FAKE NEWS DETECTION ON SOCIAL MEDIA PLATFORMS USING MACHINE LEARNING ALGORITHMS

KEE WEE BOON

UNIVERSITI TEKNOLOGI MALAYSIA

FAKE NEWS DETECTION ON SOCIAL MEDIA PLATFORMS USING MACHINE LEARNING ALGORITHMS

KEE WEE BOON

A project report submitted in partial fulfilment of the
requirements for the award of the degree of
Master of Engineering (Computer and Microelectronic Systems)

School of Electrical Engineering
Faculty of Engineering
Universiti Teknologi Malaysia

JULY 2022

DEDICATION

This project report is dedicated to my father, who taught me that the best kind of knowledge to have is that which is learned for its own sake. It is also dedicated to my mother, who taught me that even the largest task can be accomplished if it is done one step at a time.

ACKNOWLEDGEMENT

I would like to thank everybody who had contributed to the successful accomplishment of this project. First of all, I would like to express my special thanks of gratitude to my research supervisor, Dr. Mohd Azhar Bin Abdul Razak for his invaluable guidance, advice and his huge patience throughout the development of the research. I came to know about so many new things I am really thankful to them.

In addition, I would also like to extend my gratitude to my loving parents and friends who have been always encouraging and assisting me throughout the year. I have really no valuable words to express my appreciations, but my heart is still full of the favours received from every individual.

ABSTRACT

The rapid growth of advanced technology and social media platforms has changed the way of retrieving information or news for the whole world. However, it is undeniable that there is some fake news which is purposely created to disseminate false information to the public. The rapid spreading of fake news brings a significant negative impact on individuals, culture and country. Although there are some existing fake news detection methods like manual fact-checking websites and tools, they are time-consuming for fake news detection and cannot provide real time detection. Hence, an automated machine learning based approach is required for fake news detection since fake news can be disseminated rapidly through online social media platforms. This project aims to propose and construct an effective hybrid machine learning based algorithm for automated fake news detection. This algorithm will assist people in differentiating real and fake news, reduce the threats to national security by preventing the wide spread of fake news and maintain the news ecosystem's genuineness equilibrium. The project flow started with the dataset collection which includes both real and fake news. This dataset is publicly available and it is obtained from Kaggle website. Next, some text pre-processing techniques, i.e., lower case conversion, punctuation and stopwords removal and lemmatization will be applied to the raw data. Furthermore, Term Frequency – Inverse Document Frequency (TF-IDF) is the technique applied for text vectorization in this project. Two machine learning algorithms, Random Forest (RF) and Support Vector Machine (SVM) are implemented to classify the real and fake news. In addition, two hybrid ensemble models, namely voting classifier and stacking classifier are constructed by combining three individual base classifiers, i.e. RF, SVM and Logistic Regression (LR). Hybrid ensemble model is an approach that combines several models to improve the prediction accuracy. These hybrid classifiers can detect fake news in real time and effectively prevent the fake news from being disseminated widely through social media platforms. In this project, stacking classifier performs better than voting classifier, achieving an accuracy of 91.06% as compare to the classification accuracy of 90.41% obtained by voting classifier.

