Investment Portfolio Risk Manager using Machine Learning and Deep-Learning.

Vettithanam Alex Sebastian[1], Freshin Francis[2], Abin Varghese Jacob[3], Gaikar Prateek Balu[4], Prof. Dinesh S Tharwani[5]

[1,2,3,4]UG Student, Dept. of Computer Engineering, Pillai College of Engineering, New Panvel, Maharashtra, India
[5] Professor, Dept. of Computer Engineering, Pillai College of Engineering, New Panvel, Maharashtra, India

---***---

Abstract - *Portfolio Management is the concept of selecting the proportions of various assets that are to be held in a portfolio to have a good return without significant risk exposure. In financial management and investment banking, portfolio optimization is a critical component. Constructing an optimal portfolio by selecting the best possible combinations of different portfolios is a computationally challenging problem since it comes up with an exponential complexity. It has been widely accepted that public sentiment is correlated with financial markets. Different types of machine learning models have recently been applied for short-term financial market prediction with positive outcomes. However, it is observed that historical returns hardly obey the normal distribution hypothesis. But, when using long short-term memory networks, sentiment analysis in addition to historical data leads to greater returns. In this project, we aim to build a system for predicting portfolio risk using AI/ML and provide insights on how the stocks will perform. We will train our model on datasets which include historical data of top 100 companies (NIFTY 100) in NSE and BSE from 2010 to 2021, Which is obtained from Yahoo Financial API.*

Key Words: Portfolio Management, Stock Predictions, Sentiment Analysis, FII, DII.

1.INTRODUCTION

Investment Portfolio Management is a concept where the risk correlated to the investment portfolio is reduced and have also tried to maximize the profit if it is withdrawn early from the particular stock. It has also been seen that public moods towards particular stocks are related with the financial markets to a greater extent.

We plan to use long short term memory networks, wherein we firstly do the sentiment analysis of tweets – where we compare different tweets related to the stock, which helps in getting better predictions. Secondly, we compare the portfolios of different Superstars to get a better prediction. Lastly, we go through the historical data that are available and compare it, so we get predictions accurately.

2. Literature Survey

A. Correlation of Stocks and News Content using Decision Tree Classifier: Salvatore M. Carta (Member, IEEE), Sergio Consoli, Luca Piras, Alessandro Sebastian Podda, and Diego Reforgiato Recupero propose to capture the association between terms in news reports and stock price fluctuations. This is accomplished through the use of a set of lexicons that contain the most influential words for a given industry during a specific time period. Second, the created lexicons are utilized to extract a collection of attributes that characterize industry and firm-specific news on a monthly, weekly, and daily basis. The feature vectors are then loaded into a Decision Tree classifier, which determines if the daily stock price variance is high or low. The Decision Tree's rules that decide the prediction are retrieved and displayed to the user as an explanation.

B. Artificial Intelligence techniques like Genetic Algorithms and Fuzzy Logic for stock market prediction and SVM Algorithms for Sentiment Analysis: Fernando G. D. C. Ferreira, Amir H.Gandomi, (Senior Member, IEEE) and Rodrigo T. N. Cardoso study various artificial intelligence techniques like Genetic Algorithms and Fuzzy Logic to predict stock prices from historical data. The paper also discusses sentiment analysis done using SVM Algorithms. The paper is then separated into four categories: portfolio optimization, stock market prediction using AI, financial sentiment analysis, and publications combining two or more domains.

C. Portfolio Optimization by combining Deep Neural Network Models and Mean Standard Absolute Deviation : Yilin Ma, Ruizhu Han, and Weizhong Wang offer a method for evaluating the prediction performance of DMLP, LSTM neural networks, and SVR across a four-year test period (2012-2015). To completely analyze their prediction skills, this research uses five evaluation metrics: mean absolute error (MAE), mean squared error (MSE), HR, HR+, and HR. The MAE and MSE measurements are regarded as the primary indicators among all the evaluation measures since predicted errors are closely connected with prediction-based portfolio models. DMLP, LSTM neural network, and CNN are three commonly used DNNs that have been shown to outperform standard ML technologies in terms of learning ability. To evaluate risk, a semi-absolute deviation metric is preferable to variance. To demonstrate their superiority, these models are compared to three equal weighted portfolio models (DMLP+EW, LSTM+EW, and CNN+MSAD), SVR+MSAD, and SVR+MV.

