3]Priya Bhalchim, [4]Akanksha Salunke, [5]Shridhan Badve

[1]Computer,AISSMS COE,No 1,Kennedy Road,Pune,411001,Maharashtra,India.

Abstract: In this research, we examine current approaches and then suggest fresh approaches for predicting the stock market. We tackle the issue from three separate angles: fundamental analysis, technical analysis, and machine learning application. The weak form of the Efficient Market Hypothesis is supported by research showing that historical prices do not contain relevant information but that out-of-sample data may be predictive. News that is pertinent to any company listed on the stock market has an impact on future stock movement. We demonstrate how machine learning and fundamental analysis could be used to help investors make decisions. Techniques for machine learning come in handy here. Intelligent investors can use machine learning techniques to predict the behaviour of the stock market since they are aware that the examination of numerical time series produces accurate findings. Utilise the SVM algorithm to train the dataset and forecast the stock market.

Index Terms - **Machine Learning, SVM, Feature Extraction, Classification, Data Analytics, Stock Market, Decision Tree .**

I.   Introduction:

One of the top enterprises for middle class individuals to profit from is stock investing. Following that, high class investors and traders engage in actual trading activity. For investors, the most crucial factor is the company's share price, which is constantly in flux. To avoid financial loss and ultimately to make money, one must keep an eye on the share market's live price at all times and make quick decisions. You must research the company's past financial performance and planned course of action for this. You can choose to invest based on a general market and company investigation. But there are limitations to research because it is impossible to know for sure whether a study or analysis is accurate. The main factors that affect stock rate include the company's market history, propensity to continue doing business through good times and bad, policies, and announcements. Being a successful investor requires extensive experience in a challenging field of work.

II.   Problem Statement

Financial analysts are having difficulty trading since they are unsure of which stocks to buy or sell in order to maximise their profits. It would be quite difficult to manually or individually analyse all of this data. Machine learning methods can be useful in this situation. Intelligent investors can utilise machine learning techniques to predict the behaviour of the stock market by being aware that study of numerical time series yields accurate results. As a result, financial experts will be able to predict how the stock they are interested in will behave and take appropriate action.

The objective:

a.   To analyaze the stock market predication.

b.   To Predicate the Future Stock price or Flow of stock price graph of companies.

c.   The project will evaluate some existing strategies from a rigorous scientific perspective and provide a quantitative evaluation of new strategies.

d.   This project will focus exclusively on predicting the daily trend (price movement) of individual stocks.

e.   The project will make no attempt to deciding how much money to allocate to each prediction. More so, the project will analyse the accuracies of these predictions. Additionally, a distinction must be made between the trading algorithms studied in this project and high frequency trading algorithms.

III. Motivation

We are motivated by the shortcomings of the current system. The system is designed to be self-analyzing and capable of expressing predictions.

Predicting stock prices is a well-known and significant problem. With a successful stock forecasting model, we can learn more about market behaviour over time, finding trends that otherwise might not have been noticed. With the PC's inescapable computational power, AI will be a useful method to handle this problem. However, the people in general stock dataset is excessively constrained for some AI calculations, and requesting more highlights may regularly cost a huge sum of money. In order to improve our results, we will outline a structure in this paper that allows us to integrate client expectations into the current AI calculation. The idea is that the value is predictable if we have complete knowledge of the stock exchange in existence today (for all particular merchants). Therefore, if we can only gather a portion of the information, we can try to raise the current expectation level. Obtaining daily customer projections is a practical job because to the advent of the Internet, informal organisations, and online social associations1. Our goal is to construct an open system that combines historical data with client forecasts to create a more realistic model that will be advantageous to all parties.

IV. Proposed Methodology

Support Vector Machine (SVM):
The approach of supervised learning Support Vector Machine is employed to solve classification and regression issues. This SVM algorithm's primary objective is to locate the best line or decision boundary that can quickly categorise further data points in the future. This optimal decision boundary is known as a hyperplane.

