

Stock Market Analyzer

Gayathri K S¹, Megha H P², Sahana S Kotian³, Sashika V S⁴, Chethana E⁵

^{1,2,3,4} Dept of Information Science and Engineering

⁵ Asst. Professor., Dept of Information Science and Engineering

^{1,2,3,4,5} East West Institute of Technology, Bengaluru.

Abstract- In Stock Market Prediction, the aim is to predict the future value of the financial stocks of a company. The recent trend in stock market prediction technologies is the use of machine learning which makes predictions based on the values of current stock market indices by training on their previous values. Machine learning itself employs different models to make prediction easier and authentic. The paper focuses on the use of Regression and LSTM based Machine learning to predict stock values. Factors considered are open, close, low, high and volume. However, due to the high volatility and non-stationary nature of the stock market, forecasting the trend of a financial time series remains a big challenge. We introduced a new method to simplify noisy-filled financial temporal series via sequence reconstruction by leveraging motifs (frequent patterns), and then utilize a convolutional neural network to capture spatial structure of time series.

Keywords- trend prediction, machine learning, convolutional neural network, financial time series, motif extraction.

I. INTRODUCTION

Stock market plays a very important role in fast economic growth of the developing country like India. Developing nation's growth may depend on performance of stock market. If stock market rises, then countries economic growth would be high. If stock market falls, then countries economic growth would be down. In other words, we can say that stock market and country growth is tightly bounded with the performance of stock market. In any country, only 10% of the people engaging themselves with the stock market investment because of the dynamic nature of the stock market. There is a misconception about the stock market i.e. buying and selling of shares is an act of gambling. This misconception can be changed and bringing the awareness across the people. The attempt to solve the investor's problem of what to buy and when to buy led to the emergence of two distinct schools of thought regarding security valuation and stock price behavior in the early period of stock market study. They are popularly referred as the Fundamental Analysis and the Technical Analysis. The recent invention in computer science and accessibility of Internet leads to use latest methods in prediction of stock market. ARIMA model is one of the

standard models for the prediction of future direction of time series. RNN is the model used the techniques of computer science.

The Fundamental Analysts maintain that at any instant an individual security has an intrinsic value, which should be equal to the present value of the future stream of income from that security discounted at an appropriate risk-related rate of interest. The actual price of a security is considered to be a function of set of anticipated returns and anticipated capitalization rates. The real worth of a security is estimated by considering the key economic and financial variables such as earning, dividends, and growth in earning, capital structure, size of the company etc.

II. LITERATURE SURVEY

Stock Market are one of the important parts of the economy of a country. Actually, it's the most important way for the companies to raise capital. Not only the investors but also the common peoples are also finding it as an investment tool. As stock market influences individual and national economy heavily, predicting the future values of stock market is essential task while taking the correct decision whether to buy or sell the share. But it was very difficult to predict the stock price trends efficiently because many factors such as economics, politics, environment etc. were deciding parameters. The financial market is a complex, evolutionary, and non-linear dynamical system. The field of financial forecasting is characterized by data intensity, noise, non-stationary, unstructured nature, high degree of uncertainty, and hidden relationships. Many factors interact in finance including political events, general economic conditions, and traders' expectations. Therefore, predicting finance market price movements is quite difficult. Increasingly, according to academic investigations, movements in market prices are not random. Rather, they behave in a highly nonlinear, dynamic manner. The standard random walk assumption of futures prices may merely be a veil of randomness that shrouds a noisy non-linear process. Support vector machine (SVM) is a very specific type of learning algorithms characterized by the capacity control of the decision function, the use of the kernel functions and the sparsity of the solution [6]. Established on the unique theory of the structural risk minimization principle to

estimate a function by minimizing an upper bound of the generalization error, SVM is shown to be very resistant to the over-fitting problem, eventually achieving a high generalization performance.

III. EXISTING SYSTEM

Linear regression is widely used throughout Finance in a plethora of applications. Linear regression is a method used to model a relationship between a dependent variable (y), and an independent variable (x). With simple linear regression, there will only be one independent variable x. There can be many independent variables which would fall under the category of multiple linear regression. In this circumstance, we only have one independent variable which is the date. The date will be represented by an integer starting at 1 for the first date going up to the length of the vector of dates which can vary depending on the time series data. Our dependent variable, of course, will be the price of a stock. The accuracy of the prediction by Linear Regression is actually not high enough to make a good decision on stock trading. Linear Regression is limited to linear relationships. The algorithm already assumes the system is a straight-line. However, for stock trading, the values of the system could be either a raise, a drop or remain constant. The data values are scattered and fluctuated. Apart from that, Linear Regression is not a complete description of relationships among variable. It only provides the functionality to investigate on the mean of the dependent variable and the independent variable. However, it is not applicable for the situation we encountered in stock market. And hence, the prediction is actually suppressed.

