IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Stock Market Prediction based on Social Sentiments using Machine Learning

[1]Ms. Rushali Pansare , [2]Ms. Swapnali Lahane , [3]Ms. Nikita Shinde

[4]Ms. Sharda Shinde, [5]Prof. Pratibha Kashid

[1]*Ms. Rushali Pansare, Dept Of Information Technology,SIR VISVESVARAYA INSTITUTE OF TECHNOLOGY, Nashik, Maharastra, India*

[2]*Ms. Swapnali Lahane, , Dept Of Information Technology,SIR VISVESVARAYA INSTITUTE OF TECHNOLOGY, Nashik, Maharastra, India*

[3]Ms.Nikita Shinde, *Dept Of Information Technology,SIR VISVESVARAYA INSTITUTE OF TECHNOLOGY, Nashik, Maharastra, India*

[4]Ms.Sharda Shinde, *Dept Of Information Technology,SIR VISVESVARAYA INSTITUTE OF TECHNOLOGY, Nashik, Maharastra, India*

[5]Prof. Pratibha Kashid *Dept Of Information Technology,SIR VISVESVARAYA INSTITUTE OF TECHNOLOGY, Nashik, Maharastra, India*

Abstract: Machine learning and artificial intelligence techniques are being used to solve many real world problems. These techniques are highly effective, minimal effort and saving huge amount of time. Now people are invested in stock market or share market for yielding huge amount of money. Stock market is association of buyers and sellers. But in stock market, any time the stock value will grow or down according to the economic trends. So these changes could affect your share value and some times that decreases your profit. So stock market prediction is very necessary for avoiding this loss. We will propose a system that predicts the stock market value based on social sentiments using machine learning. We will collect the tweets from twitter API and Social media data to perform sentiment analysis and at same time collect data from Social media . Then find the correlation of historical data and extracted twitter data. This relational value used to determine the predicted outcome. This prediction system could greatly help stock investors in taking desired decision which could affect the profit of stock.

Key Words: machine learning, sentiment analysis, twitter API, Social Media Data, stock market prediction.

1. INTRODUCTION

Stock Market Prediction System (SMPS)is a practical system that forecasts the stock price movement of various companies. Such a prediction could greatly help stock investor in taking desired decision which would directly contribute to his profits. Nowadays, social media has become a mirror that reflects people's thoughts and opinions to any particular event or news. Any positive or negative sentiment of public related to a particular company can have effect on its stock prices. Our system predicts the stock market prices of various companies by performing sentiment analysis of the social media data such as tweets related to the respective companies. We will collect the tweets from twitter API and Social media App and perform sentiment analysis of it. Corresponding to that time period, we shall analyze the stock values from past data and use a suitable machine learning algorithm to justify a valid correlation between the tweet sentiment and the stock values. Finally, with training data, we will train our model and develop capability to produce stock predictions for future. Since the public reaction to any event is available.

Literature Survey

- **Paper 1: Improving Long Term Stock Market Prediction with Text Analysis.**
  In this thesis, we looked at the problem of forecasting stock performance. Although a substantial volume of research exists on the topic, very little is aimed at long term forecasting while making use of machine learning methods and textual data sources. We prepared over ten years worth of stock data and proposed a solution which combines features from textual yearly and quarterly filings with fundamental factors for long term stock performance forecasting. Additionally, we developed a new method of extracting features from text for the purpose of performance forecasting and applied feature selection aided by a novel evaluation function.

  Paper 2: Empirical analysis: stock market prediction via extreme learning machine.
  In this paper, we present the design and architecture of our trading signal mining platform that employs extreme learning machine (ELM) to make stock price prediction based on those two data sources concurrently. Comprehensive experimental comparisons between ELM and the state-of the-art learning algorithms, including support vector machine (SVM) and back-propagation neural network (BPNN), have been undertaken on the intra-day tick-by-tick data of the H-share market and contemporaneous news archives.

  Paper 3: Feeling The Stock Market: A Study in the Prediction of Financial Markets Based on News Sentiment.
  This study proposed a sentiment analysis system developed to infer the polarity of news articles related to a company with the purpose of predicting the stock market.

2. MACHINE LEARNING

Machine learning is a method of data analysis that automates analytical model building. It is a branch of artificial intelligence based on the idea that systems can learn from data, identify patterns and make decisions with minimal human intervention. Because of new computing technologies, machine learning today is not like machine learning of the past. It was born from pattern recognition and the theory that computers can learn without being programmed to perform specific tasks; researchers interested in artificial intelligence wanted to see if computers could learn from data. The iterative aspect of machine learning is important because as models are exposed to new data, they are able to independently adapt. They learn from previous computations to produce reliable, repeatable decisions and results. It's a science that's not new – but one that has gained fresh momentum.

