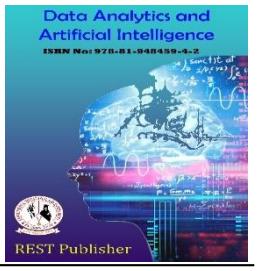
Data Analytics and Artificial Intelligence

Vol: 3(3), 2023

REST Publisher; ISBN: 978-81-948459-4-2

Website: http://restpublisher.com/book-series/daai/



Fake News Detection Using Machine Learning

***S. Sowjanya, V. Suruthi, M. Vaijayanthi**

Adhiyamaan College of Engineering(Autonomous), Hosur,Tamilnadu,India

*Corresponding author Email: **sowjanyashivanna14@gmail.com**

Abstract- *Nowadays, in some exchanges, people are spreading a lot of information every second. Fake news is a big problem all over the world. Fake news and some stock market rumors have created some application in some sector to allow us to make misleading or grossly exaggerated statements and so on. Public and cause great loss of money. Checking stock market platforms Merrill Edge, Etro or thinkers and many other platforms can accelerate the distribution of real-time information to users around the world. On the stock exchange, most users do not verify the information and spread it. Identifying fake news manually is a big problem for everyone. So the need for an automatic system that effectively detects fake news about the stock market. This project estimated a model that intuitively distinguishes fake news from a stock price news article. A new feature set for a machine learning classifier was proposed. In the experiment, the dataset used has a combination of two datasets, Google finance and API, which contain the same true and false stock price news and current stock news. Here we implement an LSTM algorithm for fake news detection in the stock market. TF bag of words and BOW TF-IDF vector from the text fields of the dataset, then apply various machine learning models including bagging and boosting methods to achieve the best accuracy of 98.02%. When the user wants to buy the stock price, if the stock price matches the actual price, then the user can buy, if not, then the user cannot buy, an email will be sent to the user using SMTP, and the data will be saved in a CSV file. Through this, all data can be displayed at any time.*

Keywords: *Stock Market Price, Deep Learning, Fake News of stock, BOW-TF, IDF.*

1. INTRODUCTION

The fight against fake news is becoming an unsolved problem in social networks in the application layer of data and information consumption, and it is becoming a serious and challenging problem in the development of information that appears in the diplomatic, economic and political sectors. The detection of false information indicates an unnecessary process in network resources. In addition, it includes content completeness and validity according to the service available. [1] These terms are becoming limited based on recent developments in machine learning, deep learning and artificial intelligence. Checking the authorship of digital content is one of the mandatory steps for information sharing. For this, block chain is a suitable and promising framework, which is a decentralized and secure platform to improve the extraction of false information. The block chain system continuously increases the number of blocks, each block having the previous block's cryptographic hash, timestamp and transaction information. [2] He proposed a message tracking system using block chain-based allocation of credits to users. News verification using this system will cause the user to request news verification from the publisher. Table 1 shows the related terms and the concept of social media fake news detection. The terms below are not considered fake news in social media [3]. Truthfulness of information is an essential part of its integrity. The fight against fake news makes the control of the integrity and truth of information about social networks and data consumption in the application layer inseparable. Posting fake content is a waste of process and network resources. Furthermore, it is a serious threat to the integrity of information and the credibility of the service provided. [4] The term fake news originally refers to false and often sensational information spread under the guise of relevant news. However, the use of the term has evolved and is now considered synonymous with spreading false information on social media. [5]

2. OBJECTIVE

This research paper studies the possibility of using deep learning techniques to discriminate fake messages on the Internet using only their text. The web and internet life has led to the entry of news data, many fake news from the stock market can be avoided. Mass social media news affects the lives of the general public, and as often happens, fake news. There are few social media that use fake news in the stock market and we can identify them. To identify various social

media sources and analyze whether a given stock article report is credible or not. The accuracy rate is limited to 95% because it was recommended to use a linguistic model. Higher accuracy will be achieved.

3. EXISTING SYSTEM

In this method, machine learning models have been implemented in trading and are trained on historical stock prices and other quantitative data to predict future stock prices. However, natural language processing (NLP) allows us to analyze financial documents such as 10-thousand forms and predict stock movements. Forms 10-k are annual reports filed by companies that provide a comprehensive overview of their financial performance (these reports are mandated by the Securities and Exchange Commission). This method uses the KNN algorithm to detect stock prices of many products. Through sentiment analysis, a subfield of natural language processing, investors can quickly understand whether the tone of a news release is positive, negative, or controversial, etc.

