International Journal of Engineering Science and Advanced Technology (IJESAT)
Vol 24 Issue 05, MAY, 2024

FORECASTING STOCK MARKET MOVEMENT DIRECTION USING SENTIMENT ANALYSIS AND SUPPORT VECTOR MACHINE

Mohd Aftab[1], Md. Ateeq Ur Rahman[2], Sridhar Gummalla[3]

1 Research Scholar, Department of Computer Science and Technology, Shadan College of Engineering and Technology, Hyderabad, Telangana, India – 500086.

2 Professor, Department of Computer Science and Engineering, Shadan College of Engineering and Technology, Hyderabad, Telangana, India – 500086.

3 Professor, Department of Computer Science and Engineering, Shadan College of Engineering and Technology, Hyderabad, Telangana, India – 500086.

ABSTRACT

On the stock market, investor sentiment is a significant factor. In addition to stock market data, user-generated textual content on the Internet is a valuable resource for reflecting investor psychology and making stock price predictions. This work incorporates sentiment analysis into a support vector machine-based machine learning technique. Additionally, we account for the day-of-week effect and create sentiment indices that are more accurate and realistic. The findings of empirical research demonstrate that 89.93% accuracy in predicting the direction of movement of the SSE 50 Index can be achieved, with an increase of 18.6% when sentiment factors are included. In the interim, our methodology assists investors in making more informed choices. These results also suggest that mood may be one of the leading predictors of the stock market and likely holds valuable information about the basic values of assets.

INTRODUCTION

The stock market is a dynamic and intricate system that is impacted by a wide range of variables, such as investor behavior, political developments, and economic indicators. Quantitative data, including past prices, trading volumes, and financial ratios, has been the mainstay of traditional stock price forecasting techniques. However, a wealth of user-generated content has been available due to the widespread use of social media and the internet, providing fresh perspectives on investor mood. The sentiments and viewpoints of investors are reflected in this textual content, which includes posts on social media, blogs, forums, and financial news stories. It can be a useful addition to traditional market data. Textual data can have sentiment extracted and quantified thanks to sentiment analysis, a subfield of natural language processing (NLP). Sentiment scores that reflect good, negative, or neutral investor sentiment can be obtained by examining the emotional tone and viewpoints presented in this content. The fluctuations of stock prices can therefore be predicted by adding these sentiment ratings as extra variables in predictive models. In order to improve the precision of stock market projections, we combine sentiment analysis with a machine learning technique in this project—more precisely, a support vector machine (SVM). Strong in high-dimensional domains and resilient to overfitting, the SVM is a potent classification technique. Our goal is to generate a comprehensive feature set for our predictive model by fusing sentiment indexes produced from textual data with technical indicators collected from historical stock prices. Furthermore, we take into account the day-of-week effect, a well-researched phenomenon in financial markets wherein regular changes in stock returns are observed according to the day of the week. Our goal is to build more accurate and realistic forecasting models by including these variations into our sentiment indexes. The SSE 50 Index, a significant stock market index in China, is the subject of our empirical investigation. We show that the addition of sentiment variables dramatically improves the model's predicting performance; at 89.93% accuracy rate, the accuracy rate is 18.6% higher than in models without sentiment data. These results imply that mood can act as a leading indicator of changes in the stock market and provides important information about the underlying values of assets. Our model is a powerful tool for investors who want to make more strategic and well-informed decisions by giving them a more detailed view of market dynamics and increasing prediction accuracy. This study demonstrates the value of integrating sentiment analysis into financial forecasting and the possibilities that arise from fusing contemporary data analytics methods with conventional quantitative data.

OBJECTIVE

The aim of this study is to obtain an effective and convincing sentiment index for stock market forecasting, taking into account day-to-day variations and closing data. Mondays were usually when the returns dropped off. The effect's existence is then demonstrated in international stock markets. One of the main causes is presumably the fact that information is generated considerably more on weekends than it is during the week.

PROBLEM STATEMENT

An important component that affects stock prices in the stock market is investor sentiment, among many other things. Customary models for predicting the movements of the stock market often overlook the psychological aspects influencing market movements, instead depending mostly on past price data and technical indicators. The accuracy of stock market predictions may be improved by using user-generated textual content from the Internet, such as news stories and social media posts, which offer insightful information about investor sentiment. In order to forecast changes in the stock market, it is still difficult to combine sentiment research with machine learning methods in an efficient manner. In addition, prediction models frequently ignore the day-of-week impact, which implies that stock market returns can change based on the day of the week.

EXISTING SYSTEM

- ANNs, or artificial neural networks, are computer programs that mimic the way the human brain processes information by drawing inspiration from biology.
- ANNs identify patterns and relationships in data to acquire knowledge; they are not programmed; rather, they learn (or are trained) by experience.
- Artificial neural networks, also known as processing elements (PE) or hundreds of individual units, are assembled into layers and coupled by weights, or coefficients, to form the neural structure of the network.

