

BMS COLLEGE OF ENGINEERING

(Autonomous College under VTU)

Bull Temple Road, Basavanagudi, Bangalore - 560019



A Project report on

“Fake News Detection on Social Media”

Submitted in partial fulfillment of the requirements for the award of degree

BACHELOR OF ENGINEERING

IN

INFORMATION SCIENCE AND ENGINEERING

By

Chiranjeevi Nayak B - 1BM20IS402

Yash Agarwal - 1BM19IS182

Voona Pranay - 1BM19IS181

Under the guidance of

K.R MAMATHA

Assistant Professor

Department of Information Science and Engineering

2022 - 2023

BMS COLLEGE OF ENGINEERING
(Autonomous Institute, Affiliated to VTU)
Bull Temple Road, Basavanagudi,
Bengaluru – 560019



Department of Information Science and Engineering

CERTIFICATE

This is to certify that the project entitled “**FAKE NEWS DETECTION ON SOCIAL MEDIA**” is a bona-fide work carried out by **Chiranjeevi Nayak B(1BM20IS402), Voona Pranay(1BM19IS181), Yash Agarwal (1BM19IS182)** in partial fulfillment for the award of degree of Bachelor of Engineering in **Information Science and Engineering** from **Visvesvaraya Technological University, Belgaum** during the year **2022-2023**. It is certified that all corrections/suggestions indicated for Internal Assessments have been incorporated in the report deposited in the departmental library. The project report has been approved as it satisfies the academic requirements in respect of project work prescribed for the Bachelor of Engineering Degree.

Under the Guidance of

K.R. Mamatha

CONTENTS

SL NO	TOPIC	PAGE NO
	Abstract	4
1	Chapter 1: Introduction	5
2	Chapter 2: Problem Statement	6
3	Chapter 3: Literature Survey	7-10
4	Chapter 4: Requirements Analysis	11-12
5	Chapter 5: Design & methodology 5.1 System Architecture 5.2 Data Flow Diagram	13-14
6	References	15-16

ABSTRACT

Consuming news from social media is becoming increasingly popular nowadays. Social media brings benefits to users due to the inherent nature of fast dissemination, cheap cost, and easy access. However, the quality of news is considered lower than traditional news outlets, resulting in large amounts of fake news. Detecting fake news becomes very important and is attracting increasing attention due to the detrimental effects on individuals and the society. The performance of detecting fake news only from content is generally not satisfactory, and it is suggested to incorporate user social engagements as auxiliary information to improve fake news detection. Thus it necessitates an in- depth understanding of the correlation between user profiles on social media and fake news. We perform a comparative analysis over explicit and implicit profile features between these user groups, which reveals their potential to differentiate fake news.

Chapter 1

INTRODUCTION

With the increasing popularity of social media, more and more people consume news from social media instead of traditional news media. Fake news is now viewed as one of the greatest threats to democracy, and freedom of expression. It has weakened public trust in government. The reach of fake news will be best highlighted all the time. The extensive spread of fake news has the potential for extremely negative impacts on individual and society. A type of yellow journalism, fake news encapsulates pieces of news that may be hoaxes and is generally spread through social media and other online media. This is often done to further or impose certain ideas and is often achieved with political agendas. Such news items may contain false and/or exaggerated claims, and may end up being verbalized by algorithms, and users may end up in a filter bubble

Chapter 2

PROBLEM STATEMENT

Given a multi-source news dataset and social contexts of news consumers (social media users), the task of fake news detection is to determine if a news item is fake or real. Formally, we define the problem of fake news detection as:

Input: News items, social contexts and associated side information

Output: One of two labels: ‘fake’ or ‘real’.

Fake news detection on social media presents unique characteristics and challenges that make existing detection algorithms from traditional news media ineffective or not applicable. First, fake news is intentionally written to mislead readers to believe false information, which makes it difficult and nontrivial to detect based on news content; therefore, we need to include auxiliary information, such as user social engagements on social media, to help make a determination. Second, exploiting this auxiliary information is challenging in and of itself as users’ social engagements with fake news produce data that is big, incomplete, unstructured, and noisy.

Chapter – 3

LITERATURE SURVEY

M. F. Mridha et al.[1] This survey covers the overall analysis of fake news classification by discussing major studies. A thorough understanding of recent approaches in fake news detection is essential because advanced frameworks are the front-runners in this domain. Thus, they analysed fake news identification methods based on NLP and advanced DL strategies. They explored different NLP techniques and DL architectures and provided their strength and shortcomings.

