
Ms. Kurhe Prajakta V  
PG Student  
Department of Computer Engineering.  
S.N.D.COE and RC, Yeola

Prof. Dhakane V. N.  
Asst. Professor  
Department of Computer Engineering.  
S.N.D.COE and RC, Yeola

ABSTRACT
The prediction of a stock market direction may serve as an early recommendation system for short term investors and as an early financial distress warning system for long term shareholders. Forecasting accuracy is the most important factor in selecting any forecasting methods. Research efforts in improving the accuracy of forecasting models are increasing since the last decade. The appropriate stock selections those are suitable for investment is a very difficult task. The key factor for each investor is to earn maximum profits on their investments. Stock is an unpredictable curve. Prediction in stock market is covered with the complexity and instability. The main aim for the persuasion of the topic is to predict the stability in the future market stocks. Many researchers have performed their research on the movement of future market evolution. Stock consists of fluctuating data which makes data as an integral source of efficiency. Impact on the same chances the efficiency of the prediction. In the recent trend of Stock Market Prediction Technologies machine learning has integrated itself in the picture for deployment and prediction of training sets and data models. Machine Learning employs different predictive models and algorithms to predict and automate things of requirement. Here while developing this system we are focusing on the use of additive regression model to predict stock values. In this paper, we investigate the predictability of financial movement with Facebook prophet (additive regression model). These methods are applied on multiple years of data retrieved from Yahoo Finance. The results will be used to analyze the stock prices and their prediction in depth in future research efforts. In this project we build a stock prediction web app in Python using streamlit, Yahoo finance, and Facebook Prophet.

Keywords
Additive regression model, Facebook Prophet, Machine Learning, streamlit, forecasting models, supervised learning.

1. INTRODUCTION
We all have heard the word stock one way or the other. Particularly stock is related with the associates and companies which are commercialized and are to settling in the world of marketization. The other word used for stock is share which is prominently used in day to day life. People even term it as an investment plan and it’s something people see as a long term investment that secures and provides an abundant funds during the retirement age. Buying a company stock is purchasing a small share of it. People invest in the same to get a long term benefit which they think is less value for now but has to potential to grow with the time. It’s an investment that provides the long time run and deals with long time goals with the fair objectives. The value of share you invest today has to give you and yield of best tomorrow but it’s not the same. Market is unpredictable so are the resources and the factors that are taken to drive it off or on the set. It’s never been on the same level and the pattern of the same is still unpredictable till the time. Some closeness and prediction method had been derived and approximates values and the rough figures are generated hoping for the best but all of the resource can’t be trusted and are still unpredictable in nature.

Knowing the market situation and researching on the same is the best way to find the reliability for which there are many agents who have taken the same as a profession and are making a fortune out of it. They predict and advise but the advisory cost and the charge is higher and the stock evaluation is never less the same. Market is changing in an instantaneous rate even in a day there are many highs and lows in the market and having said the resources and the timing the external and internal agent. Stock is a fascinating resource to start with.

2. REVIEW OF LITERATURE
Historic data are of great values and that been proved by Sathik and Sekhar [1]. They derived a hidden patterns from the dataset and have out generated a investment decision plan using different data mining technologies. They used the same output to invest on the stocks. The efficiency of the same was found to be 84.26% which was consider to be a higher hit rate.

ANN or Artificial neural networks was discovered later Liam and Jing [2]. They used the ANN techniques to classify, predict and recognize the data sets. In neural network the brain phenomenon is studied and the implementation of brain neurons are tried to be practiced. Output generated from the same were used in trading prediction and stability. In the research pages they have mentioned a seven prediction models in neural network for the higher efficiency yield. Sampling. Training and recommending are one of its features mentioned.

Neural Network was found well integrating with Linear Equation and its relation. Kun Huang and Tiffany [3] used the same to implement a time series fuzzy network model to improvise and predict the forecasting. The efficiency of the model was found to be deliberate but the computational time was higher than the expected causing it slow for prediction.

Bajkunthu and Md. Rafiul [4] approached interrelated market forecasting. The approach for the same was initiated with the help of HMM (Hidden Markov Models). Hmm is used for classification of the item set in bulk and can be even help in pattern matching. It’s a hybrid model implemented for efficiency in forecasting of stock market.