While many machine learning algorithms have been around for a long time, the ability to automatically apply complex mathematical calculations to big data – over and over, faster and faster – is a recent development. Resurging interest in machine learning is due to the same factors that have made data mining and Bayesian analysis more popular than ever. Things like growing volumes and varieties of available data, computational processing that is cheaper and more powerful, and affordable data storage.

All of these things mean it's possible to quickly and automatically produce models that can analyze bigger, more complex data and deliver faster, more accurate results – even on a very large scale. And by building precise models, an organization has a better chance of identifying profitable opportunities – or avoiding unknown risks. It is no doubt that the sub-field of machine learning / artificial intelligence has increasingly gained more popularity in the past couple of years. As Big Data is the hottest trend in the tech industry at the moment, machine learning is incredibly powerful to make predictions or calculated suggestions based on large amounts of data. Some of the most common examples of machine learning are Netflix's algorithms to make movie suggestions based on movies you have watched in the past or Amazon's algorithms that recommend books based on books you have bought before.

3. MACHINE LEARNING ALGORITHMS

1. Decision Trees: A decision tree is a decision support tool that uses a tree-like graph or model of decisions and their possible consequences, including chance-event outcomes, resource costs, and utility. Take a look at the image to get a sense of how it looks like. From a business decision point of view, a decision tree is the minimum number of yes/no questions that one has to ask, to assess the probability of making a correct decision, most of the time. As a method, it allows you to approach the problem in a structured and systematic way to arrive at a logical conclusion.
2. Naive Bayes Classification: Naive Bayes classifiers are a family of simple probabilistic classifiers based on applying Bayes' theorem with strong (naive) independence assumptions between the features. The featured image is the equation — with P(A|B) is posterior probability, P(B|A) is likelihood, P(A) is class prior probability, and P(B) is predictor prior probability.
   Some of real world examples are:
   - To mark an email as spam or not spam
   - Classify a news article about technology, politics, or sports
   - Check a piece of text expressing positive emotions, or negative emotions?
   - Used for face recognition software.
3. Ordinary Least Squares Regression: If you know statistics, you probably have heard of linear regression before. Least squares is a method for performing linear regression. You can think of linear regression as the task of fitting a straight line through a set of points. There are multiple possible strategies to do this, and "ordinary least squares" strategy go like this — You can draw a line, and then for each of the data points, measure the vertical distance between the point and the line, and add these up; the fitted line would be the one where this sum of distances is as small as possible.
4. Logistic Regression: Logistic regression is a powerful statistical way of modeling a binomial outcome with one or more explanatory variables. It measures the relationship between the categorical dependent variable and one or more independent variables by estimating probabilities using a logistic function, which is the cumulative logistic distribution. In general, regressions can be used in real-world applications such as:

a. Credit Scoring
b. Measuring the success rates of marketing campaigns
c. Predicting the revenues of a certain product
d. Is there going to be an earthquake on a particular day.

5. Support Vector Machines: SVM is binary classification algorithm. Given a set of points of 2 types in N dimensional place, SVM generates a (N—1) dimensional hyperplane to separate those points into 2 groups. Say you have some points of 2 types in a paper which are linearly separable. SVM will find a straight line which separates those points into 2 types and situated as far as possible from all those points. In terms of scale, some of the biggest problems that have been solved using SVMs (with suitably modified implementations) are display advertising, human splice site recognition, image-based gender detection, large-scale image classification.

6. Clustering Algorithms: - Clustering, like regression, describes the class of problem and the class of methods. Clustering methods are typically organized by the modeling approaches such as centroid-based and hierarchal. All methods are concerned with using the inherent structures in the data to best organize the data into groups of maximum commonalities.
   a. The most popular clustering algorithms are:
   b. k-Means
   c. k-Medians
   d. Expectation Maximization (EM)
   e. Hierarchical Clustering
   f.
7. Artificial Neural Network Algorithms: Artificial Neural Networks are models that are inspired by the structure and/or function of biological neural networks. They are a class of pattern matching that are commonly used for regression and classification problems but are really an enormous subfield comprised of hundreds of algorithms and variations for all manner of problem types. Note that I have separated out Deep Learning from neural networks because of the massive growth and popularity in the field. Here we are concerned with the more classical methods. The most popular artificial neural network algorithms are:
   a. Perceptron
   b. Multilayer perceptions (MLP)
   c. Back-Propagation
   d. Stochastic Gradient Descent
   e. Hopfield Network
   f. Radial Basis Function Network (RBFN)

4. PROPOSED SYSTEM

1.This project is quite relevant as it guides people who possess limited know-how of investments and finance into making well informed decisions regarding stock market investments. It by passes the need for hiring investment experts who command exorbitant wages to guide our financial decisions by providing a simple solution which can be accessed by anyone having a computer or a laptop and an internet connection.

