IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.918 ⁜ Vol. 11, Issue 12, December 2022

DOI: 10.17148/IJARCCE.2022.111209

Stock Data Prediction And Sentiment Analysis using Financial News Headlines

Dr. (Mrs.) S.R. Khonde [1], Adarsha Masalkar [2], Tejas Upase [3], Mayur Bhoyar[4], Viraj Nikam[5]

Guide, Computer Engineering, Modern Educations Society's College of Engineering, Pune, India [1]

Student, Computer Engineering, Modern Educations Society's College of Engineering, Pune, India [2]

Student, Computer Engineering, Modern Educations Society's College of Engineering, Pune, India [3]

Student, Computer Engineering, Modern Educations Society's College of Engineering, Pune, India [4]

Student, Computer Engineering, Modern Educations Society's College of Engineering, Pune, India [5]

Abstract: News has always been an important source of information to build perception of market investments. As the volumes of news and news sources are increasing rapidly, it is becoming impossible for an investor or even a group of investors to find out relevant news from the big chunk of news available. This forecasts real time news sentiment that reflects stock price movement trends. The aim is to create a headline scraping algorithm for sentiment analysis, choosing the most advanced and accurate algorithm for classification, and integrating the models in a single application.

Keywords: API, GUI, News Sentiment, Prediction Model

I. INTRODUCTION

Prediction of stock price variation is a very challenging task and the price movement behaves more like a random walk and time varying. In recent times, researchers have used various types of AI techniques to make trading decisions.

II. LITERATURE SURVEY

Sheta [1] used the Takagi-Sugeno (TS) technique to develop fuzzy models for two nonlinear processes. They were a software effort estimate for NASA software projects and next week's S&P 500 forecast for the stock market. The development of the TS fuzzy model can be achieved in two steps 1) determining the membership functions in the antecedents of the rules using the input data of the model; 2) estimation of consequences parameters. They used least squares estimation to estimate these parameters. The results were promising.

WengLuen Ho et al. [2] proposed an intraday financial trading system with a predictive model supported by a new brain inspired by the evolving MamdaniTakagi-Sugeno Neural-Fuzzy Inference System (eMTSFIS). The eMTSFIS predictive model possessed the synaptic mechanisms and information processing capabilities of the human hippocampus, resulting in a more robust and adaptive predictive model compared to existing econometric and neural-fuzzy techniques. The trading strategy of the proposed system was based on the moving average convergence/divergence (MACD) principle for generating buy and sell trading signals. By introducing predictive capabilities into the calculation of MACD trend signals, the lagging nature of the MACD trading rule can be addressed. Experimental results based on the S&P500 index confirmed that eMTSFIS was able to provide highly accurate forecasts and the resulting system was able to identify timely trading opportunities while avoiding unnecessary trading transactions. These attributes enabled the trading system based on eMTSFIS to bring higher multiplier returns to the investor.

Nagar and Hahsler in their research [3] presented an automated text mining approach to aggregate news from different sources and build a news corpus. The corpus is filtered for relevant sentences and analysed using natural language processing (NLP) techniques. A sentiment metric, called NewsSentiment, using the number of words with positive and negative polarity, is proposed as a sentiment measure of the overall XI news corpus. They used various open-source packages and tools to develop a news gathering and aggregation tool as well as a sentiment rating tool. They also report that the time variation of NewsSentiment shows a very strong correlation with the actual movement of stock prices.