IV. PROPOSED SYSTEM

Analysts making forecasts often have extensive domain knowledge about the quantity they are forecasting, but limited statistical knowledge. In the Prophet model specification, there are several places where analysts can alter the model to apply their expertise and external knowledge without requiring any understanding of the underlying statistics.

Capacities: Analysts may have external data for the total market size and can apply that knowledge directly by specifying capacities. Changepoints: Known dates of changepoints, such as dates of product changes, can be directly specified. Holidays and seasonality: Analysts that we work with have experience with which holidays impact growth in which regions, and they can directly input the relevant holiday dates and the applicable time scales of seasonality. Smoothing parameters: By adjusting τ an analyst

can select from within a range of more global or locally smooth models.

The seasonality and holiday smoothing parameters allow the analyst to tell the model how much of the historical seasonal variation is expected in the future. With good visualization tools, analysts can use these parameters to improve the model fit. When the model fit is plotted over historical data, it is quickly apparent if there were changepoints that were missed by the automatic changepoint selection.

Our analyst-in-the-loop modeling approach is an alternative approach that attempts to blend the advantages of statistical and judgmental forecasts by focusing analyst effort on improving the model when necessary rather than directly producing forecasts through some unstated procedure.

We find that our approach closely resembles the “transformvisualize model” loop proposed by Wickham & Goleman (2016), where the human domain knowledge is codified in an improved model after some iteration. Typical scaling of forecasting would rely on fully automated procedures, but judgmental forecasts have been shown to be highly accurate in many applications.

Our proposed approach lets analysts apply judgment to forecasts through a small set of intuitive model parameters and options, while retaining the ability to fall back on fully automated statistical forecasting when necessary. The direct result of requirements analysis is Requirements specification. Hardware requirements specifications list the necessary hardware for the proper functioning of the project. Software requirements specifications is a description of a software system to be developed, laying out functional and non-functional requirements, and may include a set of use cases that describe interactions the users will have the software. In software engineering, a functional requirement defines the function of a system and its components. A function is described as a set of inputs, the behavior, and outputs. A nonfunctional requirement that specifies the criteria that can be used to judge the operation of a system, rather than specific behavior.

The functional Requirements Specification documents the operation and activities that a system must be able to perform.

Functional requirements include:

- Descriptions of how data is collected and stored.

- Descriptions of data cleaning and pre-processing methods.
- Descriptions of work-flows performed by the system.
- Descriptions of outputs.

FLOW CHART :

It is important to complete all tasks and meet deadlines. There are many project management tools that are available to help project managers manage their tasks and schedule and one of them is the flowchart. A flowchart is one of the seven basic quality tools used in project management and it displays the actions that are necessary to meet the goals of a particular task in the most practical sequence. Also called as process maps, this type of tool displays a series of steps with branching possibilities that depict one or more inputs and transforms them to outputs.

This particular tool is very used in estimating and understanding the cost of quality for a particular process. This is done by using the branching logic of the workflow and estimating the expected monetary returns.

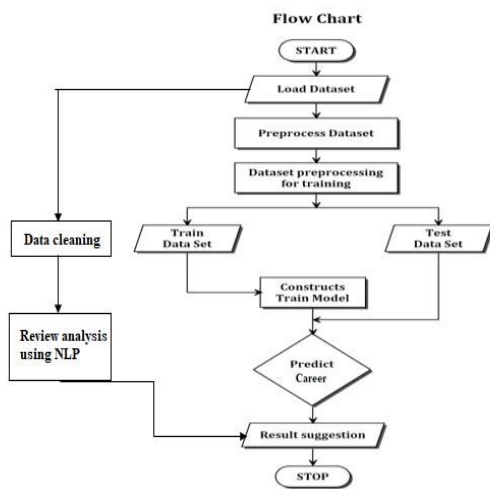


Fig 1. Flow Chart

V. ARCHITECTURE

The figure gives the overall system architecture of our project. It shows the working of our project where the data is extracted from the global dataset which undergoes the data processing. Then the filtered data is sent to the Prophet Model. Then the model predicts the test results.

