

Stock Market Prediction Using Sentiment Analysis and Machine Learning

Manjusha Kolhe¹, Shubhangi More², Snehal Pimpale³, Pratibha Sanvatsarkar⁴

Prof. Rohit Nikam⁵

^{1,2,3,4}Students, Sanjivani COE, Kopargaon, Tal: Kopargaon, Dist: Ahmednagar.

⁵Assistant Professor at Sanjivani COE kopargaon, Dept. of Information Technology, Sanjivani COE, kopargaon, Tal-Kopargaon, Dist-Ahmednagar, State-Maharashtra, India.

Abstract - Investor sentiment plays an important role on the stock market. User-generated textual content on the Internet provides a precious source to reflect investor psychology and predicts stock prices as a complement to stock market data. This paper integrates sentiment analysis into a machine learning method based on support vector machine. Stock Market Prediction is the challenging but the important task in the area of Finance and Economics. Stock Market System is non-linear and dynamic in nature. Financial news has large impact on stock price prediction and investor sentiments constitute a key factor of the financial market. This paper includes Sentiment Analysis into a Machine Learning method based on Support Vector Machine.

Hence an automated approach is required to get the knowledge from large text document. Sentiment Analysis is used to automatically extract the views, emotions from the opinionated contents. All this information along with the support vector machine will be taken as feature to train the model. This model will predict the near future price of the stock.

Key Words: Support Vector Machine, Sentiment analysis.

1. INTRODUCTION

Stock Market prediction and analysis is process of determining future value of the company stock and other financial instruments on trading exchange. Stock market is the important part of economy of the country and plays important role in the growth of industry and commerce of the country that affect the economy of the

country. Both Investors and industry are involved in the stock market and wants to know stock will rise or fall. It is based on the demand and supply strategy as demand increases price increases. The demand and supply law affect the stock market. Typically low availability and high demand boost the price of an item. And high availability and low demand reduces its price. Stock market prediction is very important from business point of view. Stock Market Prediction is the act of trying to determine the future value of a company stock. The successful prediction of stock price could yield significant profit. It is important from company and investors point of view. Stock market movement can be predicted by various factors such as economic growth of the company, Management, related market, price to earnings ratios. Beneficial news related to any company can attract more investors to invest in that company share which may lead to increase in demand.

Hence it is very necessary for the investors to read news, reviews of the company on internet and social media. This can be automated using sentiment analysis which is one of the leading indicators in stock market. Using sentiment analysis we can mine textual contents from the internet and take proper decision from it. Our system will contain two major modules: Sentiment analysis and SVM. For sentiment analysis we will take twitter's data in the form of tweets related to stock. SVM will take historical stock price data from Yahoo Finance websites. By combining results of both the modules system will predict approximate stock price.

2. LITERATURE SURVEY

Previous work done in the area of stock market prediction is as follows:

Rui Ren, Desheng Dash Wu, Tianxiang Liu[1], This paper aims to forecast stock market movement direction by not only using financial market data, but also combining them with sentiment features that incorporate investor psychology. The features are extracted from unstructured news data automatically and then are expressed as sentiment indexes. Moreover, the prediction results are used to instruct investment decisions, and the performance of three different trading strategies are evaluated and compared.

Kalyani Joshi, Bharati H.N., Jyoti Rao[2], Stock Trend Prediction Using News Sentiment Analysis. This project is about taking non quantifiable data such as financial news articles about a company and predicting its future stock trend with news sentiment classification. Assuming that news articles have impact on stock market, this is an attempt to study relationship between news and stock trend. To show this, they have created three different classification models which depict polarity of news articles being positive or negative.

Dev Shah, Haruna Isah, Farhana Zulkernine[3], Due to the unavailability of dictionaries specific to the financial markets, they have manually created a dictionary that contains words and phrases which carried specific meaning pertaining to pharmaceutical companies. The dictionary was created by leveraging author's domain expertise and thorough analysis of news articles over the years. The dictionary consisted of 100 words which cover a broad spectrum of topics specific to the pharmaceutical sector. Each entry in the dictionary was tagged with positive, negative, or neutral class of sentiment. The generated n-grams were then compared against the dictionary, and if a match was found then each matching word was assigned a positive or negative polarity. The positive and negative words were then counted and the

sentiment scores were generated based on their frequency.