ABSTRAK

Pertumbuhan teknologi canggih dan platform media sosial yang pesat telah mengubah cara untuk mendapatkan maklumat atau berita bagi seluruh dunia. Namun begitu, memang tidak dapat dinafikan bahawa terdapat beberapa berita palsu yang sengaja dicipta untuk menyebarkan maklumat palsu kepada awam. Penyebaran berita palsu yang pantas membawa kesan negatif yang ketara kepada individu, budaya dan negara. Walaupun terdapat beberapa kaedah pengesanan berita palsu yang sedia ada seperti laman web dan alatan semakan fakta manual, ia memakan masa untuk pengesanan berita palsu dan tidak dapat memberikan pengesanan masa nyata. Oleh itu, pendekatan berasaskan pembelajaran mesin automatik diperlukan untuk pengesanan berita palsu kerana berita palsu boleh disebarkan dengan cepat melalui platform media sosial dalam talian. Projek ini bertujuan untuk mencadangkan dan membina sebuah algoritma berasaskan pembelajaran mesin hibrid yang berkesan untuk pengesanan berita palsu automatik. Algoritma ini akan membantu orang ramai untuk membezakan berita sebenar dan palsu, mengurangkan ancaman kepada keselamatan negara dengan menghalang penyebaran berita palsu secara meluas dan mengekalkan keseimbangan ketulenan ekosistem berita. Aliran projek bermula dengan pengumpulan set data yang merangkumi kedua-dua berita sebenar dan palsu. Seterusnya, beberapa teknik pra-pemprosesan teks, iaitu, penukaran huruf kecil, penyingkiran tanda baca dan kata henti serta lemmatisasi akan digunakan pada data mentah. Tambahan pula, *Term Frequency – Inverse Document Frequency* (TF-IDF) ialah teknik yang digunakan untuk vektorisasi teks dalam projek ini. Dua algoritma pembelajaran mesin, *Random Forest* (RF) dan *Support Vector Machine* (SVM) digunakan untuk mengklasifikasikan berita sebenar dan palsu. Selain itu, dua model ensembel hibrid, iaitu pengelas undian dan pengelas tindanan dibina dengan menggabungkan tiga pengelas asas individu, iaitu RF, SVM dan *Logistic Regression* (LR). Model ensemble hibrid ialah pendekatan yang menggabungkan beberapa model untuk meningkatkan ketepatan ramalan. Pengelas hibrid ini boleh mengesan berita palsu dalam masa nyata dan berkesan menghalang berita palsu daripada disebarkan secara meluas melalui platform media sosial. Dalam projek ini, pengelas tindanan berprestasi lebih baik daripada pengelas undian, mencapai ketepatan 91.06% berbanding ketepatan pengelasan 90.41% yang diperolehi oleh pengelas undian.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AdaBoost	-	Adaptive Boosting
BiLSTM	-	Bidirectional Long Short-Term Memory
BoW	-	Bag-of-Words
CNN	-	Convolutional Neural Network
DNN	-	Dense Neural Network
DT	-	Decision Tree
FN	-	False Negatives
FP	-	False Positives
GB	-	Gradient Boosting
GRU	-	Gated Recurrent Unit
IDE	-	Integrated Development Environment
IDF	-	Inverse Document Frequency
KNN	-	K-Nearest Neighbor
LR	-	Logistic Regression
LSTM	-	Long Short-Term Memory
NB	-	Na ïve Bayes
NLP	-	Natural Language Processing
OOB	-	Out-of-bag
Regex	-	Regular expression
RF	-	Random Forest
RNN	-	Recurrent Neural Network
SVM	-	Support Vector Machine
TF	-	Term Frequency
TF-IDF	-	Term Frequency – Inverse Document Frequency
TN	-	True Negatives
TP	-	True Positives
XGBoost	-	Distributed Gradient Boosting

LIST OF SYMBOLS

C	-	Regularization parameter
n	-	Number of words in all documents
n_estimators	-	Number of trees in Random Forest
v_k	-	TF-IDF value for k^{th} word

LIST OF APPENDICES

CHAPTER 1

INTRODUCTION

1.1 Background Study

With the rapid growth and improvement of the internet, people prefer to retrieve or share information through online social media platforms such as Facebook, WhatsApp, Twitter, YouTube and many more. Nowadays, people can easily access any interested information by searching keywords on search engines. As a result, all the information or news articles can be retrieved through online platforms at the touch of a button. Unfortunately, there are some news which are purposely created to disseminate false information to the public and these are known as fake news.

For instance, there is some fake news related to COVID-19 pandemic spread through online social media platforms recently. Social media platforms serve as the main source of information and news especially during COVID-19 outbreak period. Some people will take this opportunity to create and publish fake news on social media platforms. This fake news is intended to mislead people to have a distorted opinion. Another reason for creating fake news is for the purpose of gaining some economic or political benefits (Qawasmeh et al., 2019). Fake news will bring significant negative impacts to individuals and countries if they are widely disseminated. Due to this reason, fake news has become an increasingly matter of contention in both industry and academia.

1.2 Motivation

There are some existing fake news detection methods used to detect and prevent the spreading of fake news on social media platforms. However, these existing fake news detection methods have their own limitations while detecting the fake news.

For instance, manual fact-checking such as Politifact and Snopes requires a significant amount of time in order to detect fake news (Liu, 2019). In other words, these manual fact-checking approaches cannot detect fake news in real time. As a result, fake news has already been disseminated through social media platforms before it is being detected.