3. SYSTEM ARCHITECTURE
The dataset we use for the proposed project is been taken from Yahoo Finance. But, this data set is in raw format. The data set is a collection of valuation of stock market information about some companies. The initial step is to convert raw data into
processed data. Which is done by feature extraction, since the raw data collected have multiple attributes but only some of those attributes are needed for the prediction. Feature extraction is a reduction process.

The structure, behavior and views of a system is given by structural model.

Here as described in the figure above, the proposed system will have an input from the dataset which will be extracted featured wise and Classified underneath. The classification technique used is supervised and the various techniques of machine level algorithms are implemented on the same. Training Dataset are created for training the machine and the test cases are derived and implemented to carry out the visualization and the plotting’s. The result generated are passed and visualized in the graphical form.

3.1 Module Description

Data set
This is the fundamental module before starting of the project. The dataset is a group of data that are mended together to show the data variations in a time span to undergo further estimation and the source of the resources and its outcome for the later time of evaluation. It generates the result optimization and gives a feasible time period to customize and get the flow to the derivation. This increases and are used in the level of research and finding the best suitable resource out of the same the resources has to be finely estimated and derived for the best possible outcome and the finest the value become the better is the level of extraction and closure is the best yield values that needs to be considered.

Data abstraction
Abstraction is the finding of the resource to its best to categorize the above dataset and learning the best out of it. Abstraction of the data is the integral part to the flow. All the data are a huge set of chunks which on processing can limitize the yield result and the computational mean too. Thus with the available resources the data yield had to be derivative. Abstraction of the dataset is to customize the data set and finding the best suitable constraints to take into consideration and the unwanted resources are the dump which will be dumped and the supreme cluster is created with the valuable constrains and a pattern is needed to be derived from the same.

Training dataset
After the abstraction of the data and clustering of the same. The machine had to be trained for which the training data plays the important role. There are thousands of machine learning algorithms that are into place and evolving with the same. The best to the practice of machine learning is to yield the result and the content to derive what’s needed with the time frame. This is a supervised learning form where the input are passed so that the system learns from the same. Various variants of inputs are passed which were stored in the dataset. Every resource is considered and taken into consideration. After considering the whole set of information and the resource the machine tries to learn from the passed dataset. The dataset has to be wide and versatile. After considering the learning it tries to integrate with the same type and flow like the same as the human mind and creates a pattern and the links between the same.

Test dataset
These are the sets of data that gives the result after learning from the data. This is the test generation with the output result. Results are generated in each phase of testing. This is also termed as the testing phase. Now a new set of datasets are passed which are deliberately like the training dataset and the efficiency of the same is calculated.
Over-Fitting of the dataset. Validation of the same with the effective constraints and hyper parameters are checked. This phase is training and the output is evaluated with the set of training. After each process of computation the set of data are trained and efficiency of the same is measured and is evaluated with the others.

**Result Evaluation**
This is the main part for any implementation of the project. Evaluation of the key point to the success. All the categorization of the work and the best to know the resource fundamentals and again establishing the same to check the validity and the work flow and check on the output is must. The evaluation, utilization and implementation undergoes a various level of extraction and evaluation.

### 4. ALGORITHMIC STRATEGY

These are the Machine Learning Algorithms implemented during the building of the project.

**Linear regression**
One of the well-known algorithm used in machine learning is the linear regression. It is covered under both statistical as well as in machine learning. It is used for analyzing the dependency between two variables one is known dependency which value is known and the other is unknown. The value of the unknown dependency is checked with the known dependencies and the result is found ad derived on its basis. The dependency of the variable chances and are categorized into two types. Positive Linear Regression is the regression flow when both the dependencies shows the growth rate and both are totally depended and supportive with the changes flow. Negative Regression is the regression flow where one dependency cancels the growth of the other. If one dependency shows the tendency to grow whereas the other one is decreasing then this graph flow comes in picture.