K.ArunKumar et al.[2] Proposed a different approach for false news detection using one of the best trained deep neural techniques and LSTM. The LSTM helps in training faster with large datasets compared to other neural network models. Texts or images containing information are processed and recognized as reliable or unreliable using a deep neural network model. The preprocessing of the image are done using LSTM and it also proves to be more efficient than the existing systems. Analyzed image results are converted into text and then processes it . The LSTM model has trained and validated with more than 1000 images collected from various online sources.

Federico Monti et al.[3] In this paper, presented a geometric deep learning approach for fake news detection on Twitter social network. There solution achieves very high accuracy and robust behaviour in several challenging settings involving large-scale real data, pointing to the great potential of geometric deep learning methods for fake news detection.

Jayashree M Kudari et al.[4] Paper discusses on comprehensively review, summarize, compare and evaluate the current research on fake news. By this paper uses passive aggressive and TF-IDF vectorizer and it is efficient as we obtained 90% of accuracy from this model.

Kai Shu et al.[5] In this article, they explored the fake news problem by reviewing existing literature in two phases: characterization and detection. In the characterization phase, they introduced the basic concepts and principles of fake news in both traditional media and social media. In the detection phase, they reviewed existing fake news detection approaches from a data mining perspective, including feature extraction and model construction. They also further discussed the datasets, evaluation metrics, and promising future directions in fake news detection.

Ankit Kesarwani et al.[6] describes a simple approach for fake news detection with the help of K- Nearest Neighbor classification algorithm. They collect and analyse the particular features of users in various aspects like implicit and explicit. Prediction is done on the basis K most similar instances by searching the entire training dataset using Euclidean distance. After testing our model,they got 79% accuracy.

Abdulrahman et al.[7] The approach used in this study is divided into three stages. In the first stage (pre-processing), this study used certain data filtering and cleaning techniques to extract semantic features from the raw dataset. During the second stage (extracting numerical features), techniques were used to convert semantic features into feature vectors, and the last stage (classifiers) employed machine learning and deep learning classifiers to group the items in the dataset. These were the best techniques through which we obtained quality results, with an average rating accuracy of 91.23%.

Anoud et al.[8] developed a semantic model for Context-based,it approaches focus on contextual data for news articles.3 major social context areas can be observed across research.These are users, generated posts and networks. With this, both real and fake news datasets can be analysed by automatically mapping social context data for each news to

corresponding classes within the domain, thereby fostering analysis for both and subsequently, identifying metrics for classification of news articles as either real or fake.

Pakindessama et al.[9] build a semi supervised learning model that is able to classify a news at the early stage of its spread. They extracted the opinion expressed in the replies to the tweets using LSTM and Global vector about a given news, evaluated the credibility of the users who posted a tweet or replied about the news and evaluated the relationship between these users to determine either a given news is fake or real with an accuracy 71.10%.

Shikun et al.[10] proposed an approach of employing Doc2vec to vectorize dataset and feeds feature vectors to Decision Tree to train and identify the real data. Extracted four parts of the JSON document, which are 'URL', 'text', 'author', 'title'. The method chosen was to split the data by comma, and the first step was to collect the four features which include 'URL', 'text', 'author', 'title'. Precision and Accuracy of both Decision tree and SVM are calculated. Decision Tree classifier has a 5% enhance than SVM. So the model can help us to judge if a news is real or fake, and the accuracy was more than 95 percent.

Akash Dnyandeo Waghmare et al.[11] In this research, a novel idea of the machine Learning-based Model to Detect Fake News using Blockchain is presented. There are several methods experimented within the existing systems but the proposed system deals with custom blockchain and machine learning classification algorithms. In experiment, it is found that minimum computation time taken by custom blockchain during the module execution. At present, the system framework is partly deployed. We plan to introduce a fully working real-time framework to identify and eliminate fake news as part of our future research.

