Stock Market Prediction and Risk Analysis using NLP and Machine Learning

Nishant Verma, S G David Raj, Ackley J Lyimo, Kakelli Anil Kumar

Abstract: The stock market has been an instrument of investment for more than 200 years now. The price movements in the stock market have been an enigma for many financial analysts and they have tackled this problem with very little success. The phenomenal advancement in technology led to increased storage systems, higher processing speed and better algorithms. Thus, it is more possible now to develop a system for predicting stock markets. People have this taboo that only big investors can profit from the stock market and it is a trap for retail investors or small players. The solution we propose here is easy to understand and implement. We first do the sentiment analysis of the selected stock and then suggest whether to buy, sell or hold. Secondly, we calculate the maximum risk involved in the investment using a threefold approach; market risk, sector risk and stock risk. Finally, using Support vector Regression (SVR) with three different approaches, we calculate expected return and compare them with actual returns.

Index Terms: Machine Learning, Risk Analysis, Stock Market, Stock Prediction, Support Vector Regression (SVR).

I. INTRODUCTION

The investments in the stock market has become very risky nowdays because of increased volatility and lack of patience among youngsters. The investing is one of the most difficult decisions which may result in huge profit or loss according to the investors' analysis. It is very crucial that the extent of human errors in these pressure situations is reduced so that the profit can be maximized. The technical analysts believe that the future price can be forcasted using the past price movements. They think the past is repeated again and again after certain cyces. The research people have used historical datasets to forecast the price movements of the stock market. The results are shocking and it proves that historical data has strong predictive sense. The analysts say that it is because of long term support and resistance or demand and suply zones in the charts.[1] We should try to understand that the security value doesn't primarily depend upon the historical values but also on the company's present projects and public sentiment. The sentiment analysis decreases the risk factor by informing the investors about the intricacies of the decision they are about to make.[2] Machine learning is a branch of computer science which has become important in almost every field or domain nowdays.

Nishant Verma, PG Scholar, School of Computer Science Engineering Specialization in Information Security, Vellore Institute of Engineering and Technology, Vellore, Tamil Nadu, India.

S G David Raj, PG Scholar, School of Computer Science Engineering Specialization in Information Security, Vellore Institute of Engineering and Technology, Vellore, Tamil Nadu, India.

Ackley Lyimo, PG Scholar, School of Computer Science Engineering Specialization in Information Security, Vellore Institute of Engineering and Technology, Vellore, Tamil Nadu, India.

Dr. Kakelli Anil Kumar, Associate professor, School of Computer Science Engineering specialization in Network Security, Vellore Institute of Engineering and Technology, Vellore, Tamil Nadu, India.

The stock closing prices for some future date could be predicted by training the machine learning models by providing the stock prices for previous dates. Sentiment analysis means going through a given text and calculating the sentiment of the text in terms of score whether it's positive or negative in the given context. When sentiment analysis is applied on stocks in news from moneycontrol.com regarding the public sentiment or opinion on that stock. Then, it becomes evident that whether to invest in that stock or not. The proposed system includes all the previously mentioned modules in an efficient and compact manner. The usage of sentiment analysis and machine learning in a novel and ingenious way makes the system robust and scalable for practical implementation.

II. PROCEDURE FOR PAPER SUBMISSION

A. Related Work

Exhaustive research has been done in the field of stock market analysis in the past but the required levels of accuracy has never been attained. Machine learning has been applied not only on the previous closing stock prices but also on various other features such as volume of stocks, standard deviations etc. to enhance the efficiency in predicting the values. There have been attempts to include sentiment analysis as well to improve the viability of the readings. A lot of work has also been done in the field of making these predictors in a portable version for ensuring ease of access. Neural networks and sentiment analysis libraries are used in collaboration to obtain more accurate readings. Very little research has been done to assign a risk to the investment as major focus is given to try and predict the precise value of the stock rather than asssociating risk based on trend and public sentiment. The system proposed in this paper assigns a risk value on the investment along with various other functionalities on a portable platform.

B. Motivation

The existing systems only focuses on huge financial organizations and not on the common people [3]. The common man who is not well equipped with the knowledge of the stock market deserves a solution which not only informs the investor regarding the historic and predicted trend of the company along with the public opinion but also suggests them the investments by placing a maximum risk percentage on each investment.[4] There is a lack of compact and simple platform to view information regarding the stock market. The system proposed in this paper provides all of these functionalities at zero cost and is available easily.