4. DISADVANTAGES

Since it uses historical data and only forecasts, current price accuracy will not be achieved. Investors will not be able to get a clear overview of the current share price. The KNN algorithm used in this method is time-consuming and provides an accuracy of 60 percent proposed system In the proposed system, the user should log in to the site by entering a username and password. We connect with Google finance and API address to cross check stock price and fake stock price. Implementation of LSTM algorithm for accurate message data. The LSTM algorithm is a type of recurrent neural network (RNN) that is specifically designed for processing sequential data such as time series and text. The accuracy of LSTM algorithm will be 92%, this algorithm is implemented for more accurate and correct data to help investors to make the right decisions, also it will not send an email if the user tries to buy the wrong stock price through SMTP protocol. Bag-of-words, tied produces a normalized count where each word count is divided by the number of documents in which that word occurs. In this proposed system, we have avoided the fake financial stock market news that are published in some news to avoid the loss of people who invest their money in the wrong stock market, they can compare the real value of the stock brand with google finance stock market. Advantages The algorithm implemented in this project will be accurate and correct. The time required to implement this algorithm will be shorter and also less complex. By implementing this project, the costs will be lower compared to the existing system. This method gives investors a clear knowledge of tracking current stock prices.

5. BLOCK DIAGRAM

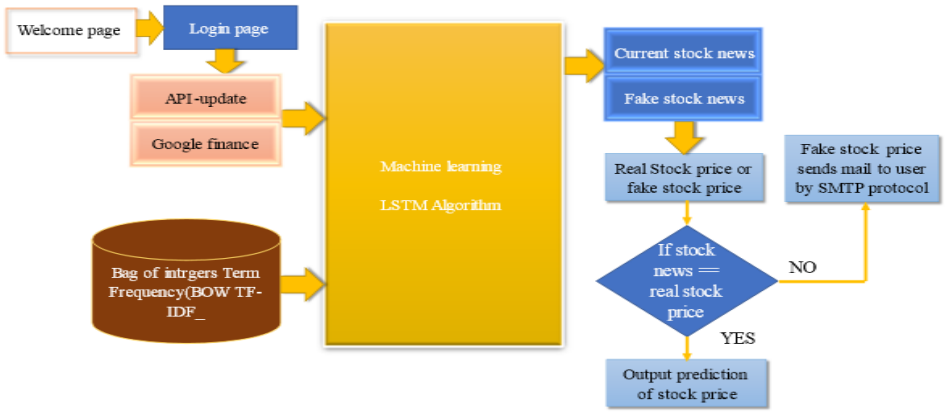


FIGURE 1. Block Diagram

5. MODULE

Login Module: A login module is a portal module that allows users to enter a login username and password. You can add this module to any module tab to allow users to log into the system. **Google Finance Module:** Google Finance provides an easy way to search for financial security data (stocks, mutual funds, indices, etc.), currency exchange rates and cryptocurrencies ("Finance Data"). Financial data is collected from various data providers and sources into a unified format available for provision to users.

LSTM Algorithm Module: It is a type of Recurrent Neural Network (RNN) that is specially designed for processing sequential data such as time series, speech and text

SQLite Module: SQLite is a continuous library that implements a server less, configuration-free standalone transactional SQL database engine.

PyQt5 Module: PyQt5 is the latest version of the GUI toolkit developed by Riverbank Computing. It is a Python interface for Qt, one of the most powerful and popular cross-platform GUI libraries. PyQt5 is a mixture of the Python programming language and the Qt library.

Output Module: In this output module, if the user buys the actual stock price, the output will be predicted because the actual value of the stock price will be displayed. If the user is not buying the wrong stock price, then the email will be sent as a fake stock price via SMTP and the data will be stored in CSV files.

6. SYSTEM FUCTION

API stands for Application Programming Interface, which is a set of communication protocols and routines used by different programs to communicate with each other. A programmer can use various API tools to make his program easier and simpler. The API also facilitates an efficient way for programmers to develop their software programs. So, in simpler terms, an API helps two programs or applications communicate with each other by providing them with the necessary tools and functions. It takes the request from the user and sends it to the service provider and then resends the result generated by the service provided to the requested user. API is extensively used by the developer in his software to implement various functions using API calls without writing complex codes for the same. We can create an API for an operating system, a database system, a hardware system, a JavaScript file, or similar object-oriented files. Also, API is similar to GUI (Graphical User Interface) with one main difference. Unlike a GUI, an API helps software developers access web tools, while a GUI helps users understand the program more easily.