Anjali jain, Avinash Shakya et al.[12] In the paper, the components for recognizing Fake news are discussed. A mindfulness that not all, the fake news will propagate via web-based networking media. Currently, to test out the proposed method of Nave Bayesclassifier, SVM, and NLP are used. In future, ensuing algorithm may provide better results with hybrid approaches for the same purpose fulfilment. The mentioned system detects the fake news on

the based on the models applied. Also it had provided some suggested news on that topic which is very useful for any user. In the future, the efficiency and accuracy of the prototype can be enhanced to a certain level, and also enhance the user interface of the proposed model

Marwan Albahar et al.[13].A hybrid model based on an RNN and a SVM to detect rumours in news content in FakeNews datasets composed of two subparts, the PolitiFact and GossipCop datasets. RNN was utilised for encoding news content and comments to feature representation. The features were given as input to an SVM with a Gaussian kernel to detect rumours (real or fake news) in the input data. Results in terms of the accuracy and F1 score of our proposed model were compared with existing state-of-the-art models. Various performance measures show that our work has outperformed those existing models.

M. Sreedevi et al.[14].This paper had about five different classification models which are compared by the accuracies. SVM got the highest accuracy and it is about 99.48%. In this paper we consider only twitter data for prediction of fake tweets. Our future scope is to predict fake news against political affairs, economic affairs and Health affairs on facebook and other social sites data.

Ugur Mertgolu ,Burkay Genc et al.[15]. In this study we presented a novel framework which offers a practical architecture of an integrated system for identifying fake news. We have tried to illustrate how digital libraries can be a service authority to promote media literacy and fight against fake news. Because librarians are trained to critically analyze information sources, their contributions to our proposed model are critical. Accordingly, we see this work as an encouraging effort for the next collaborative studies among the communities of LIS and CS.

Chapter 4

REQUIREMENTS ANALYSIS

3.1 Functional Requirements

- **Comparison of predictor variable with other variables:** Making inferences from comparing other variables under consideration.
- **Evaluation Metrics:** Multiple evaluation metrics to be used such as F1-Score, Precision, Recall, etc. to conclude on an optimally sound model.

3.2 Non-Functional Requirements

- **Performance :** The system should respond to a user query and return a result in less than 5 seconds.
- **Safety and Security Requirements :** System should be safe and meet the minimum security requirements
- **Software Quality Attributes :** The required Attributes to be present in software Quality

3.3 Operating Environment

- Windows XP/Vista/7/8/10 (All Docker compatible versions of Windows)
- Mac OS X 10.13 or newer
- Linux (Ubuntu/Debian/CentOS)

3.4 Technical Requirements

- Google Colab
- Jupyter Notebooks
- Scikit-Learn
- Numpy, Pandas, Matplotlib, Seaborn
- Dataset

Chapter 5

DESIGN & METHODOLOGY

5.1 System Architecture

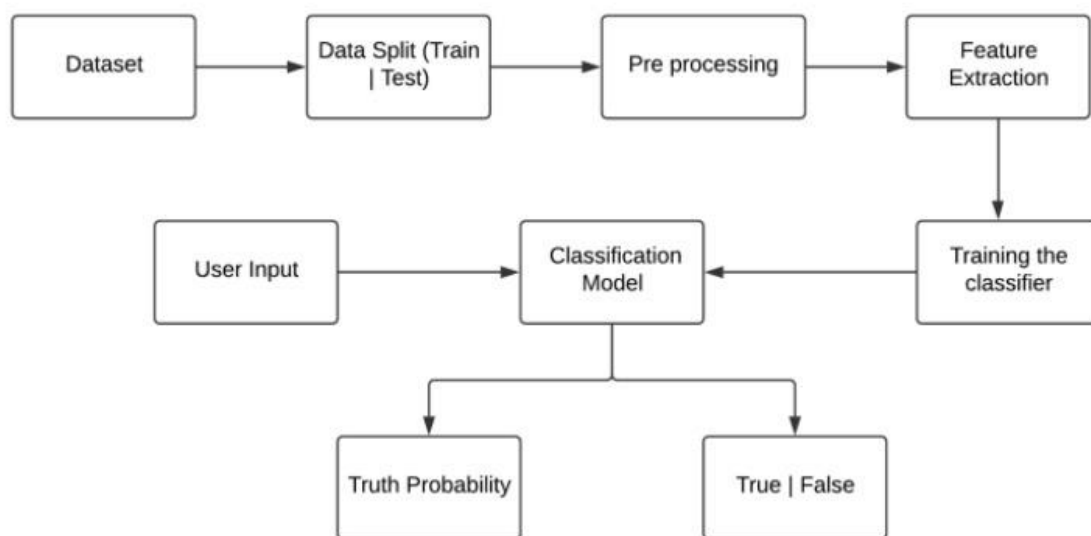


Figure 5.1: This diagram demonstrates a high-level view of the underlying architecture used in the application.