Any individual can make an informed decision with the help of this product which will be ready for consumer use and easy to understand.[5]
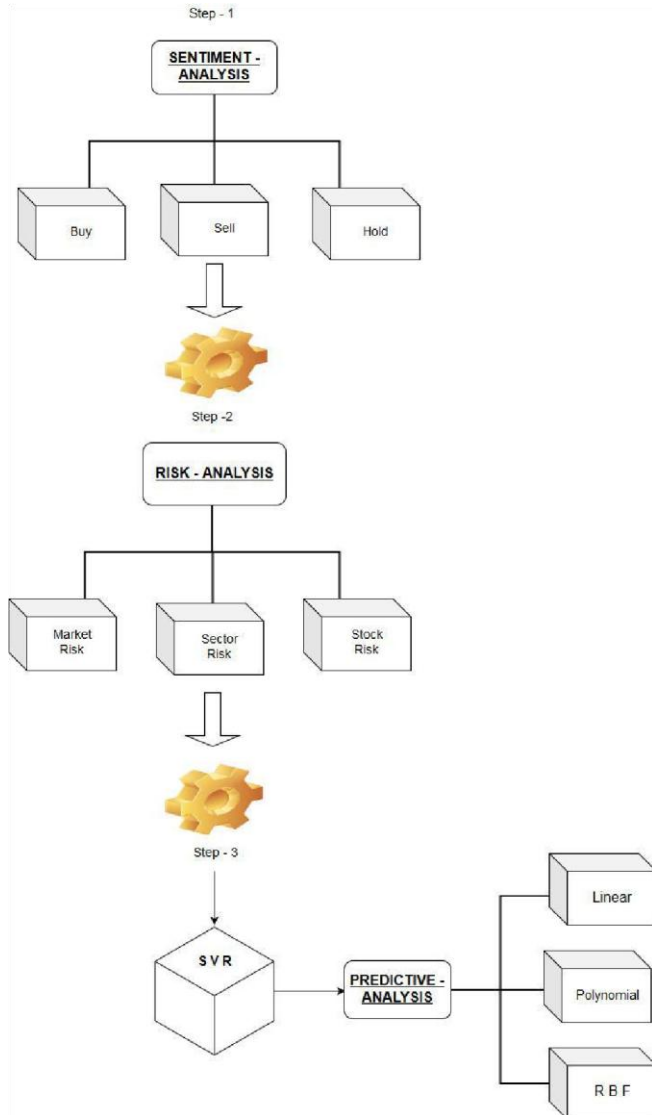
C. Figures

Fig 1: Implementation flow diagram.

III. IMPLEMENTATIONS

The implementation can be divided into three modules as they are sequential in nature as shown in figure 1.

A. Sentimental analysis using NLP

We do the sentiment analysis of the selected stock and then suggest whether to buy, sell or hold.[6] It is done using the inbuilt package of python nltk for natural language processing.[9]

Codes:

```
#print the final list of sentimental Analysis
positive = 0
negative = 0
for item in df:
        positive = positive + item [1]
        negative = negative + item [2]

sentiment = positive – negative
```

```
if sentiment > 0:
        print ("BUY THE STOCK")
elif sentiment < 0:
    print ("SELL THE STOCK IF YOU HAVE")
else:
    print ("HOLD THE STOCK IF YOU HAVE")
```

B. Risk Analysis

We calculate the maximum risk involved in the investment using a threefold approach; market risk, sector risk and stock risk.[6] Each risk is calculated using the volatility of the selected security calculated from the historical data.

Codes:

```
# investment period
Days = 10
#market volatility and risk
quotes = market ['Close'][-days:]
log_returns = np.log (quotes / quotes. Shift (1))
alpha = np.sqrt (days*log_returns.var ())
market_volatility = alpha*100
market_risk_per = market_volatitlity * market_corr
# sector volatility and risk
quotes = market ['Close'][-days:]
log_returns = np.log (quotes / quotes. Shift (1))
beta = np.sqrt (days*log_returns.var ())
volatility = beta*100
sector_risk_per = sector_volatitlity * sector_corr
# stock volatility and risk
quotes = market ['Close'][-days:]
log_returns = np.log (quotes / quotes. Shift (1))
gamma     =     np.sqrt     (days*log_returns.var     ())
stock_volatility = gamma*100
stock_risk_per = market_volatitlity * market_corr
max_risk_per    =    market_risk    +    sector_risk_per+
stock_risk_per
print ("MAXIMUM RISK: ", max_risk_per)
```
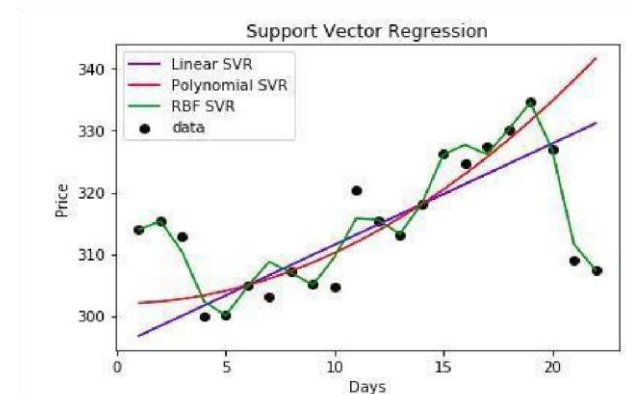
C. Predictive Analysis

Fig 2. SVR Graphical Representation

The third module corresponds to predicting the stock market closing price for future dates. This was done with the help of a machine learning model.