Decision Tree Algorithm:
Classification and regression issues can be resolved using the supervised learning technique known as a decision tree, but it is frequently preferred for doing so. It is a tree-structured classifier, with each leaf node denoting the classification result and inner nodes denoting the dataset's features. The Decision Node and Leaf Node are the two nodes of a decision tree. Decision nodes are used to make decisions and have many branches, whereas Leaf nodes are the outcomes of decisions and do not have any more branches.

Data Preprocessing:
The process of converting unprocessed data into something a machine learning model can use is known as data preparation. It is both the initial and most important phase in the development of a machine learning model. It is not always the case that we are given the clean and ready data when working on a machine learning project. Every time you work with data, you must also prepare and clean it. When working on a machine learning project, it is not always the case that we are presented with the clean and prepared data. Additionally, you must format and clean up your data every time you work with it. So, for this, we employ a data pretreatment activity.

Feature extraction:
A large amount of raw data is sorted into smaller, more manageable groupings using the feature extraction approach. These enormous data sets are similar in that they contain a large number of variables that require a lot of processing capacity. "Feature extraction" refers to methods for selecting and/or combining variables into features, which greatly decreases the quantity of data that needs to be processed while accurately and completely describing the initial data set.
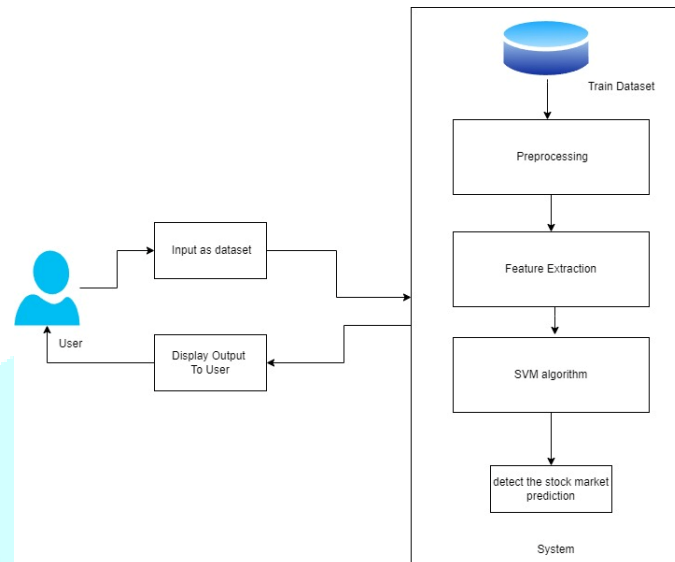
Segmentation:
This is made possible via segmentation, a technique for grouping customers according to traits or behaviours. Other marketing responsibilities including upselling strategies, product recommendations, and pricing can benefit from customer segmentation.
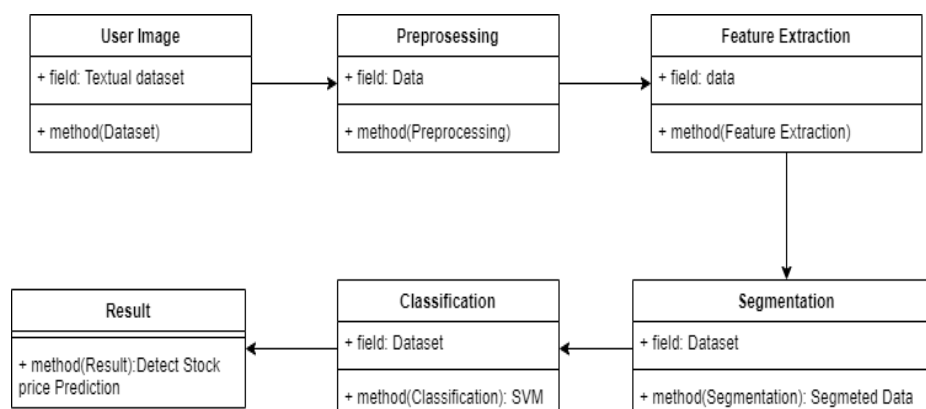
Classification:

Regression and classification algorithms can be used to classify the procedure for supervised machine learning. Regression approaches have been used to anticipate the outcomes for continuous values, but classification algorithms are needed to forecast the outcomes for categorical values. On the basis of training data, new observations are classified using the classification algorithm, a Supervised Learning technique. In classification, a programme determines how to categorise fresh observations into various classes or groupings based on the dataset or provided observations. Examples include: cat or dog, yes or no, 0 or 1, spam or not spam, etc. Targets, labels, or categories can all be used to indicate classes. In contrast to regression, classification creates an output variable that is a category rather than a value, such as "Green or Blue," "fruit or animal," etc. The Classification method requires labelled input data because it is a supervised learning technique with input and output data.

- **Convolution Neural Network (CNN)**

- Convolutional Neural Networks with a focus on image and video recognition applications. CNN is mostly utilised for image analysis applications like segmentation, object detection, and picture recognition.

- There are Four types of layers in Convolutional Neural Networks:

- **1) Convolutional Layer:** Each input neuron in a typical neural network is connected to the following hidden layer. Only a small portion of the input layer neurons in CNN are connected to the hidden layer of neurons.

- **2) Pooling Layer:** The feature map's dimensionality is decreased using the pooling layer. Inside the CNN's hidden layer, there will be numerous activation and pooling layers.

- **3) Flatten:** Data is flattened when it is made into a 1-dimensional array for input into the following layer. We flatten the convolutional layer output to produce a solitary, lengthy feature vector.

- **4) Fully-Connected layer:** The final several layers of the network are known as Fully Connected Layers. The output from the last pooling or convolutional layer is passed into the fully connected layer, where it is flattened before being applied.

Related Work:

In this project[1], This paper mainly studies the specific mechanism of investor senti- ment affecting stock market volatility. With the help of Pollet and Wilson's theory of volatility decomposition, it performs a comparative analysis based on big data strat- egy and sources.

In this project[2], This research work emphases on the prediction of future stock marketindex values based on historical data. The experimental evaluation is based on historicaldata of 10 years of two indices, namely, CNX Nifty and SP Bombay Stock Exchange (BSE) Sensex from Indian stock markets.

In this project[3] ,This literature review is aiming to explore the use Artificial Neural Network (ANN) techniques in the field of stock market prediction. Design: Content analysis research technique.

In this project[4], Stock market forecasting is very important in the planning of business ac- tivities. Stock price prediction has attracted many researchers in multiple disciplines including computer science, statistics, economics, finance, and operations research.

In this project[5], The categorization of high dimensional data present a fascinating challengeto machine learning models as frequent number of highly correlated dimensions or at- tributes can affect the accuracy of classification model. In this paper, the problem of high dimensionality of stock exchange is investigated to predict the market trends by applying the principal component analysis (PCA) with linear regression.

In this project[6], In this paper we present a novel data miming approach to predict long term be-havior of stock trend. Traditional techniques on stock trend prediction have shown theirlimitations when using time series algorithms or volatility modelling on price sequence.

In this project[7], Stock price is one of intricate non-linear dynamic system. Typically, Elman neural network is a local recurrent neural network, having one context layer that mem-orizes the past states, which is quite fit for resolving time series issues.

In this project[8], Given a financial time series such as SP 500, or any historical data in stock markets, how can we obtain useful information from recent transaction data to predict the ups and downs at the next moment? Recent work on this issue shows initial evidencethat machine learning techniques are capable of identifying (non-linear) dependency inthe stock market price sequences.