With the advancement of technology, a lot of fake news is created and propagated widely on social media platforms for someone's benefit. However, manual approaches are not able to handle well with the growing number of freshly created fake news. Other than that, some existing fake news detection methods even require more external information such as the source and the comments of the news to determine the veracity of that news. Unfortunately, this information may sometimes not be available in some of the news articles which limits the efficiency and effectiveness of detecting fake news (Liu, 2019).

Due to the limitations of existing methods, an automated fake news detection algorithm is necessary to detect and prevent the spreading of fake news since fake news can be disseminated rapidly through online platforms. Machine learning and deep learning algorithms have been proved to be promising and able to effectively detect fake news through many studies. Thus, this project proposed to implement machine learning approaches in detecting fake news on social media platforms.

1.3 Problem Statement

A study proved that it is a hard decision for humans to differentiate between a fact and lie statement when they are overloaded with misleading information (Zhou et al., 2020). In other words, people tend to believe every published news article regardless of the truthfulness of the news article. This is because people always tend to process the information and make a quick decision through mental shortcuts. As a result, they share the misleading information without even verifying the information's veracity.

The rapid spreading of fake news through social media platforms brings a significant negative impact on individuals, culture and country. For instance, fake news will affect the news ecosystem's genuineness equilibrium. Fake news is also intended to persuade someone to have misleading views towards certain truths or facts (Abbad et al., 2021). Furthermore, the spread of fake news has a chance to instigate social unrest which can bring threats to national security (Liu, 2019).

1.4 Aim and Objectives

The main aim of this project is to propose and construct an effective hybrid machine learning based algorithm for automated fake news detection.

The objectives of this project are as follows:

(a) To assist people in differentiating real and fake news before they directly share without verifying the veracity of news.

(b) To reduce the threats to national security by preventing the wide spreading of fake news.

(c) To maintain the news ecosystem's genuineness equilibrium and avoid people having misleading views due to fake news.

1.5 Research Scope

The objectives of this project are as follows:

(a) This project utilized publicly available dataset from Kaggle website.

(b) TF-IDF technique is used for text vectorization.

(c) Two machine learning algorithms, i.e., RF and SVM are implemented for classifying real and fake news.

(d) Two hybrid classifiers are constructed to further improve the classification accuracy.

(e) Spyder, an Integrated Development Environment (IDE) is used for implementing pre-processing, feature extraction and classification steps.

1.6 Chapter Summary

This project is organized as follows. Chapter 1 is introducing the background study of research especially the fake news on social media platforms. Besides, the motivation, problem statements, aim and objectives and research scope of this project are also clearly mentioned in this chapter.

Next, Chapter 2 provides a detailed literature review after reading several articles and journals. This chapter introduces some existing fake news detection methods, including manual and computational approaches. Besides, the definition and significance of dataset, text pre-processing techniques, text features and machine learning algorithms are mentioned in this chapter. A summary of related works is provided at the end of this chapter.

Chapter 3 explains the project's methodology. This chapter started with a brief introduction which included all the steps involved for fake news detection. Furthermore, all the steps are further explained in detail in each subsection. Two Gantt charts showing the project timelines are also provided at the end of this chapter.

In Chapter 4, all the results and discussions are clearly mentioned. All the results including classification accuracy, precision, recall and F1-score are tabulated for all the classification algorithms implemented in this project. Chapter 5 summarized the whole project. Besides, there are also some recommendations provided for future works.

REFERENCE

Abbad, M., Kumar, G., Samiullah, M., &Kumar, N. S. (2021). A Predominant Advent to Fake News Detection using Machine Learning Algorithm. *2021 International Conference on Intelligent Technologies, CONIT 2021*. https://doi.org/10.1109/CONIT51480.2021.9498436

Agarwal, A., &Dixit, A. (2020). Fake News Detection: An Ensemble Learning Approach. *Proceedings of the International Conference on Intelligent Computing and Control Systems, ICICCS 2020*, 1178–1183. https://doi.org/10.1109/ICICCS48265.2020.9121030

Agrawal, R. (2021). *Must Known Techniques for text preprocessing in NLP Table of Contents Introduction to NLP Text Preprocessing Libraries used to deal with NLP problems Techniques for Text Preprocessing*. 1–10.

Anello, E. (n.d.). *A Friendly Guide to NLP : TF- IDF With Python Example*. 1–17.