**Polynomial regression**
It is a form of Non-Linear Regression. In this form of regression the two constraints one having known dependencies whereas the other part is unknown and is generalized with the help of nth polynomial value. Research is a wide level of scope that one’s involved. There are various curves and lines estimation which can’t be normally fixed and plotted with the limitation of linear regression if trying to do so there will be the higher error ratio which will bring down the integrity and the reliability of the system in itself.

Thus to overcome this barrier and to represent the most of the curve in every way possible either it be a straight line, or hysteresis curves. This regression helps to analyze the curve in every possible format and helps to reduce the redundancy and the trial points of errors alongside with optimized cost factor which is a great boost to the algorithm itself.

Its widely used for the complexity to solve and takes the particular values which are unique in nature and peculiar values that needs to be considered before to set the outcome. The natural uses of the same is found in epidemics growth and to see the growth ratio of the tissue.

**Long short term memory (lstm)**
Sequence prediction problems have been around for a long time. They are considered as one of the hardest problems to solve in the data science industry. These include a wide range of problems; from predicting sales to finding patterns in stock markets’ data, from understanding movie plots to recognizing your way of speech, from language translations to predicting your next word on your iPhone’s keyboard. With the recent breakthroughs that have been happening in data science, it is found that for almost all of these sequence prediction problems, long short Term Memory networks, LSTMs have been observed as the most effective solution.

### 5. Implementation

#### 5.1 Methodology & Programming Platform

**Python: programming platform**
To implement this “Stock Market Prediction Using Machine Learning” project we are using python programming platform. Python is an interpreter, object-oriented, high-level programming language with dynamic semantics. Its high-level built in data structures, combined with dynamic typing and dynamic binding, make it very attractive for Rapid Application Development, as well as for use as a scripting or glue language to connect existing components together. Python’s simple, easy to learn syntax emphasizes readability and therefore reduces the cost of program maintenance. Python supports modules and packages, which encourages program modularity and code reuse. The Python interpreter and the extensive standard library are available in source or binary form without charge for all major platforms, and can be freely distributed.

**Streamlit**
We are developing “Stock Market Prediction Using Machine Learning” as web application and to deploy this web application we are using streamlit python library. Streamlit is an open-source python framework for building web apps for Machine Learning and Data Science. We can instantly develop web apps and deploy them easily using Streamlit. Streamlit allows you to write an app the same way you write a python code. Streamlit makes it seamless to work on the interactive loop of coding and viewing results in the web app.

**Yahoo finance**
We use yahoo finance to fetch historical data of various stocks and index for training and prediction purpose. Yahoo Finance provides free stock quotes, up-to-date news, portfolio management resources, international market data, social interaction and mortgage rates that help you manage your financial life.

**Facebook prophet**
Just like Support Vector Machines (SVM) and Artificial Neural Networks (ANN) are widely used for prediction of stock prices and its movements, Facebook prophet is also one of the best. Prophet is indeed machine learning. On its own it is an algorithm that generates no value. Combine it with data and you have something that can predict the future. Prophet is a procedure for forecasting time series data based on an additive model where non-linear trends are fit with yearly, weekly, and daily seasonality, plus holiday effects. It works best with time series that have strong seasonal effects and several seasons of historical data.
6. RESULT & DISCUSSION
Analysis of stocks using Machine Learning will be useful for new investors to invest in stock market based on the various factors considered by the software.

Stock market includes daily activities like Sensex and Nifty calculation, exchange of shares. The exchange provides an efficient and transparent market for trading in equity, debt instruments and derivatives.

Our aim is to create software that analyzes previous stock data of certain companies from Indian share market as well as International share market like S&P 100 etc., with help of certain parameters that affect stock value. We are going to implement these values in Using Facebook prophet Machine learning methodology.

This will also help us to determine the values that particular stock will have in near future. We will determine the Month’s High and Low with help of Machine learning algorithms.

6.1 Output & Screenshot

Step: 01 (In this step we have to set Python Virtual Environment and Run Web App Using Streamlit Tool)
- What Is a Virtual Environment? At its core, the main purpose of Python virtual environments is to create an isolated environment for Python projects.

- This means that each project can have its own dependencies, regardless of what dependencies every other project has.