Aditya Bhardwaja, Yogendra Narayanb[4], in this paper sentiment analysis for stock market is demonstrated by fetching Sensex and Nifty live server data values on different interval of Stock Market Prediction Using Sentiment Analysis and Machine Learning time that can be used for predicting the stock market status. For this purpose they have used python scripting language which have fast execution environment and this will help out the investors in order to make a prediction of on what shares money should be invested, it will also help in maintaining the economical balance of share market.

3. MOTIVATION

Stock market prediction is very important form business point of view. Financial experts make prediction about the stock market by using various technical indicators, news about the company, company performance, investor sentiment, demand and supply strategy, etc. It is a hectic work for a person and which can be made more easier by using computer software. The Stock market process is full of uncertainty and it's affected by many factors such as company news and performance, industry performance, investor sentiment, economic factors etc. that can cause the price of a stock to rise or fall.

The common problem faced by the investors include high market volatility, loss of money, stock market crash, poor investment skills and lack of market knowledge. These reasons may lead to wrong decisions. If investor makes wrong decision while selling and buying of the shares then they may face loss. Hence, before investing money, it is very important for investors to predict the stock market. Hence the Stock market prediction is one of the important exertions in business and finance. Accuracy is the most important factor in predicting the stock prices that too high degree of accuracy.

By identifying issues related to stock market prediction we have proposed this system which will take investor’s sentiments in the form various tweets from twitter related to any NSE stock and historical prices of the same stock.

4. SYSTEM ARCHITECTURE

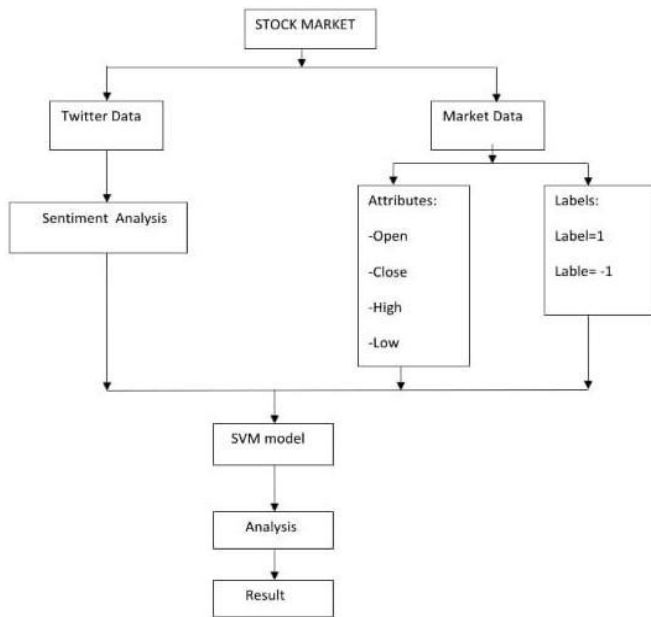


Fig.4.1.Proposed System architecture

5. SVM MODEL

Support Vector Machine (SVM) :

SVM offers very high accuracy compared to other classifiers such as logistic regression, and decision trees. It is known for its kernel trick to handle nonlinear input spaces. The classifier separates data points using a hyperplane with the largest amount of margin. SVM finds an optimal hyperplane which helps in classifying new data points. Support vector machines can be employed for classification and regression problems. In our system we are using regression approach.

The SVM algorithm is implemented in practice using a kernel. A kernel transforms an input data space into the required form. SVM uses a technique called the kernel trick. Here, the kernel takes a low-dimensional input space and transforms it into a higher dimensional space. In other words, you can say that it converts

nonseparable problem to separable problems by adding more dimension to it. It is most useful in non-linear separation problem. Kernel trick helps you to build a more accurate classifier. There are various types of kernels among which we have used linear kernel method

- Linear Kernel Linear Kernel is used when the data is Linearly separable, that is, it can be separated using a single Line. It is one of the most common kernels to be used. It is mostly used when there are a Large number of Features in a particular Data Set..A linear kernel can be used as normal dot product any two given observations. The product between two vectors is the sum of the multiplication of each pair of input values.

- $K(x,xi) = \sum(x*xi)$

Steps for implementation of SVM in Python using scikit-learn

1.Loading Data :first load the required dataset you will use.

2. Exploring Data After you have loaded the dataset, you might want to know a little bit more about it. You can check feature and target names.

3.Splitting Data To understand model performance, dividing the dataset into a training set and a test set is a good strategy.

4.Generating Model build support vector machine model.fit your model on train setand perform prediction on the textst set

5.Evaluating the Model estimation of how accurately the classifier or model can predict the stock price of particular stock.Accuracy can be computed by comparing actual test set values and predicted values.