D. Latent Dirichlet Allocation for sentiment extraction from tweets and hypothesis check using Granger Causality Test : Salah Buoktif, Ali Fiaz (Member, IEEE), and Mamoun Awad offer a machine learning strategy that consists of five basic steps to anticipate stock movements. We scrape two sets of stock data from web sites. In step two, use NLP approaches to pre-process data before extracting various important aspects from tweets. Fit machine learning models as described in Section III and evaluate model accuracies in Step III. Step IV is activated if the accuracy falls below a particular threshold T. Step IV involves selecting and transforming features, and then refitting the machine learning models with a better collection of features. The accuracy of the model is then compared to that of the preceding stage. Finally, in step V, use a regularized model stacking to try to improve classification performance even more. In the Augmented Textual Features-Based Stock Market, Latent Dirichlet Allocation (LDA) is employed for topic modeling to assess the legitimacy of the tweet corpus. Granger Causality analysis is done to see if there is a statistically meaningful link between tweet-driven sentiment and stock returns.

E. Prediction from stock based images using Deep Q Networks : To efficiently train our NN model utilizing RL for stock prediction, Jinho Lee, Raehyun Kim, Yookyung Koh, and Jaewoo Kang adopt the DQN framework rather than traditional supervised learning. For the stock prediction issue, using RL (Q-learning) instead of traditional supervised learning gives a number of advantages. First, we use rewards to train our model. Assigning binary labels (e.g., True or False) to activities is insufficient because we are dealing with the stock price prediction problem. For example, if a model decides to take a Long action, it is preferable to receive a 10.0 reward for a future price change of +10 percent and a 1.5 reward for a subsequent price change of +1.5 percent. Receiving True in both situations does not distinguish between the two scenarios. Second, rather than using only immediate rewards to train an agent, RL uses cumulative rewards. Finally, a trained model can employ an action value in Q-learning, which is the expected cumulative rewards of the related action. So, after training, our model not only understands which action to take, but it can also anticipate how much profit the action would bring, allowing us to discern between strong and weak patterns.

F. Combining Sharpe Ratio and Genetic Algorithms for selecting profitable stocks: To answer the problem of stock selection, the Yao-Hsin Chou, Shu-Yu Kuo, Yi-Tzu Lo technique combines the Sharpe ratio and GA. We use stock prices to determine portfolio fund standardization, which improves risk evaluation. Funds standardization can accurately depict portfolio risk while also taking into account all of the interconnections between equities in a portfolio. Because this article does not limit the amount of stocks in a portfolio, there are a lot of possibilities in the search area. In a large search space, GA is used to locate the portfolio with the lowest risk and highest return. In addition, to prevent the over-fitting problem, we employ the sliding windows to select the best portfolio in the training period and to trade in the testing period.

2.1 Summary of Related Work

The Summary of methods used are given in Table.

Literature	Stock Prediction from Historical Data	Stock Prediction Using Sentiment Analysis	Hybrid
Salvatore M. Carta et al. 2021 [1]	No	Yes	No
Fernando G. D. et al. 2021 [2]	Yes	Yes	Yes
Yilin Ma. et al. 2020 [3]	Yes	No	No
Salah Buoktif, et al. 2020 [4]	No	Yes	Yes
Jinho Lee, et al 2019 [5]	Yes	No	No
Yao-Hsin, et al. 2017 [6]	Yes	No	No

Table- 1: Summary of literature survey

3. Proposed Work

The model proposes to take various inputs from different sources to build the effective risk manager. When input is given, all the relevant news from credible and available sources will be fetched along with top few tweets from twitter, Open interest (if available) and FIIS and DIIS data of current day , markets closing point price will be stored in the database. After the first stage of information collection is done the processing of data is initiated by sending NEWS and Tweets to the classifier to predict the general sentiment of the people and to classify them into Positive to Negative sentiments. The sentiments, current day market closing price of stock and FIIS and DIIS buying and selling activity will be given as an input to the LSTM Neural network. The predicted output will be shown to the user along with General Market sentiment and Open interest of the underlying asset.