Using SVR with three different approaches, we calculate expected return and compare them with actual returns.

```
stock = get_history(symbol='SBIN',start=date(2015,2,2),end=date(2015,3,1))
start = stock['Close'][0]
end = stock['Close'][10]

returnval = (start - end)/start*100
print("Actual Return: ",returnval)

Actual Return:  0.260162601626

return_lin = (start - pred_lin)/start*100
print("Linear SVR Return: ",return_lin)

return_poly = (start - pred_poly)/start*100
print("Polynomial SVR Return: ",return_poly)

return_rbf = (start - pred_rbf)/start*100
print("RBF SVR Return: ",return_rbf)

Linear SVR Return:  -13.0162601626
Polynomial SVR Return:  -25.4772358219
RBF SVR Return:  -3.46401493926
```

Fig 3: SVR investment return analysis

IV. RESULTS

We can be confidently say that the proposed system differs from the existing system in a sense that, unlike other systems, it does not try to predict the accurate stock price of the company but rather assigns a risk percentage corresponding to each company based on sentiment analysis and risk analysis.[8] The SVR model has been used to predict the future price. The RBF SVR gives the best result compared to linear and polynomial approach as shown in Figure. 3.

V. CONCLUSION

The current implementation is restricted in terms of speed and efficiency due to constraints on computation power but to handle concurrent users and massive requests to the server, better hardware or processors can be used. The system can be made such that it can learn over time the investing habits of the users such as investing schedule, preferences and willingness to take risk and then provide suggestions which are more suitable to their needs [7]. The interface can be developed in different languages so that people from all regions can fully utilize the benefits of the web app. Localization of the processing can be done by performing machine learning and calculation on the local device itself by using the TensorFlow software. This will enable the user to enjoy the benefits of the app without internet connection as well.

ACKNOWLEDGMENT

REFERENCES

1. S. Asur, & B. A. Huberman, "Predicting the future with social media," International Conference on Web Intelligence and Intelligent Agent Technology (WI- IAT),2010.
2. A. W. Lo, & A. C. MacKinlay, "Stock market prices do not follow random walks: Evidence from a simple specification test," Review of financial studies, vol. 1, no. 1, pp. 41-66,1988.
3. S. Shen, H. Jiang, & T. Zhang, "Stock market forecasting using machine learning algorithms,2012.
4. Efthymios Kouloumpis, Theresa Wilson, "Twitter Sentiment Analysis: The Good the Bad and the OMG!," Proceedings of the fifth International AAAI Conference on Weblogs and Social Media.
5. Chen, Tao, & Bahsoon, Rami. (2013). "Self-adaptive and sensitivity-aware QoS modeling for the cloud". Paper presented at the Proceedings of the 8th International Symposium on Software Engineering for Adaptive and Self-Managing Systems, San Francisco, CA,USA.
6. A. Pak and P. Paroubek. "Twitter as a Corpus for Sentiment Analysis and Opinion Mining". In Proceedings of the Seventh Conference on International Language Resources and Evaluation, 2010, pp.1320–1326.
7. W. Huang Research paper: Forecasting stock market movement direction with support vector machine. Journal: Computers & Operations Research.
8. Yauheniya Shynkevich, T.M. McGinnity, Sonya Coleman, Ammar Belatreche, Predicting Stock Price Movements Based on Different Categories of News Articles, 2015 IEEE Symposium Series on Computational Intelligence.
9. Anurag Nagar, Michael Hahsler, Using Text and Data Mining Techniques to extract Stock Market Sentiment from Live News Streams, IPCSIT vol. XX(2012) IACSIT Press, Singapore.

AUTHORS PROFILE

Nishant Verma, PG Scholar, School of Computer Science Engineering Specialization in Information Security, Vellore Institute of Engineering and Technology, vellore-632014, TamilNadu, INDIA.

S G David Raj, PG Scholar, School of Computer Science Engineering Specialization in Information Security, Vellore Institute of Engineering and Technology, vellore-632014, TamilNadu, INDIA.

Ackley Lyimo, PG Scholar, School of Computer Science Engineering Specialization in Information Security, Vellore Institute of Engineering and Technology, vellore-632014, TamilNadu, INDIA.

Dr. Kakelli Anil Kumar, Associate professor, School of Computer Science Engineering specialization in Network Security, Vellore Institute of Engineering and Technology, vellore-632014, TamilNadu, INDIA.