Step: 02 (In this step we can see visualization of project on Localhost by using Streamlit tool.)
- Streamlit is an all-in-one tool that encompasses web serving as well as data analysis. Use Streamlit if you want a structured data dashboard with many of the components you’ll need already included.

- Here in this step, we have to choose company stock for which we want to do prediction. After choosing stock its historical data will be get fetched.

- Next we can use Range slider to, using which we can set for how many years we want to do prediction.

- Raw data get load and its graph is also get displayed.

- Here we have selected GOOGLE company stock for 03 years of future prediction.

Step: 03 (In this step we can see Stock Prediction or Forecasting for next 03 years using Facebook Prophet Analysis Model of Machine Learning.)
- Prophet is used in many applications across Facebook for producing reliable forecasts for planning and goal setting.

- Here we are using Prophet Methodology, using which training of historical data will be done and analysis will result into prediction.

- At its core, the Prophet procedure is an additive regression model with four main components: A piecewise linear or logistic growth curve trend.

- Prophet automatically detects changes in trends by selecting change points from the data. A yearly seasonal component modeled using Fourier series.

- Also forecasting Plot will be get displayed.

Step: 03 (In this step we can see visualization of project on Localhost by using Streamlit tool.)
- Streamlit is an all-in-one tool that encompasses web serving as well as data analysis. Use Streamlit if you want a structured data dashboard with many of the components you’ll need already included.

- Here in this step, we have to choose company stock for which we want to do prediction. After choosing stock its historical data will be get fetched.

- Next we can use Range slider to, using which we can set for how many years we want to do prediction.

- Raw data get load and its graph is also get displayed.

- Here we have selected GOOGLE company stock for 03 years of future prediction.

Step: 03 (In this step we can see Stock Prediction or Forecasting for next 03 years using Facebook Prophet Analysis Model of Machine Learning.)
- Prophet is used in many applications across Facebook for producing reliable forecasts for planning and goal setting.

- Here we are using Prophet Methodology, using which training of historical data will be done and analysis will result into prediction.

- At its core, the Prophet procedure is an additive regression model with four main components: A piecewise linear or logistic growth curve trend.

- Prophet automatically detects changes in trends by selecting change points from the data. A yearly seasonal component modeled using Fourier series.

- Also forecasting Plot will be get displayed.

Fig No 3: Login Page
Step: 04 (In this step we can see Stock Prediction or Forecasting components.)

*Here forecasting components are graphically displayed as a weekly and yearly trends for selects stocks.*

7. APPLICATIONS

The application will predict the stock prices for the next trading sessions. The requirements and the functionality of this application correlates it to the class.

CONCLUSION

To conclude stock is an unpredictable mechanism which follows the segments of chain and the dependencies of the same are unpredictable. It is defined to be a curve which keeps on changing and turning the price from low to high and vice-versa.

As the integration of the same is higher with other dependencies so leaving one dependencies compromises the level of accuracy. Accuracy is not the term used over in stock as the actual prediction is not possible for any fiscal days it keeps on changing and turning the tables day and night. Having higher component assets and the dependencies makes it more flexible and useful in nature causing it even harder to predict. The approx. value are taken into consideration and the hit or profit of the gain rate is calculated for the same.

In the project various high level machine learning algorithms are implemented and integrated and the output is generated from the same making a user visible with the outputs in the form of graph which makes it easier for them to see and interpret what’s the scenario and they can decide on the same to invest and get the benefit out of it.

The proposed software takes the raw set of data from the dataset or the .csv file and process it. The cleaning and cleansing of data is done and then further processed to gain the effective outcomes. After the computational mean the output is displayed in the screen in the form of graph. The project can be further continued to gain the effectiveness of the prediction with addition implementations of the content that can involve real time scenario and the way of executing and processing the real time scenario. Various constrains has to be added and performance of the same can be acylated in the future time for the effective results. The expected form of the display is graph whereas from the same the more appearance and setting of the display can be integrated and a pie-chart and a custom graph can further me implemented on the same. Obstacle sensor by which accident can be avoided.

REFERENCES


on Artificial Intelligence (ICAI). The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing (WorldCom), 2012.