Disadvantage of Existing System

- Nature of a black box
- Heavy computational load

PROPOSED SYSTEM

- A separating hyperplane serves as the formal definition of a discriminative classifier, or support vector machine, or SVM.
- To put it another way, the algorithm generates an ideal hyperplane that classifies fresh cases based on labeled training data (supervised learning).
- This hyperplane is a line that splits a plane into two sections in two dimensions, with each class lying on each side.

Advantages of Proposed System

- Data that is semi-structured
- Unstructured is excellent when we don't know anything about it.

RELATED WORKS

Prior research has indicated that sentiment analysis plays a major role in forecasting changes in the stock market. Based on sentiment data, several machine learning algorithms have been used to predict stock prices. Because of its efficiency in solving classification difficulties, Support Vector Machine (SVM) is an often used option. Many academics have studied the day-of-week effect, which shows that stock returns fluctuate based on the day of the week.

METHODLOGY OF PROJECT

In order to predict the direction of movement of the SSE 50 Index, this study combines sentiment analysis with a support vector machine (SVM) model. Sentiment scores are obtained and corrected for the day-of-week effect by gathering and preprocessing user-generated textual content and historical stock data. The SVM model's feature set is comprised of these sentiment indices and technical stock indicators.

MODULES DESCSRIPTION:

1) A web scraper

Our objective is to construct a web crawler that can autonomously retrieve specific text documents from the Internet and save them in a database for subsequent analysis. With the seeds, which are a list of URLs, the web crawler gets started. Choosing the order and removing duplicate entries from the queue of URLs is done by the scheduler. The downloader is then in charge of gathering the web pages from the Internet and giving them to the spider so that it may parse the pages and extract the desired contents.

2) Getting ready

Preparing data for data mining involves taking certain actions to improve its suitability. Typically, there are two types of steps involved in data preprocessing:

• Choosing the characteristics and data items for the analysis.

• Adding or modifying the qualities.

3) Everyday attitude

The textual data is processed over a predetermined time period using a sentence-based sentiment analysis technique. Since a sentence can represent a relatively comprehensive meaning and help address the issue of ambiguity, we view a sentence as a unit to interpret the meaning of the entire document rather than a single word. Consequently, a document is organized into sentences first. The phrases are then broken up into individual words, the words are projected onto the sentiment space, and the number of positive and negative words is counted.

4) A changed attitude

Among the most well-known anomalies in finance is the day-of-week effect, which indicates that the average return on a Monday is much less than that on other days

of the week. A major portion of the news is reported over the weekend or on Friday just after the market closes, which is one of the reasons. On Mondays, investors are highly likely to be more flexible and take action, given the substantial and valuable information at their disposal.

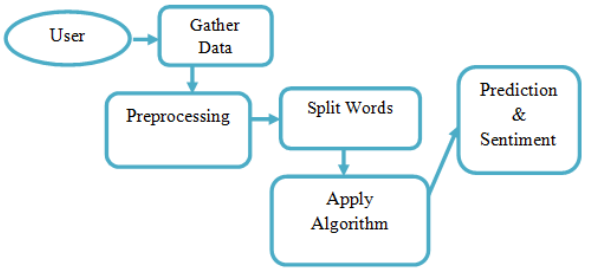
5) Forecasting

When you try to predict the likelihood of a specific result using new data, the output of an algorithm that has been trained on a previous dataset is referred to as a prediction. For each record in the new data, the algorithm will produce likely values for an unknown variable, enabling the model builder to determine what that value will most likely be. Here, we are employing the SVM algorithm to forecast the information related to the stock market.

ALGORITHM USED IN PROJECT

- Support vector machine
- A separating hyperplane serves as the formal definition of a discriminative classifier, or support vector machine (SVM).Stated differently, the method produces an optimum hyperplane that classifies fresh cases based on labeled training data (supervised learning).
- This hyperplane is a line that splits a plane into two sections in two dimensions, with each class lying on each side.