3.1 System Architecture

The system architecture is given in Figure 1.

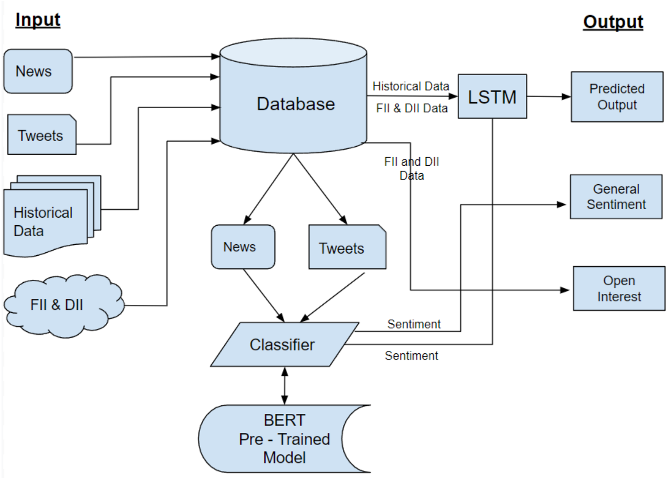

Figure 1: Proposed System Architecture

A. News:

Most of the time the stock price movement is in response to the speculation. It is highly regulated by the financial news and sentiments present in the market. As we know, the financial market is volatile, and this reflects in the share prices movements. By analyzing the News, we can approximate the stock market movements.

B. Tweets:

It has been observed that stock movements are directly related to sentiment of people. Twitter is such a place where people can tweet about what they think about a particular topic. This is of great use for us as it summarizes their sentiment towards any stock which we are interested in. The data are collected by Twitter Search API, where a search query consists of the stock cash-tag (e.g., “$NKE” for Nike).

C. FII and DII’s data:

Stock markets are primarily driven by institutional money. DIIs and FIIs account for the bulk of the liquidity in the market. Tracking FIIS and DIIS inflows and outflows can help predict the broader trends in the market. FIIS invest in domestic markets in large and bulk quantities. As FIIS do thorough research, and they do have their advanced in-house AI recommendations thus we can depend on those numbers.

D. Historical Data:

Historical data are day wise open, high, low and close of the given stock. This Historical data is used by many analysts and investors to back test and make investment decisions by mining patterns in data. These data are much reliable because the historical patterns may sometimes repeat in similar kinds of stock rally and fall

E. Database:

MongoDB:

MongoDB is a cross platform document oriented database program which is classified as a NoSQL database. MongoDB uses JSON-like documents along with optional schemas. MongoDB is developed by MongoDB Inc. MongoDB can provide many benefits as it is flexible schema which makes it easy to evolve and store data in a way that is easy to work with. MongoDB is built to scale up quickly and supports all the features of modern databases such as transactions.

F. Classifier:

The Sentiment Classification is a type of the Text Classification problem in NLP. We would be performing Binary text classification. The Classifier reads the input given and processes it to give positive and negative sentiment. By classifying the sentiment, we can get a moderately clear picture of the general sentiment of people which will further help in increasing the efficiency of the prediction.
For our use we will use BERT (Pre-Trained Model) as our classifier.