The architecture of the proposed system:

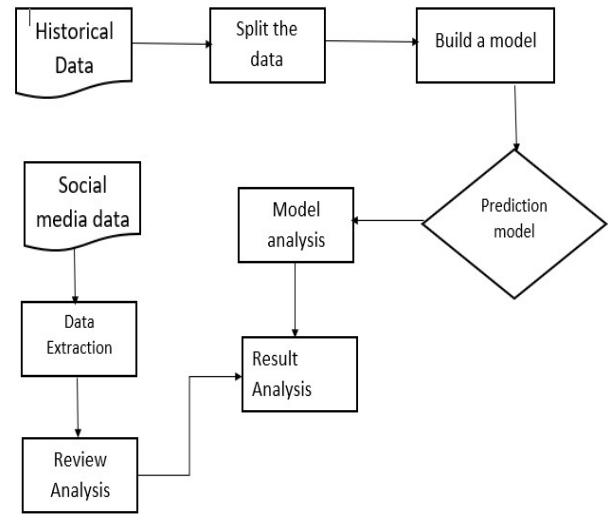


Fig 2. Architecture

VI. IMPLEMENTATION

A. Data collection

Data collection is the process of gathering and measuring information on targeted variables in an established system, which then enables one to answer relevant questions and evaluate outcomes. Data collection is a component of research in all fields of study including physical and social sciences, humanities, and business. While methods vary by discipline, the emphasis on ensuring accurate and honest collection remains the same. The goal for all data collection is to capture quality evidence that allows analysis to lead to the formulation of convincing and credible answers to the questions that have been posed.

Quandl's data products come in many forms and contain various objects, including time-series and tables. Through our APIs and various tools (R, Python, Excel, etc.), users can access/call the premium data to which they have subscribed. (Our free data can be accessed by anyone who has registered for an API key.)

B. Data preprocessing

It is a data mining technique that transforms raw data into an understandable format. Raw data (real world data) is always incomplete and that data cannot be sent through a model. That would cause certain errors. That is why we need to pre-process data before sending through a model.

Steps in Data Pre-processing

1. Import Libraries
2. Read Data
3. Checking for Missing Values
4. Checking for Categorical Variable
5. Standardized the data

C. NORMALIZATION

Normalization is a technique often applied as part of data preparation for machine learning. The goal of normalization is to change the values of numeric columns in the dataset to a common scale, without distorting differences in the ranges of values. For machine learning, every dataset does not require normalization. It is required only when features have different ranges.

D. Model fitting

Keras is an incredible library: it allows us to build state-of-the-art models in a few lines of understandable Python code. Although other neural network libraries may be faster or allow more flexibility, nothing can beat Keras for development time and ease of use. With the training and validation data prepared, the network built, and the embeddings loaded, model to learn how to write patent abstracts.

E. Testing/validation

Test Dataset: The sample of data used to provide an unbiased evaluation of a final model fit on the training dataset. The Test dataset provides the gold standard used to evaluate the model. It is only used once a model is completely trained (using the train and validation sets).

Validation Dataset: The sample of data used to provide an unbiased evaluation of a model fit on the training dataset while tuning model hyperparameters. The evaluation becomes more biased as skill on the validation dataset is incorporated into the model configuration.

F. Main packages

Some of the main packages used in this project are as mentioned below:

Prophet follows the sklearn model API. An instance of the Prophet class is created and then call its fit and predict methods.

PyTrends Unofficial API for Google Trends (fork). Allows simple interface for automating downloading of reports from

Google Trends. Main feature is to allow the script to login to Google on your behalf to enable a higher rate limit.

Pandas is a software library written for the Python programming language for data manipulation and analysis. In particular, it offers data structures and operations for manipulating numerical tables and time series.

NumPy is a package in Python used for Scientific Computing.

NumPy package is used to perform different operations.

From pyTrends. request import TrendReq It is used to connect to google.

Matplotlib. pyplot as plt Matplotlib is a Python 2D plotting library which produces publication quality figures in a variety of hardcopy formats and interactive environments across platforms.

Quandl The API can be used to deliver more complex datasets. This call gets the quarterly percentage change in AAPL stock between 1985 and 1997, closing prices only, in JSON format.

Forecasting techniques play an important role in stock market which can search uncover and hidden patterns and increasing the certain level of accuracy, where traditional and statistical methods are lacking. There is huge amount of data are generated by stock markets forced the researchers to apply forecasting to make investment decisions. The credit of originating the concept of investment value goes to John B. Williams [1938], who has also presented an actual formula for determining the intrinsic value of stocks. However, the concept of intrinsic value was popularized by B. Graham and D. Dodd [1934] in their classic book 'Security Analysis'. Many researchers have suggested further development in the theory of intrinsic value.

Fundamentalist forecast stock prices on the basis of economic condition of the industry and company statistics. In a major study covering the period 1927 - 1960. King [1966] has found that about 50% of the variance of the average stock returns is explained by the overall market factors. In another study over the period January 1966 to June 1970 Livingston [1977] observed that approximately 23% of the variance in stock returns was accounted by the market effect. Elton and Gruber [1987] also reported that about 25% to 50% of the variations in a company's annual earnings are due to the state of the overall economy.