Sentiment Analysis:

Sentiment analysis is the interpretation and classification of emotions (positive, negative and neutral) within text data using text analysis techniques. Sentiment analysis models detect polarity within a text (e.g. a positive or negative opinion), whether it's a whole document, paragraph, sentence, or clause. Understanding people's opinion regarding any particular stock is essential for investors before investing in that stock. By automatically analyzing people's views, from social media conversations, investors can take proper decision about selling and buying of shares.

In our system we are taking tweeter tweets to extract positive and negative reviews of people. Tweets are downloaded from tweeter and stored in csv file.

Preprocessing is done on that csv file which is again processed to assign polarity to each tweet. Frequency of positive and negative tweets is calculated means how many positive tweets and how many negative tweets are calculated. Data from csv file is divided into training data and testing data to perform SVM classification. Linear kernel is used for SVM. effectively classifying the text into positive and negative polarities by using various preprocessing methods like eliminating stop words and tokenization which increases the performance of sentiment analysis in terms of accuracy and time taken by the classifier. The features obtained after applying feature extraction techniques on the text sentences are trained and tested using the support vector machines.

Step 1: Crawl Tweets Against Hash Tags: To have access to the Twitter API, you'll need to login the Twitter Developer website and create an application. Enter your desired Application Name, Description and your website address making sure to enter the full address including the http://. You can leave the callback URL empty. This Twitter Crawler allows you to scrape tweets against hash tags and store the tweets into a csv.

Step 2. Analyzing Tweets for Sentiment: This module can classify particular stock related tweets into positive and negative. In this module, text data is preprocessed

and stop words are filtered out before applying the support vector machine algorithm.

Step 2.1: Preprocess Tweets

Before we start building the analyzer, we first need to remove noise and preprocess tweets by using the following steps:

1. Lower Case - Convert the tweets to lower case.
2. URLs - Eliminate all of these URLs via regular expression matching or replace with generic word URL.
3. @username - Remove "@username" via regex matching or replace it with generic word AT_USER.
4. #hashtag - replace hashtags with the exact same word without the hash (hash tags may provide some useful information).
5. Punctuations and additional white spaces - remove punctuation at the start and ending of the tweets

Step 2.2: Extract Feature Vectors: One important step in building a classifier is deciding what features of the input are relevant, and how to encode those features. we can use the presence or absence of words that appear in tweet as features. In the training data, we can split each tweet into words and add each word to the feature vector. Some of the words might not indicate the sentiment of a tweet and we can filter them out. Then merge individual feature vector into a large list that contains all the features and remove duplicates in this list.

Step 2.3: Training the Classifiers

Support Vector Machines: The Support Vector Machine (SVM) algorithm tries to find a hyperplane which separates the data in two classes as optimally as possible. **Step 3: Visualizing the Results:** A bar chart is generated having two bars showing frequency of positive and negative tweets.

6. CONCLUSIONS

Predicting Stock Market is very difficult because of its non-linear, dynamic nature. Predicting the direction of

movements of the stock market index is important and of great interest because successful prediction may promise attractive benefits. It usually affects a financial trader's decision to buy or sell an instrument. This study attempted to predict the direction of stock price movement in the National Stock Exchange.

A prediction model has been built that uses social media data such as Twitter and machine learning to periodically predict the trend about stock markets. Model shows that sentiment analysis of the social data complements proven to the technical analysis methods such as support Vector Machine. However the algorithm used for sentiment analysis uses summative assessment of the sentiments in a particular tweets, this could be improved for better sentiment calculations, which would improves for the accuracy of the prediction.

6. REFERENCES

- [1] Rui Ren, Desheng Dash, Tianxiang Liu, "Forecasting Stock Market Movement Direction Using Sentiment Analysis and Support Vector Machine" IEEE 2018
- [2] Kalyani Joshi¹, Prof. Bharathi H. N.², Prof. Jyothi Rao³, "Stock Trend Prediction Using Sentiment Analysis" IJCSIT 2016
- [3] Dev Shah, Haruna Isah, Farhana Zulkernine, "Predicting the Effects of News Sentiments on the Stock Market" IEEE
- [4] Aditya Bhardwaja, Yogendra Narayanb, Vanrajc, Pawana, Maitreyee Duttaa "Sentiment Analysis for Indian Stock Market Prediction Using Sensex and Nifty" ScienceDirect
- [5] Alenxder Porsher, Ilya Redkin, Alexey Shevchenko, "Machine Learning in Prediction of Stock Market Indicators Based on Historical Data" IEEE 2013