Ching Long Su et al. [4] developed a self-organized five-layer neuro-fuzzy model for modelling stock market dynamics using technical indicators. The effectiveness of the model in prediction and forecasting was verified with a dataset containing four indicators: Stochastic Oscillator, Volume Adjusted Moving Average (VAMA) and Ease of Movement

DOI: 10.17148/IJARCCE.2022.111209

(EMV) from TAIEX (Taiwan Stock Exchange Capitalization Weighted Stock Index). A modified moving average method can be proposed to predict the input set for the neuro-fuzzy model in stock price forecasting. The simulation results showed that the model was effective in prediction and accurate in forecasting. The input error from the modified moving average method prediction was greatly attenuated by the neuro-fuzzy model to achieve better prediction results. M.H. Fazel, Zarandi [5] developed an expert system based on type 2 fuzzy rules for stock price analysis. An interval type 2 fuzzy logic system allowed rule uncertainties to be modelled, and each feature membership value was itself an interval. The proposed type 2 fuzzy model applied technical and fundamental indices as input variables. The model can be tested to predict the stock price of an automobile manufacturer in Asia. Through intensive experimental testing, the model successfully predicted price fluctuations for stocks from various sectors. The results were very encouraging and were implemented in a real-time trading system to predict stock prices during the trading period.

Kelvin Sim [6] proposed a 3D subspace clustering method to generate rules for selecting potential undervalued stocks. 3D subspace clustering is efficient in processing high-dimensional financial data and is adaptable to new data. The obtained results are not influenced by human biases and emotions and are easy to interpret. Experimenting extensively in the stock market for 28 years (from 1980 to 2007), XII found that using rules generated by 3D subspace clustering algorithms, CAT Seeker and MIC, resulted in 60 percent higher profits than using Graham's rules alone.

Xiaodong Li, Xiaodi Huang, Xiaotie Deng and Shanfeng Zhu [7] used MKSVR, the system quantitatively analyzes and integrates intraday market news and stock price. Experiments were conducted using market reports and tick data on the Hong Kong stock market for one full year. The results showed that MKSVR is better able to exploit the hidden information in news articles and historical stock prices than models that simply use news articles to predict stock prices. MKSVR has also been shown to outperform those models that use only one information source.

The profits generated by the eMAQT implemented by Qing Li, Tie Jun Wang, Ping Li, Ling Liu, Qixu Gong and Yuanzhu Chen [8] further prove previous claims in finance that public information events are subject to different interpretations by investors. This result presents profitable trading opportunities for qualified investors to generate profit after the release date of the report. This means that stock markets are sensitive to public information in the era of social media.

The contribution of the study conducted by S.F. Crone and C. Koeppel [9] is summarized as follows. First, while previous research considered overall sentiment in documents, this research proposed a method using topic sentiment to predict the stock market. Second, we proposed two methods to capture these sentiment-topic associations. One is a JST-based method that relies on an existing topic model, the other is an aspect-based sentiment method where topics and sentiments are identified by the proposed method. Finally, this is the first research to demonstrate the effectiveness of incorporating sentiment analysis using research on large-scale test data. From a practical point of view, although the average accuracy is only 54.41%, the proposed method can predict the stock price movement with more than 60% accuracy for several stocks and performs much better than other methods for stocks that are difficult to predict with only past prices.

Research conducted by Daisuke Katayama and Kazuhiko Tsuda[10] confirmed that sentiment, which assessed news according to the polarity dictionary, has some effect on the stock market. Specifically, they created three hypotheses and tested each one. Hypothesis 1 is that if there is positive news, the company's stock price will increase afterwards, the result is as predicted. Many investors believe that they judge the content of the news and make investment decisions. Another hypothesis is that the effect of Hypothesis 1 will increase for a front-page article. The result of the regression analysis confirmed that this hypothesis is correct.

With G Chowdhury, S Routh, S Chakrabarti [11] their study using a predictive model collected news at market close and then analysed the sentiment of each company. This 4-week sentiment track reveals a strong correlation between the original stock price curve and the effective positive index curve generated by the predictive news mining model introduced in this article. This study reveals that there is a very strong correlation of about 67% between news sentiment and the original stock price curve. This easily meets the efficient market hypothesis test and shows that sentiment is positively reflected in stock price movements.

László Nemes, Attila Kiss [12] used various sentiment analysis tools to emotionally analyze and classify various economic news headlines and investigate their impact on various stock market value changes even without their full context. Emotions were divided into the usual positive negative and neutral categories. Neutral categories emerged for the TextBlob and NLTK-VADER Lexicon tools, but not for the Recurrent Neural Network (RNN). The results of various sentiment analyses were compared with the result of BERT as a benchmark.