G. LSTM:

Long Short-Term Memory (LSTM) is one of many forms of Recurrent Neural Network RNN that can capture input from previous stages and use it to make predictions in the future. Because RNN isn't good for long-term memory, we chose LSTM, which has proven to be quite good at forecasting with long-term data. A cell, an input gate, a forget gate, and an output gate make up a typical LSTM unit. The three gates control the flow of information into and out of the cell, and the cell remembers values for arbitrary time intervals. Time series data is well-suited to LSTM networks for processing, classification, and prediction. LSTMs were created to solve the problem of vanishing gradients that can occur when training traditional RNNs.

H. Output:

Stock Price Forecast: Predicted Output Using Historical Data and Sentiment
Open Interest: Shows the Highest number of Call and Put option Strike price
General sentiment: Shows the general Sentiment of NEWS and Tweets
Because in stock market sometimes the markets may react differently from the General Sentiment

4. Requirement Analysis

The implementation detail is given in this section.

4.1 Software

Operating System	Windows 10
Programming Language	Python
Database	MongoDB
Framework	Django

4.2 Hardware

Processor	2 GHz Intel
HDD	180 GB
RAM	4 GB

4.3 Dataset and Parameters

The dataset for prediction is using the NIFTY 50 historical dataset from Yahoo Financial. FIIS and DIIS data are to be obtained from the NSE Official Website. Tweets are fetched from Twitter API. News are Fetched from Google News which appear on top search results.

5. Conclusion

The Combination of Machine Learning (LSTM), Technical(Option Chain), Sentiment analysis(News, Twitter) along with Foreign and Domestic Institutional Investors analysis gives the user high accuracy in predicting the short-term market efficiently.

Acknowledgement

It is our honor to thank our project coordinator Prof. K S. Charumathi and guide Prof. Dinesh Tharwani for their invaluable contributions, able guidance, encouragement, wholehearted cooperation, and constructive criticism during the course of this project. Dr. Sharvari Govilkar, our Department Head, and Dr. Sandeep M. Joshi, our principal, deserve our heartfelt gratitude for encouraging and permitting us to share this work.

References

[1] *Salvatore M. Carta, (Member, IEEE), Sergio Consoli, Luca Piras, Alessandro Sebastian Podda, Diego Reforgiato Recupero* "Explainable Machine Learning Exploiting News and Domain-Specific Lexicon for Stock Market Forecasting", Institute of Electrical and Electronics Engineers (IEEE) , Feb 2021

[2] *Fernando G. D. C. Ferreira, Amir H.Gandomi, (Senior Member, IEEE), Rodrigo T. N. Cardoso* "Artificial Intelligence Applied to Stock Market Trading: A Review" , Institute of Electrical and Electronics Engineers (IEEE) , Jan 2021

[3] *Yilin Ma, Ruizhu Han, Weizhong Wang* "Prediction-Based Portfolio Optimization Models Using Deep Neural Networks", Institute of Electrical and Electronics Engineers (IEEE), June 2020

[4] *Salah Buoktif, Ali Fiaz, (Member, IEEE), Mamoun Awad* "Augmented Textual Features-Based Stock Market Prediction" , Institute of Electrical and Electronics Engineers (IEEE) , Feb 2020

[5] *Jinho Lee, Raehyun Kim, Yookyung Koh, Jaewoo Kang* "Global Stock Market Prediction Based on Stock Chart Images Using Deep Q-Network" , Institute of Electrical and Electronics Engineers (IEEE) , Nov 2019

[6] *Yao-Hsin Chou, Shu-Yu Kuo, Yi-Tzu Lo* "Portfolio Optimization Based on Funds Standardization and Genetic Algorithm", Institute of Electrical and Electronics Engineers (IEEE), Nov 2017.

BIOGRAPHIES

Vettithanam Alex Sebastian

Freshin Francis

Abin Varghese Jacob

Gaikar Prateek Balu

Prof. Dinesh S Tharwani.
He is a Assistant professor at PCE, New Panvel. Having more than 13 years of teaching experience in the field of Computer Engineering. He has published about 4 research papers in various national and international conferences.