DATA FLOW DIAGRAM



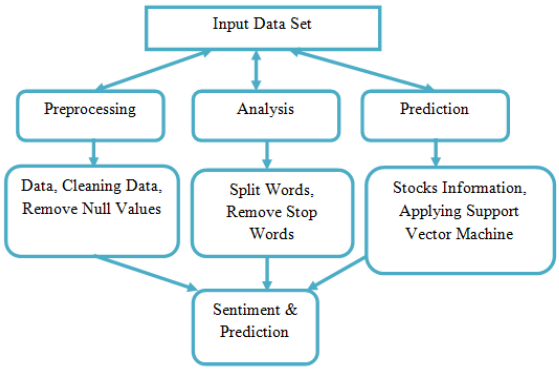


Fig: 7 Flow Diagrams of Modules

SYSTEM ARCHITECTURE

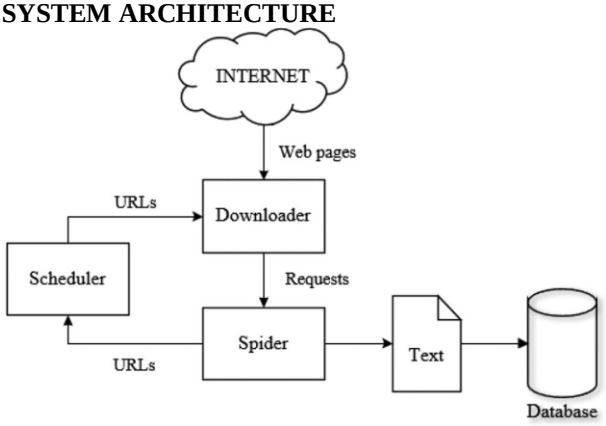
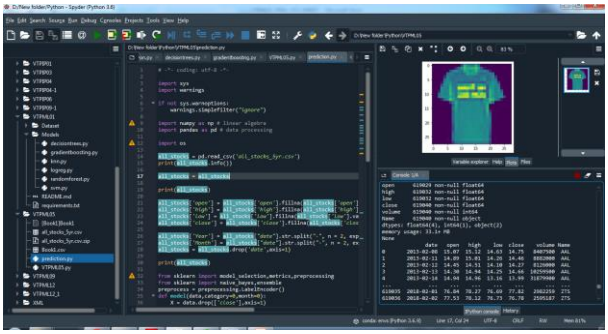
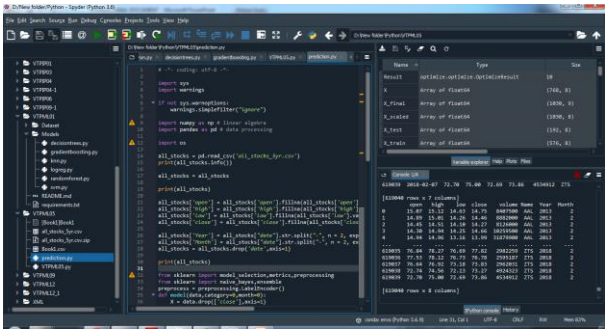


Fig: 8 SYSTEM ARCHITECTURE OF PROJECT

RESULTS





FUTURE ENHANCEMENT

Future work will likely involve crawling more textual content from the Internet in addition to extending the time interval. Improving the method's efficiency is also essential for processing large amounts of data quickly.

CONCLUSION

Our goal is to use investor sentiment to predict the direction of stock market movement by highlighting the importance of investors. The psychology of investors influences the stock market, which is relevant to our study. Consequently, online user-generated material offers a priceless resource for reflecting

investor psychology. Sentiment analysis is the process that creates daily sentiment indexes from unstructured text data. In addition, the financial anomaly day-of-week effect—which implies that Mondays have a significantly lower average return than the other days of the week—likely affects the sentiment indexes' accuracy. Accordingly, we modify the indexes by adding an exponential function based on historical weekend sentiment changes before extrapolating the results to holidays. Consequently, two common financial websites—Sina Finance and East money—were chosen as test platforms to gather data from financial reviews into a corpus. Next, using a realistic rolling window technique and fivefold cross validation, the machine learning model SVM is used to predict the SSE 50 Index, a very significant index in China.

REFERENCES:

[1] R. J. Shiller, Irrational Exuberance. Princeton, NJ, USA: Princeton Univ. Press, 2000.
[2] I. Perikos and I. Hatzilygeroudis, "Recognizing emotions in text using ensemble of classifiers," Eng. Appl. Artif. Intell., vol. 51, pp. 191–201, 2016.
[3] B. Wu, X. Zhou, Q. Jin, F. Lin, and H. Leung, "Analyzing social roles based on a hierarchical model and data mining for collective decision-making support," IEEE Syst. J., vol. 11, no. 1, pp. 356–365, Mar. 2017.
[4] B. Liu and L. Zhang, "A survey of opinion mining and sentiment analysis," Mining Text Data. New York, NY, USA: Springer, 2012.
[5] R. J. Shiller, "From efficient markets theory to behavioral finance," J. Econ. Perspectives, vol. 17, no. 1, pp. 83–104, 2003.
[6] F. C. Kelly, Why You Win or Lose: The Psychology of Speculation. North Chelmsford, Massachusetts, USA: Courier Corp., 2003.
[7] E. D. Maberly, "Eureka! Eureka! Discovery of the monday effect belongs to the ancient scribes," Financial Anal. J., vol. 51, pp. 10–11, 1995.
[8] J. Zhang, Y. Lai, and J. Lin, "The day-of-the-week effects of stock markets in different countries," Finance Res. Lett., vol. 20, pp. 47–62, 2017.
[9] W. Huang, Y. Nakamori, and S.-Y. Wang, "Forecasting stock market movement direction with support vector machine," Comput. Oper. Res., vol. 32, no. 10, pp. 2513–2522, 2005.
[10] L. Yu, H. Chen, S. Wang, and K. K. Lai, "Evolving least squares support vector machines for stock market trend mining," IEEE Trans. Evol. Comput., vol. 13, no. 1, pp. 87–102, 2009.