

Fake News Detection Using Pattern Recognition

By

Name: Aniket Sinha Roy

University Roll No: 12200119019

Name: Subham Tripathi

University Roll No: 12200119018

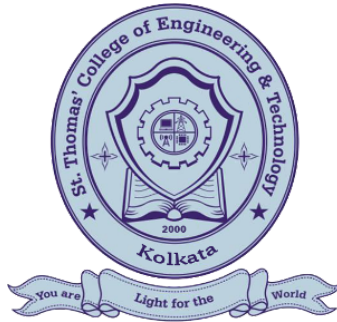
Name: Surajeet Sai

University Roll No: 12200119013

Under the esteemed guidance of

Ms. Ranjita Chowdhury

(Assistant Professor of Computer Science & Engineering Department)



A project synopsis submitted for the partial fulfillment of

Bachelor of Technology in the

Department of Computer Science and Engineering

St. Thomas' College of Engineering and Technology

Affiliated to

Maulana Abul Kalam Azad University of Technology,

West Bengal

St. Thomas' College of Engineering and Technology
Department of Computer Science and Engineering

DECLARATION

We declare that this written submission represents our ideas in our own words and we have adequately cited and referenced the original sources. We also declare that we have adhered to all the principles of academic honesty and integrity and have not misrepresented, fabricated, or falsified any idea/data/fact/source in our submission. We understand that any violation of the above will be cause for disciplinary action by the institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

Name: Aniket Sinha Roy
University Roll No: 12200119019

Name: Subham Tripathi
University Roll No: 12200119018

Name: Surajeet Sai
University Roll No: 12200119013

Table of Contents

S.No	Topic	Page. No
I	Preamble	i
	I.I Vision and Mission	i
	I.II Program Outcome (PO) and Program Specific Outcome (PSO)	ii
	I.III PO and PSO mapping with justification	iii
	Abstract	
1	Introduction	1
	1.1 Problem Statement	1
	1.2 Objective	1
	1.3 Literature Survey	1
	1.4 Brief Discussion on Problem	4
	1.5 Planning	4
	1.6 Organization	6
2	Concepts And Problem Analysis	6
	2.1 Problem Analysis	6
	2.2 Concepts	6
	2.3 Proposed Algorithms	8
3	Conclusion	10
4	Reference	10

I. Preamble

I.I Vision and Mission

Vision of the Institute

To evolve as an industry-oriented, research-based Institution for creative solutions in various engineering domains, with an ultimate objective of meeting technological challenges faced by the Nation and the Society.

Mission of the Institute

- To enhance the quality of engineering education and delivery through accessible, comprehensive, and research-oriented teaching-learning-assessment processes in the state-of-art environment.
- To create opportunities for students and faculty members to acquire professional knowledge and develop managerial, entrepreneurial, and social attitudes with highly ethical and moral values.
- To satisfy the ever-changing needs of the nation with respect to evolution and absorption of sustainable and environment-friendly technologies for effective creation of knowledge-based society in the global era.

Vision of the Department

To continually improve upon the teaching-learning processes and research with a goal to develop quality technical manpower with sound academic and practical experience, who can respond to challenges and changes happening dynamically in Computer Science and Engineering.

Mission of the Department

- To inspire the students to work with the latest tools and to make them industry ready.
- To impart research-based technical knowledge.
- To groom the department as a learning centre to inculcate advanced technologies in Computer Science and Engineering with social and environmental awareness.

I.II Program Outcome and Program-Specific Outcome

Program Outcome (PO)