5.2 Data Flow Diagram

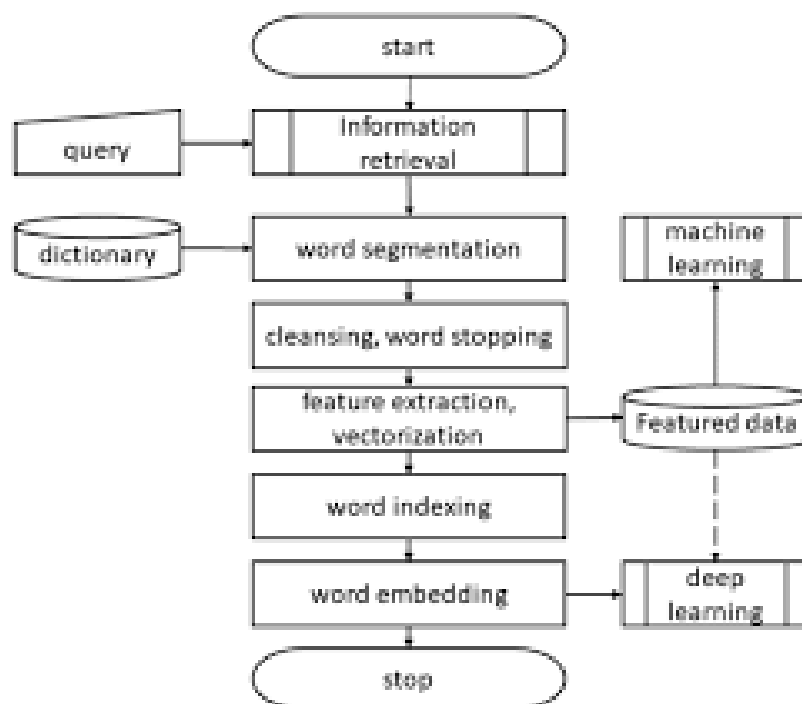


Figure 5.2: This diagram represents the flow of data within the application

Chapter 6

REFERENCES

- [1] M. F. MRIDHA, A Comprehensive Review on Fake News Detection With Deep Learning, date of publication November 18, 2021, date of current version November 30, 2021.
- [2] K.ArunKumar, A Study of Fake News Detection using Machine Learning Algorithms, ISSN: 0976-1345.
- [3] Federico Monti, Fake news detection on social media using geometric deep learning
- [4] Jayashree M Kudari, Fake News Detection using Passive Aggressive and TF-IDF Vectorizer, Issue: 09 | Sep 2020.
- [5] Kai Shu, Fake News Detection on Social Media: A Data Mining Perspective.
- [6] Ankit, Sudakar, Anil,2020. Fake News Detection on Social Media using K-Nearest Neighbor Classifier.
- [7] Abdulrahman, Muhammet,2020. Fake News Detection Using Machine Learning and Deep Learning Algorithms.
- [8] Anoud, Oluwasegun, Elhadj, Munir, Feras,2021. A Semantic Model for Context-Based Fake News Detection on Social Media.
- [9] Pakindessama, Rui, Siyuan, Toussida, Jose, Lin,2021. A Deep Learning Model for early Detection of Fake News on Social Media.
- [10] Shikun, Dan,2021. Fake News Detection by Decision Tree.
- [11] Carlos Castillo, Marcelo Mendoza, and Barbara Poblete. Information credibility on twitter. In WWW'11.
- [12] Abhijnan Chakraborty, Bhargavi Paranjape, Sourya Kakarla, and Niloy Ganguly. Stop clickbait: Detecting and preventing click baits in online news media. In ASONAM'16.

[13] Yimin Chen, Niall J Conroy, and Victoria L Rubin. Misleading online content: Recognizing clickbait as false news. In Proceedings of the 2015 ACM on Workshop on Multimodal Deception Detection, pages 15–19. ACM, 2015.

[14] Justin Cheng, Michael Bernstein, Cristian Danescu Niculescu-Mizil, and Jure Leskovec. Anyone can become a troll: Causes of trolling behavior in online discussions. In CSCW '17.

[15] Zi Chu, Steven Gianvecchio, Haining Wang, and Sushil Jajodia. Detecting automation of twitter accounts: Are you a human, bot, or cyborg? IEEE Transactions on Dependable and Secure Computing, 9(6):811–824, 2012.