Dataset Information:

Dataset Name	S$P500 Stock Market
Dataset Source	YFinance (Online wedsite)
Number of Classes	11
Data Format	Text

Result-

The result of stock market prediction using deep learning models on stock market dataset were promising. In addition, various companies have developed AI-powered trading platforms that use machine learning algorithms to analyze market trends and make predictions. For example, Trade Ideas is a platform that uses machine learning to analyze market data and provide traders with real-time trading signals.

Overall, while machine learning algorithms have shown promise in predicting stock market trends with high accuracy and efficiency, it is important to note that the stock market is highly unpredictable and past performance does not guarantee future results.

Symbol	AAPL
Name	Apple Inc. Common stocck
Last Sale	$142.55
Net change	0.699%
Market capital	247143817000.0
Country	United States
IPO Year	1980.0
Volume	13006443
Sector	Technology
Industry	Computer Manufacturing

Overall, while machine learning algorithms have shown promise in predicting stock market trends with high accuracy and efficiency, it is important to note that the stock market is highly unpredictable and past performance does not guarantee future results.

Screenshots:

The image below shows the live price of stock market .

The above image show us the predicted price of the stock market.

Conclusion and Future Work:

Conclusion:
Here, we discovered that machine learning technology can be used to predict stock market movements. A graph of any company's stock price cannot be read in detail by a single individual. In the real world, we must analyse data at a massive scale with many different businesses. As a result, we may use machine learning techniques to make considerably better predictions. We may utilise the SVM algorithm to greatly increase the accuracy of our predictions.

Future Scope:
We are here covering multiple aspects in terms of getting better and updated results from our research based system. In future we can add more features. Multiple Country data can be tracked for international investments and multinational banking. Mutual Funds also can be tracked and can be given suggestion regarding mutual fund investments. Every activity , National Level effective Decisions and Decision makers can be tracked for early prediction.

References:

1)University of Bradford, 2017. H. Isah, "Social Data Mining for Crime Intelligence: Contributions to Social Data Quality Assessment and Prediction Methods."

2)University of Bradford, 2017. H. Isah, "Social Data Mining for Crime Intelligence: Contributions to Social Data Quality Assessment and Prediction Methods."

3)Expert Systems with Applications, vol. 83, pp. 187–205, 2017. E. Chong, C. Han, and F. C. Park, "Deep learning networks for stock market analysis and prediction: Methodology, data representations, and case studies."

4)A unique data-driven stock price trend prediction system was developed by J. Zhang, S. Cui, Y. Xu, Q. Li, and T. Li in 2018. Expert Systems with Applications, vol. 97, pp. 60–69.

5)Expert Systems with Applications, vol. 105, pp. 11-22, 2018. L. S. Malagrino, N. T. Roman, and A. M. Monteiro, "Forecasting stock market index daily direction: A Bayesia Network approach."

6)"Stock Price Prediction Using Artificial Neural Network," by M. B. Patel and S. R. Yalamalle, International Journal of Innovative Research in Science, Engineering, and Technology, vol. 3, June 2014, pp. 13755–13762.

7)Wen Fang, Jie Wang, and Jun Wang. Elman Recurrent Random Neural Networks for Financial Time Series Prediction[J]. Neuroscience and Computational Intelligence, 2016(12):1–14

8)Twitter is a wonderful site to spot health concerns, according to V. M. Prieto, S. Matos, M. Alvarez, F. Cacheda, and J. L. Oliveira. 2014, PLoS ONE, vol. 9, no. 1, article e86191.

9)"Overcoming data scarcity of Twitter: Using tweets as bootstrap with application to autism-related topic content analysis," A. Beykikhoshk, O. Arandjelovi, D. Phung, and S. Venkatesh, in Proc. IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining, Paris, France, August 2015, pp. 1354–1361.

10)Deep learning- and word embedding-based heterogeneous classifier ensembles for text categorization, Z. H. Kilimci and S. Akyokus, Oct. 2018, Complexity, vol. 2018, art. no. 7130146.