- **PO1:** Engineering Knowledge- Apply the knowledge of mathematics, science, engineering fundamentals and engineering specialization to the solution of complex engineering problems.
- **PO2:** Problem Analysis- Identify, formulate, review research literature, and analyse complex engineering problems reaching substantiated conclusions using first principles of mathematics, natural sciences, and engineering science.
- **PO3:** Design & Development of Solutions- Design solutions for complex engineering problems and design system components, analysis and interpretation of data, and synthesis of the information to provide valid conclusions.
- **PO4:** Conduct Investigations of Complex Problems- Use research-based knowledge and research methods including design of experiments, analysis and interpretation of data, and synthesis of the information to provide valid conclusions.
- **PO5:** Modern Tool Usage- Create, select, and apply appropriate techniques, resources, and modern engineering and IT tools including prediction and modelling to complex engineering activities with an understanding of the limitations.
- **PO6:** The Engineer and Society- Apply reasoning informed by the contextual knowledge to assess societal, health, safety, legal and cultural issues and the consequent responsibilities relevant to the professional engineering practice.
- **PO7:** Environment and Sustainability- Understand the impact of professional engineering solutions in social and environmental context and demonstrate the knowledge of, and need for sustainable development.
- **PO8:** Ethics- Apply ethical principles and commit to professional ethics and responsibilities and norm of engineering practice.
- **PO9:** Individual and Team Work- Function effectively as an individual and as a member or leader in diverse teams and in multi-disciplinary settings.
- **PO10:** Communication- Communicate effectively on complex engineering activities with the engineering community and with society at large, such as, being able to comprehend and write effective reports and design documentation, make effective presentations and give and receive clear instruction.

- **PO11:** Project Management and Finance- Demonstrate knowledge and understanding of the engineering and management principles and apply these to one's own work, as a member and leader in a team to manage projects and in multi-disciplinary environments.
- **PO12:** Life-long Learning- Recognize the need for, and have the preparation and ability to engage in independent and life-long learning in the broadest context of technological change.

Program Specific Outcome (PSO)

- **PSO1:** Programming skills- Apply fundamental knowledge and programming aptitude to identify, design and solve real-life problems.
- **PSO2:** Professional skills- Students shall understand, analyze and develop software solutions to meet the requirements of industry and society.
- **PSO3:** Competency- Students will be competent for competitive examinations for employment, higher studies and research.

I.III PO and PSO mapping with justification

Fake News Detection Using Pattern Recognition	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7	PO 8	PO 9	PO 10	PO 11	PO 12	PS O 1	PS O 2	PS O 3
	2	3	3	3	-	3	-	3	2	3	2	3	3	3	2

Justification

- **PO 1:** The project uses the application of supervised learning algorithms hence statistics are being used. Hence the knowledge of mathematics and engineering has been applied.
- **PO 2:** Literature survey has been done to understand the performance of different supervised learning algorithms on the same dataset.
- **PO 3:** This project provides a solution to identify whether a news article is fake or not. An algorithm has been designed to solve the problem.

- **PO 4:** During the literature survey, we have gathered different research-based knowledge which has been implemented in this project.
- **PO 6:** This project will help society to find whether the news is fake or not as a result it will help society not to get influenced by fake news.
- **PO 8:** We shall follow professional ethics and not submit the work of other individuals in the project. We shall develop our own project by fair means and credit authors wherever required.
- **PO 9:** The project has required the team members to function effectively as individuals as well as communicate and coordinate among themselves to ensure a smooth flow of work.
- **PO 10:** Team discussion at different stages has been greatly beneficial to design this project. Good communication has been necessary to avoid any conflicts or confusion during the course of the project.
- **PO 11:** In this project, we have broken down the project into project management five stages which helped in the easy continuation of the problem.
- **PO 12:** Identifying fake news will always be needed by the society and this project has a scope to develop in the future.
- **PSO 1:** Programming knowledge has been required in this project to implement machine learning algorithms to obtain the desired result.
- **PSO 2:** The solution provided by the project will meet the requirements of better and more accurate prediction of fake news detection.
- **PSO 3:** Further research work can be done on this particular topic.

Abstract

Most smartphone users prefer to read the news via social media over the internet. The news websites are publishing the news and provide the source of authentication. The question is how to authenticate the news and articles which are circulated among social media like WhatsApp groups, Facebook Pages, Twitter, and other microblogs & social networking sites. It is harmful to society to believe in rumors and pretend to be news. The need for an hour is to stop the rumors, especially in developing countries like India, and focus on the correct, authenticated news articles. This project demonstrates a model and methodology for fake news detection. With the help of Machine learning, the news is determined whether the news is real or fake. The proposed models are working well and define the correctness of the results with 97 % of accuracy or above.