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified ⁜ Impact Factor 7.918 ⁜ Vol. 11, Issue 12, December 2022
DOI: 10.17148/IJARCCE.2022.111209

Yi Zuo, Eisuke Kita [13] In this study, a P/E ratio forecasting algorithm using a Bayesian network was presented. They digitized the P/E ratio values by clustering the frequency distribution of the P/E ratio using uniform clustering or Ward's method. A Bayesian network for the dependence between the prior distribution of the P/E ratio is determined from the digitized values of the P/E ratio. The forecasting accuracy and correlation coefficient with respect to the actual stock price are compared with traditional time series forecasting algorithms such as AR, MA, ARMA and ARCH models.

S. F. Crone and C. Koeppel [14] proved that sentiment indicators are able to describe market behaviour. Considering continuous returns, the nonlinear descriptive model can achieve a directional accuracy of 75.64\% in the validation set and a real-world rate of 60.26\% in the generalization set. Empirical results concluded that nonlinear models outperform linear regressions as well as simple benchmark models such as the naive method. Bivariate analysis also showed that the relationship between exchange rate and various sentiment increases after market movements. The analysis revealed the ability to describe financial time series using sentiment indicators.

T. Sidogi, R. Mbuvha and T. Marwala [15] investigated the effect of sentiment from financial news headlines on stock price predictability using Long-Term Short-Term Memory (LSTM) networks. The investigation is conducted on intraday data with specific delays between published article headlines and realized share prices. They performed a principled comparison of the performance of LSTM models for forecasting the stock market under the same conditions, but with an objective assessment of the importance of incorporating sentiment from financial news as model inputs.

III. SOFTWARE REQUIREMENT SPECIFICATION

A. Purpose and Scope of Document

A software requirements specification (SRS) is a document that is created when a detailed description of all aspects of the software to be built must be specified before the project is to commence. It is important to note that a formal SRS is not always written. In fact, there are many instances in which effort expended on an SRS might be better spent in other software engineering activities.

B. Scope of Project

Prediction of stock price variation is a very challenging task and the price movement behaves more like a random walk and time varying. In recent times, researchers have used various types of AI techniques to make trading decisions. News plays a vital role in the overall direction of the stock market. The main scope of this project is to find the dependency of the news reports and articles on stock price.

C. Usage Scenario

Usage scenario of the project is in the finance sector. The project can be used by investors, stock traders as well as hedge for stock price prediction and sentiment analysis for getting the statistics and analysis of stocks based on news.

- User profiles
- Actors: User, Admin
- Use-cases Users

use the platform to analyse the market sentiment before market open time. The user input is optional, and the NLP model will generate a report regarding the sentiment. The news data will be pulled from financial news websites and will be used in the NLP model. Traders and investors will use the platform to get a statistical and analytical insight of stock and general market sentiment.

D. Assumptions & Dependencies

The data will be coming from the news websites. We are assuming that the data provided by the news websites is accurate. The whole output of the system depends on the accuracy of data.

IV. FUNCTIONAL REQUIREMENTS

A. Frontend Subsystem

- Description

The frontend module of our project is the part which will be visible to the user. The user will exclusively interact with this module. The Frontend will be a GUI Web Application developed in Python Flask framework for visual aesthetics.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.918 ⁜ Vol. 11, Issue 12, December 2022

DOI: 10.17148/IJARCCE.2022.111209

There will be an option to provide input (Login, input for stock specific scans which is optional) and this input will be passed to the backend and backend will perform the necessary computations on it and give the desired output which will be passed back to the frontend.

Stimulus/Response Sequences the Frontend will be loaded when the user inputs the address/domain of the project and no additional stimulus is required to load this module. Functional Requirements

REQ-1: JavaScript enabled Browser
REQ-2: Internet Connectivity.