Investors are familiar with the saying, "buy low, sell high" but this does not provide enough context to make proper

investment decisions. Before an investor invests in any stock, he needs to be aware how the stock market behaves. Investing in a good stock but at a bad time can have disastrous results, while investment in a mediocre stock at the right time can bear profits. Financial investors of today are facing this problem of trading as they do not properly understand as to which stocks to buy or which stocks to sell in order to get optimum profits. Predicting long term value of the stock is relatively easy than predicting on day-to-day basis as the stocks fluctuate rapidly every hour based on world events.

The direct result of requirements analysis is Requirements specification. Hardware requirements specifications list the necessary hardware for the proper functioning of the project. Software requirements specifications is a description of a software system to be developed, laying out functional and non-functional requirements, and may include a set of use cases that describe interactions the users will have the software. In software engineering, a functional requirement defines the function of a system and its components. A function is described as a set of inputs, the behavior, and outputs. A non-functional requirement that specifies the criteria that can be used to judge the operation of a system, rather than specific behavior.

VII. RESULTS

In the future, the stock market prediction system can be further improved by utilizing a much bigger dataset than the one being utilized currently. This would help to increase the accuracy of our prediction models. Furthermore, other models of Machine Learning could also be studied to check for the accuracy rate resulted by them.



Fig 3. Historic data, Train data and Predicted data



Fig 4. Accuracy vs Test data graph

VIII. CONCLUSION

The popularity of stock market trading is growing rapidly, which is encouraging researchers to find out new methods for the prediction using new techniques. The forecasting technique is not only helping the researchers but it also helps investors and any person dealing with the stock market. In order to help predict the stock indices, a forecasting model with good accuracy is required. In this work, we have used one of the most precise forecasting technologies using Recurrent Neural Network and Long Short-term Memory unit which helps investors, analysts or any person interested in investing in the stock market by providing them a good knowledge of the future situation of the stock market.

IX. ACKNOWLEDGEMENT

Any Achievement, be it scholastic or otherwise does not depend solely on the individual efforts but not on the guidance, encouragement and cooperation of intellectuals, elders and friends.

We would like to take this opportunity to thank them all. First and foremost we would like to thank Ms. Chethana E, Asst. Professor, EWIT, for her moral support and valuable suggestions and expert advice towards completing our project work and constantly guiding me to organize the project in a systematic manner.

We thank our parents, and all the Faculty members of Department of Information Science & Engineering for their constant support and encouragement. Last, but not the least, we would like to thank our parents and friends who provided us with valuable suggestions to improve our project.

REFERENCES

- [1] I. Svalina, V. Galzina, R. Luji, and G. Imunovi, "An adaptive network- based fuzzy inference system (ANFIS) for the forecasting: The case of close price indices," *Expert systems with applications*, vol. 40, no. 15, pp. 60556063, 2013.
- [2] M. A. Boyacioglu and D. Avci, "An adaptive network-based fuzzy inference system (ANFIS) for the prediction of stock market return: the case of the Istanbul stock exchange," *Expert Systems with Applications*, vol. 37, no. 12, pp. 79087912, 2010.
- [3] E. F. Fama and K. R. French, "Common risk factors in the returns on stocks and bonds," *Journal of financial economics*, vol. 33, no. 1, pp. 356, 1993.
- [4] T.-J. Hsieh, H.-F. Hsiao, and W.-C. Yeh, "Forecasting stock markets using wavelet transforms and recurrent neural networks: An integrated system based on artificial bee colony algorithm," *Applied soft computing*, vol. 11, no. 2, pp. 25102525, 2011.
- [5] Hall JW. Adaptive selection of US stocks with neural nets. In: Deboeck GJ, editor. *Trading on the edge: neural, genetic, and fuzzy systems for chaotic financial markets*. New York: Wiley; 1994. p. 45–65.
- [6] Tay FEH, Cao LJ. Application of support vector machines in financial time series forecasting. *Omega* 2001; 29:309–17.
- [7] Eugene F. Fama "The Behavior of Stock Market Prices", the *Journal of Business*, Vol 2, No. 2, pp. 7–26, January 1965.
- [8] Cao LJ, Tay FEH. Financial forecasting using support vector machines. *Neural Computing Applications* 2001; 10:184–92.
- [9] Zhen Hu, Jibe Zhu, and Ken Tse "Stocks Market Prediction Using Support Vector Machine", *International Conference on Information Management, Innovation Management and Industrial Engineering*, 2013.M.
- [10] Wei Huang, Yoshie Nakamori, Shou-Yang Wang, "Forecasting stock market movement direction with support vector machine", *Computers & Operations Research*, Volume 32, Issue 10, October 2005, Pages 2513–2522.
- [11] N. Ancona, *Classification Properties of Support Vector Machines for Regression*, Technical Report, RIIESI/CNR- Nr. 02/99.