2.Stock market trends for a given time frame can be analysed easily even by the uninformed. Popularizing this machine learning option provides cheap alternative to various stock market investment guidance agencies which are in vogue today. The project puts in a small effort to assist the inexperienced investors and prevent from suffering heavy capital loss.

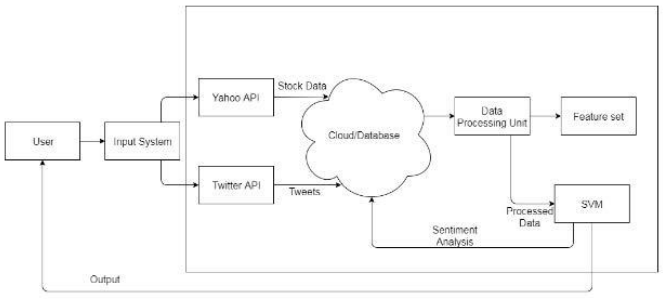

Fig. System Architecture

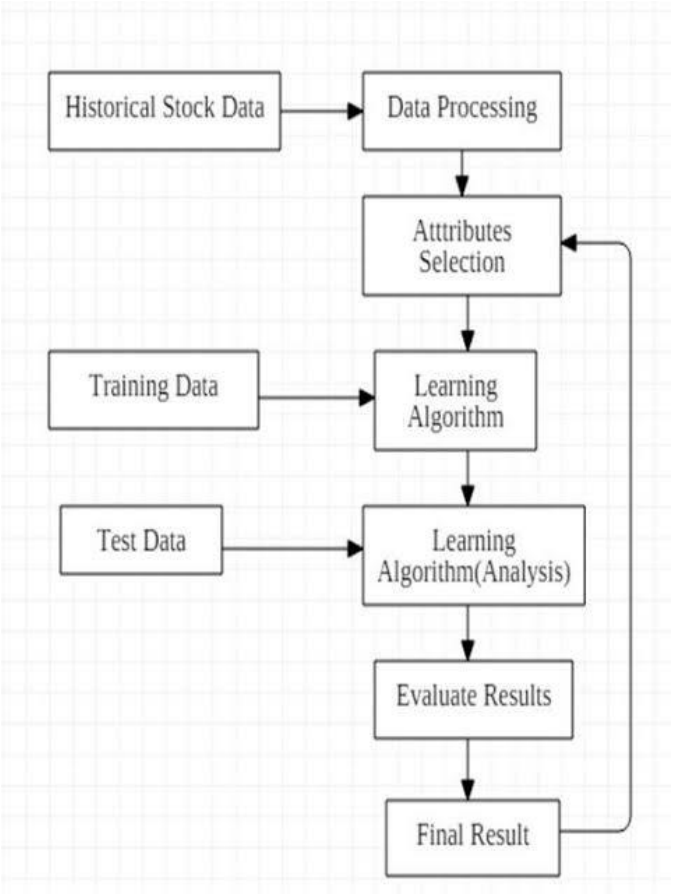

Fig. System model

5. Conclusion

Using Machine Learning technique and sentiment analysis for prediction purposes is inexpensive compared to other models. Support Vector Machine provide to be the most efficient and feasible model in predicting the stock price value. Cloud services will enable us to collect large amount of data and also store it in real time when we will get the data directly from the social media API. Collection of tweets and classification of tweets as positive, negative and neutral gives a good overview of public mood. The proposed system.

6. References

[1] Bohn, Tanner A. "Improving Long Term Stock Market Prediction with Text Analysis." (2017).

[2] Li, Xiaodong, et al. "Empirical analysis: stock market prediction via extreme learning machine." *Neural Computing and Applications* 27.1 (2016): 67-78.

[3] Sorto, Max, Cheryl Aasheim, and Hayden Wimmer. "Feeling The Stock Market: A Study in the Prediction of Financial Markets Based on News Sentiment." (2017).

[4] Alostad H,Davulcu H(2015) Directonal prediction of stock prices using breaking news on Twitter,In:IEEE/WIC/ACM international conferences on WI-IAT 1,pp 523-530.