B. Backend Subsystem

- Description

Natural language processing model will be developed using Python. Pandas library will be used to convert the dataset in csv format to data frames. The dataset could be having some values missing, some incorrect information i.e. the data is not clean. Therefore, the data cleaning of the data set will be handled by the Python backend. Scikit-learn (Sklearn) library is used as it provides a selection of efficient tools for machine learning and statistical modelling including classification, regression, clustering and dimensionality reduction. These models are provided input from the Dataset during the training phase and from the frontend of the web application during normal use.

Django is a high-level python web framework that enables rapid development of application without compromising the security of the application. Security features are enabled by default in this framework. As our users will XVII be providing their personal data, the first and foremost responsibility as a developer is to develop a robust and secured website which can be done using Django. For storing the data, MySQL database is used.

Stimulus/Response Sequences The Backend module will be loaded when the user inputs the required data and clicks on submit button. The backend module is not loaded by default and is loaded only when the frontend makes a function call to backend. Functional Requirements

REQ-1: JavaScript enabled Browser
REQ-2: Internet Connectivity.

V. External Interface Requirements

A. User Interfaces

As this project is a GUI Web Application there will be an option to provide input to user, this input will be passed to the backend and backend will perform the necessary computations on it and give the desired output which will be passed back to the frontend. The Frontend will now display the output received from the backend to the user. Depending on who is the user different outputs will be shown.

B. Hardware Interfaces

The Project itself does not use any Hardware Interface but still it requires a system to run itself. As it is a Machine Learning on Web Based Implementation, this project can be run on any device with internet support. Supported device types include Mobile Phones, Smart Television, Desktops, Laptops etc. Communication protocols to be used is HTTPS as the web application will be hosted on a Web Server.

C. Software Interfaces

Python 3.8.3, fyers-apiv2, Flask, BeautifulSoup, Scrapy, JavaScript, HTML, CSS.

D. Communication Interfaces

The requirements associated with communications functions required by this project, include any web browser with JavaScript support for running the model in backend. The main communication, i.e. between the website and database, is done using a database connectivity.

VI. System Requirements

A. Database Requirements

Data objects that will be managed/manipulated by the software are described in this section. The database entities or files or data structures required to be described. For data objects details can be given as below.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.918 ⁜ Vol. 11, Issue 12, December 2022

DOI: 10.17148/IJARCCE.2022.111209

B. Software Requirements

[For User]: Any Browser with JavaScript Connectivity, Internet Connectivity [For Development]: Scikit-Learn, Python, Pandas, Numpy, Matplotlib, Flask, HTML, JavaScript, CSS

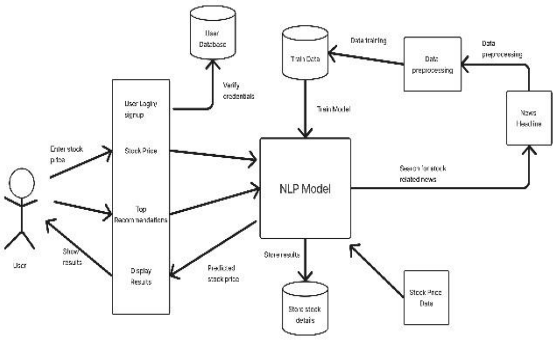
C. Hardware Requirements

[For User]: Any Device with Internet Connectivity and ability to run websites. [For Development]: GPU, CPU.

VII. ANALYSIS MODEL

We will be using Agile Model in our development. We will be discussing and implementing necessary changes from time to time with each other and with our project guide. We will be distributing the work among each other and working on different modules simultaneously.

VIII. SYSTEM ARCHITECTURE



IX. CONCLUSION

We have proposed a Natural Language Processing based Sentiment Analysis model to predict and study the correlation between stock news and stock price. In order to find correlation between sentiment predicted from news, the NIFTY 50 stocks will be used. Also, we can also study other factors ranging from social to legal issues to check whether at all they have any impact on overall market scenario. Using this proposed model, we can do several financial modelling like portfolio management, risk estimation models, and several other strategic modelling etc. where perception plays an important role.