Brownlee, J. (2017). Data, Learning and Modeling. *Machine Learning Mastery*, 1–19. https://machinelearningmastery.com/data-learning-and-modeling/

Dagdagan, J. (n.d.). *Fake News Detection : Text Pre- Processing*. 1–15.

Dimitriadis, S. I., &Liparas, D. (2018). How random is the random forest? Random forest algorithm on the service of structural imaging biomarkers for Alzheimer's disease: From Alzheimer's disease neuroimaging initiative (ADNI) database. *Neural Regeneration Research*, *13*(6), 962–970. https://doi.org/10.4103/1673-5374.233433

Draelos, R. (2019a). Best Use of Train/Val/Test Splits, with Tips for Medical Data. *Machine Learning and Medicine*, 1. https://glassboxmedicine.com/2019/09/15/best-use-of-train-val-test-splits-with-tips-for-medical-data/

Draelos, R. (2019b). Measuring Performance: The Confusion Matrix. *Glass Box*, *1*, 2019–2021. https://glassboxmedicine.com/2019/02/17/measuring-performance-the-confusion-matrix/

Elnemr, H. A., Zayed, N. M., &Fakhreldein, M. A. (2015). *Feature Extraction Techniques*. 264–294. https://doi.org/10.4018/978-1-4666-8654-0.ch013

Gandhi, R. (2018). Support Vector Machine — Introduction to Machine Learning

Algorithms. *Towards Data Science*, 1. https://towardsdatascience.com/support-vector-machine-introduction-to-machine-learning-algorithms-934a444fca47

Jiang, T., Li, J. P., Haq, A. U., Saboor, A., &Ali, A. (2021). A Novel Stacking Approach for Accurate Detection of Fake News. *IEEE Access*, *9*, 22626–22639. https://doi.org/10.1109/ACCESS.2021.3056079

Lin, J., Tremblay-Taylor, G., Mou, G., You, D., &Lee, K. (2019). Detecting Fake News Articles. *Proceedings - 2019 IEEE International Conference on Big Data, Big Data 2019*, 3021–3025. https://doi.org/10.1109/BigData47090.2019.9005980

Liu, H. (2019). A Location Independent Machine Learning Approach for Early Fake News Detection. *Proceedings - 2019 IEEE International Conference on Big Data, Big Data 2019*, 4740–4746. https://doi.org/10.1109/BigData47090.2019.9005962

Mahir, E. M., Akhter, S., &Huq, M. R. (2019). 2019 7th International Conference on Smart Computing and Communications, ICSCC 2019. *Detecting Fake News Using Machine Learning and Deep Learning Algorithms*, 1–5.

Patwa, P., Sharma, S., Pykl, S., Guptha, V., Kumari, G., Akhtar, M. S., Ekbal, A., Das, A., &Chakraborty, T. (2021). Fighting an Infodemic: COVID-19 Fake News Dataset. *Communications in Computer and Information Science*, *1402 CCIS*, 21–29. https://doi.org/10.1007/978-3-030-73696-5_3

Priya, A., &Kumar, A. (2021). Deep Ensemble Approach for COVID-19 Fake News Detection from Social Media. *2021 8th International Conference on Signal Processing and Integrated Networks (SPIN)*, 396–401. https://doi.org/10.1109/spin52536.2021.9565958

Qawasmeh, E., Tawalbeh, M., &Abdullah, M. (2019). Automatic Identification of Fake News Using Deep Learning. *2019 6th International Conference on Social Networks Analysis, Management and Security, SNAMS 2019*, 383–388. https://doi.org/10.1109/SNAMS.2019.8931873

Rajalakshmi, C., Subika, T., Vaishali, K., &Shana, J. (2021). Fake News Prediction On COVID Dataset Using Machine Learning. *2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, 1–7. https://doi.org/10.1109/icccnt51525.2021.9579543

S, U. C. (2021). *How sklearn ' s Tfidfvectorizer Calculates tf-idf Values What is Term Frequency (tf) Inverse Document Frequency (idf) Term Frequency-Inverse*

Document Frequency (tf-idf) Python program to generate tf-idf values. *26000*, 1–8.

Training Set vs Validation Set vs Test Set How to Split N-Fold Cross-Validation. (n.d.). 1–5.

Zhou, X., Jain, A., Phoha, V.V., &Zafarani, R. (2020). Fake News Early Detection. *Digital Threats: Research and Practice*, *1*(2), 1–25. https://doi.org/10.1145/3377478