FIGURE 2. API Connection with Google finance

LSTM Algorithm: LSTM networks are well-suited for classification, processing, and forecasting based on time series data because there may be unknown time lags between important time series events. Long short-term memory (LSTM) networks are a type of recurrent neural network that can learn the order dependence of sequence prediction problems. This is a necessary operation for complex problem areas such as machine translation and speech recognition. Long Short-Term Memory is a type of recurrent neural network. It is a special kind of recurrent neural network that is able to learn long-term dependencies in data. This is achieved because the repeating module of the model has a combination of four layers that interact with each other.

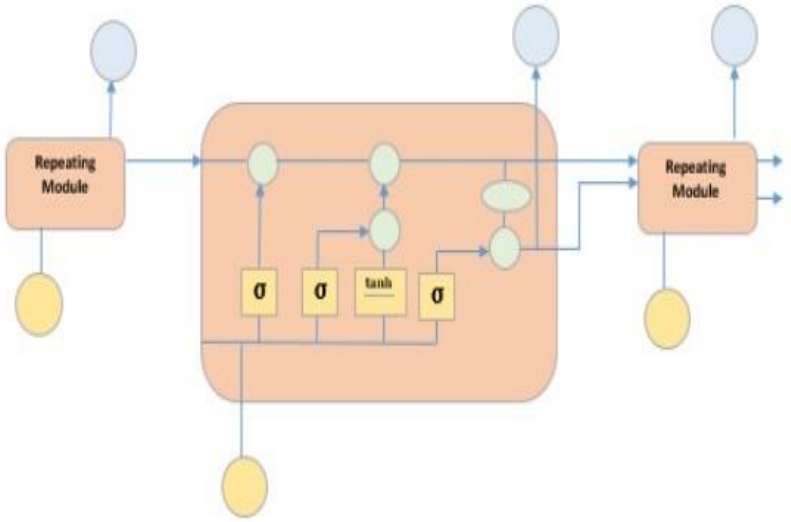


FIGURE 3. LSTM Algorithm

Data base Management System: A DBMS is a collection of data and information that is stored in a dataset. It is a small computer system and provides ways for users or other programs to modify or extract data. DBMS stands for Database Management System. We can break it down like this: DBMS = Database + Management System. A database is a collection of data and a management system is a set of programs for storing and retrieving that data. Based on this, we can define a DBMS as follows: A DBMS is a collection of interrelated data and a set of programs for storing and accessing that data in an easy and efficient way.

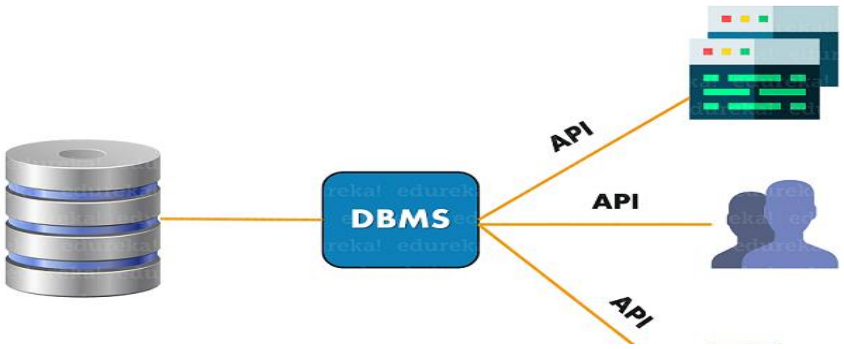


FIGURE 4. Data Base Management System

Google Finance: Google Finance provides an easy way to find data about financial securities (stocks, mutual funds, indices, etc.), currency exchange rates, and crypto-currencies ("Financial Data"). Financial data is aggregated from various data providers and sources into a unified format available for delivery to users. Google Finance ranks search suggestions based on three main elements: exact query matches, visibility on Google Search, and visibility on Google Finance. Exact matches to queries are prioritized and then appear in Google Search and Google Finance, which are given equal importance.

FIGURE 5. Google Finance

7. RESULTS

This can be interpreted as testing the significance over time of the abnormal price effect of stock fake news, cumulative after the disclosure shock. The most robust of the empirical tests performed, these results account for potential cross-correlation bias, unrepresentative excessively volatile observations, and non-standard normal distributions.