The project can be taken to larger scale by adding few more news websites. With improvements in the ML algorithms, the classifiers can be updated timely. Larger datasets can help in providing better accuracy in future.

REFERENCES

[1]. A Sheta, "Software Effort Estimation and Stock Market Prediction Using Takagi-Sugeno Fuzzy Models", In Proceedings of the IEEE International Conference on Fuzzy Systems, pp.171-178, Vancouver, BC, 2006.

[2]. WengLuen Ho, Whye Loon Tung and Chai Quek, "Brain-Inspired Evolv- ing Neuro-Fuzzy System for Financial Forecasting and Trading of the S&P500 Index", Lecture Notes in Computer Science, Vol.6230, pp.601- 607,2010

[3]. 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

[4]. Su, Ching & Chen, Chuen & Yang, Shih. (2010). A self-organized neuro-fuzzy system for stock market dynamics modeling and forecast- ing. WSEAS Transactions on Information Science and Applications. 7. 1137-1149.

[5]. Zarandi, Mohammad Hossein Fazel, Babak Rezaee, I. Burhan Tu¨rksen and Elaheh Neshat. "A type-2 fuzzy rule-based expert system model for stock price analysis." Expert Syst. Appl. 36 (2009): 139-154.

[6]. Sim, Kelvin & Gopalkrishnan, Vivekanand & Phua, Clifton & Cong, Gao. (2014). 3D Subspace Clustering for Value Investing. Intelligent Systems, IEEE. 29. 52-59. 10.1109/MIS.2012.24.

DOI: 10.17148/IJARCCE.2022.111209

[7]. Xiaodong Li, Xiaodi Huang ,Xiaotie Deng and Shanfeng Zhu "Enhancing quantitative intra-day stock return prediction by integrating both market news and stock prices information" Neurocomputing Volume 142, 22 October 2014, Pages 228-238

[8]. Qing Li ,Tie Jun Wang ,Ping Li ,Ling Liu ,Qixu Gong and Yuanzhu Chen "The effect of news and public mood on stock movements" Book Information Sciences Volume 278, 10 September 2014, Pages 826-840

[9]. Thien Hai Nguyen ,Kiyoaki Shirai and Julien Velcin "Sentiment analysis on social media for stock movement prediction" Journal Expert Systems with Applications Volume 42, Issue 24, 30 December 2015, Pages 9603-9611

[10]. Daisuke Katayama and Kazuhiko Tsuda "A Method of Measurement of The Impact of Japanese News on Stock Market" Journal Procedia Computer Science Volume 126, 2018, Pages 1336-1343

[11]. Spandan Ghose Chowdhury , Soham Routh and Satyajit Chakrabarti " News Analytics and Sentiment Analysis to Predict Stock Price Trends" International Journal of Computer Science and Information Technologies, Vol. 5 (3) , 2014, 3595-3604

[12]. László Nemes \& Attila Kiss (2021) Prediction of stock values changes using sentiment analysis of stock news headlines, Journal of Information and Telecommunication, 5:3, 375-394, DOI: 10.1080/24751839.2021.1874252

[13]. Yi Zuo, Eisuke Kita "Stock price forecast using Bayesian network, Expert Systems with Applications" Volume 39, Issue 8, 2012, Pages 6729-6737

[14]. S. F. Crone and C. Koeppel, "Predicting exchange rates with sentiment indicators: An empirical evaluation using text mining and multilayer perceptrons," 2014 IEEE Conference on Computational Intelligence for Financial Engineering \& Economics (CIFEr), 2014, pp. 114-121, doi: 10.1109/CIFEr.2014.6924062

[15]. T. Sidogi, R. Mbuvha and T. Marwala, "Stock Price Prediction Using Sentiment Analysis" 2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC), 2021, pp. 46-51, doi: 10.1109/SMC52423.2021.9659283