FIGURE 6. Stock User Interface

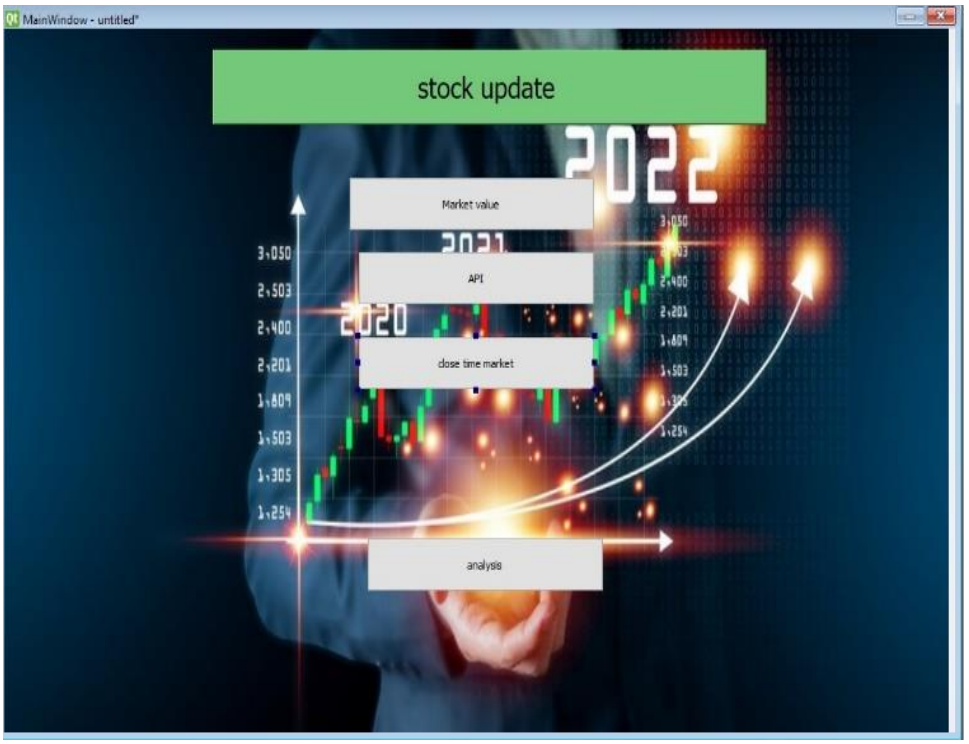


FIGURE 7. Stock Update

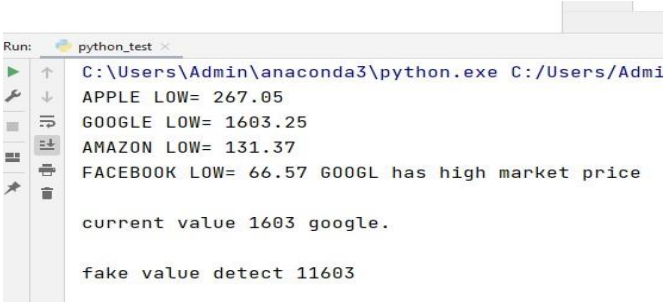


FIGURE 8. Stock Value

10. CONCLUSION

In this work, we analyze the fake news robustness of stock price detectors to stop believing fake news about the stock price. For this purpose, we use four different architectures – multi-layer Perceptron (LSTM), Long short-term memory (LSTM) algorithm, LSTM proposed Hybrid Fake News Stock Detector – and multiple datasets – BOW TF-IDF dataset training, BOW TF dataset and BOW TF-IDF. We vary the complexity of the detectors and experiment with different input lengths. Our findings indicate the real price of the stock, so people can stop investing their money on the false price of the stock and get lost.

REFERENCES

[1]. Nicollas R de Oliveira, Dianne SV Medeiros, and Diogo MF Matto, A sensitive stylistic approach to identify fake news on social networking, 2020.

[2]. Arvind Narayanan, Joseph Bonneau, Edward Felten, Andrew Miller, and Steven Goldfeder , Bitcoin and cryptocurrency technologies ,2019.
[3]. Xinyi Zhou and Reza Zafarani , A survey of fake news: Fundamental theories, detection methods, and opportunities. ACM Computing Surveys, 2020.
[4]. de Oliveira, N.R.; Medeiros, D.S.V.; Mattos, D.M.F. , A Sensitive Stylistic Approach to Identify Fake News on Social Networking , 2019.
[5]. Sharma, K.; Qian, F.; Jiang, H.; Ruchansky, N.; Zhang, M.; Liu, Y., Combating fake news: A survey on identification and mitigation techniques , 2019.
[6]. H. Ali, S. Nepal, S. S. Kanhere, and S. Jha, ''HaS-Nets: A heal and select mechanism to defend DNNs against backdoor attacks for data collection scenarios,'' 2021.
[7]. A. C. Serban and E. Poll, ''Adversarial examples—A complete characterisation of the phenomenon,'' 2020.
[8]. H. Ali, F. Khalid, H. A. Tariq, M. A. Hanif, R. Ahmed, and S. Rehman, ''SSCNets: Robustifying DNNs using secure selective convolutional filters,'' IEEE Des. Test. Comput., Apr. 2020.
[9]. J. X. Morris, E. Lifland, J. Y. Yoo, J. Grigsby, D. Jin, and Y. Qi, ''Textattack: A framework for adversarial attacks, data augmentation, and adversarial training in NLP,'' in Proc. Conf. Empirical Methods Natural Lang. Process., Syst. Demonstrations, (EMNLP), Q. Liu and D. Schlangen, Eds. Stroudsburg, PA, USA: Association for Computational Linguistics, Nov. 2020, pp. 119–126.
[10]. F. Khalid, H. Ali, M. A. Hanif, S. Rehman, R. Ahmed, and M. Shafique, ''FaDec: A fast decision-based attack for adversarial machine learning,'' in Proc. Int. Joint Conf. Neural Netw. (IJCNN), Jul. 2020, pp. 1–8
[11]. A. Athalye, N. Carlini, and D. Wagner, ''Obfuscated gradients give a false sense of security: Circumventing defenses to adversarial examples,'' in Proc. 35th Int. Conf. Mach. Learn. (ICML) (PMLR), vol. 80, J. Dy and A. Krause, Eds. Stockholm, Sweden: PMLR, Jul. 2020, pp. 274–283.
[12]. Z. Zhou, H. Guan, M. Bhat, and J. Hsu, ''Fake news detection via NLP is vulnerable to adversarial attacks,'' in Proc. 11th Int. Conf. Agents Artif. Intell., Feb. 2019.
[13]. D. Su, H. Zhang, H. Chen, J. Yi, P. Chen, and Y. Gao, ''Is robustness the cost of accuracy—A comprehensive study on the robustness of 18 deep image classification models,'' in Proc. 15th Eur. Conf. Comput. Vis. (ECCV) in Lecture Notes in Computer Science, vol. 11216, V. Ferrari, M. Hebert, C. Sminchisescu, and Y. Weiss, Eds., Munich, Germany: Springer, Sep. 2019, pp. 644–661.
[14]. C. Wardle and H. Derakhshan, ''Information disorder: Toward an interdisciplinary framework for research and policy making,'' Council Eur. Rep., vol. 27, pp. 1–107, Sep. 2020.
[15]. M. Usama, J. Qadir, and A. Al-Fuqaha, ''Adversarial attacks on cognitive self-organizing networks: The challenge and the way forward,'' in Proc. IEEE 43rd Conf. Local Comput. Netw. Workshops (LCN Workshops), Oct. 2021.
[16]. D. Song, K. Eykholt, I. Evtimov, E. Fernandes, B. Li, A. Rahmati, F. Tramer, A. Prakash, and T. Kohno, ''Physical adversarial examples for object detectors,'' in Proc. 12th USENIX Workshop Offensive Technol. (WOOT), 2020, p. 1.
[17]. M. Usama, M. Asim, S. Latif, and J. Qadir, ''Generative adversarial networks for launching and thwarting adversarial attacks on network intrusion detection systems,'' in Proc. 15th Int. Wireless Commun. Mobile Comput. Conf. (IWCMC), Jun. 2019.
[18]. M. Usama, A. Qayyum, J. Qadir, and A. Al-Fuqaha, ''Black-box adversarial machine learning attack on network traffic classification,'' in Proc. 15th Int. Wireless Commun. Mobile Comput. Conf. (IWCMC), Jun. 2019.
[19]. S. Latif, R. Rana, and J. Qadir, ''Adversarial machine learning and speech emotion recognition: Utilizing generative adversarial networks for robustness,'' 2019.