1. Introduction

The advent of the World Wide Web and the rapid adoption of social media platforms paved the way for information dissemination that has never been witnessed in human history before. Besides other use cases, news outlets benefitted from the widespread use of social media platforms by providing updated news in near real time to its subscribers. The news media evolved from newspapers, tabloids, and magazines to a digital form such as online news platforms, blogs, social media feeds, and other digital media formats. It became easier for consumers to acquire the latest news at their fingertips. Facebook referrals account for 70% of traffic to news websites. These social media platforms in their current state are extremely powerful and useful for their ability to allow users to discuss and share ideas and debate over issues such as democracy, education, politics and health. However, such platforms are also used with a negative perspective by certain entities creating biased opinions, manipulating mindsets, and spreading satire or absurdity. The phenomenon is commonly known as fake news. Fake News contains misleading information that could be checked. Here we are using different techniques and algorithms to detect if a news is fake or real.

1.1 Problem Statement

There is a lot of fake news on social media. The problem here is to find whether a given news article is real news or it is fake news with the help of a dataset.

1.2 Objective

The objective here is to find out if a news article is fake or real using some supervised learning algorithms. Here we are using two algorithms. One is logical regression and the other one is by using tensorflow.

1.3 Literature Survey

In[1] There are four types of models that have been applied here. Two different types of pre-processing are done on the dataset. In one pre-processing step count, a vectorizer is used. A simple way to both tokenize a collection of text documents and build a vocabulary of known words is provided by it. New documents using that vocabulary are encoded by it. In another pre-processing technique, TFIDF vectorizer is used that transforms text to feature vectors that can be used as input to estimator vocabulary that converts each token (word) to feature index in the matrix, each unique token gets a feature index.

Two different types of training models are used to train these two different pre-processed dataset, one is Naive Bayes Classifier and another one is Passive-Aggressive Classifier. A Naive Bayes classifier is a probabilistic machine learning model that's used for classification tasks. The crux of the classifier is based on the Bayes theorem. The passive-aggressive algorithms are a family of algorithms for large-scale learning. They are similar to the Perceptron in that they do not require a learning rate. The Naive Bayes classifier combined with the count vectorizer has given an accuracy level of 87%. The Naive Bayes classifier combined with the TFIDF vectorizer has given an accuracy level of 80%. The passive-aggressive classifier combined with the count vectorizer has given an accuracy level of 89%. The passive-aggressive classifier combined with the TFIDF vectorizer has given an accuracy level of 92%.

In[2] The process of making the model is done in six stages. At first, the data management phase is done where the dataset is collected and all the pre-procession is done. Term Frequency-Inverse Document Frequency (TF-IDF) is used to convert all the sentences of articles into a structured format. TFIDF works by proportionally increasing the number of times a word appears in the document but is counterbalanced by the number of documents in which it is present. Words that are very common are not given high priority. In the second stage, the model is trained with the help of logistic regression on the pre-processed data. Logistic regression estimates the probability of an event occurring, based on a given dataset of independent variables. The outcome is a probability, so the dependent variable is bounded between 0 and 1. Logistic regression uses a sigmoid function to convert the score into probability. The trained model is then saved using a .sav file. In the third stage, the model is evaluated by seeing the accuracy rate on test data and F1 score. In the fifth stage, the model has been integrated with the web portal to allow entry of news contents or news URL. The saved model is passed through API service which acts as a web service to the portal. In the last stage, the model is applied to new or fresh data and outcomes are monitored. The model gives a training accuracy of 99.2% and testing accuracy of 97% according to the report.

In[3] At first, two datasets, i.e. Fake and True are cleaned. Empty rows and the rows with some other errors are removed. The extra columns are also removed from the datasets. Then a class level column is added inside the datasets. After that these datasets are merged together. Then all the contents of the merged dataset is converted into lowercase characters. All special characters are also removed from the dataset. The 'title' and 'text' fields are merged together into a single column. After that the data is being prepared and transformed into a context which the machine can understand. For that the data are transformed into a list of vectors. Then the texts are tokenized by turning it into a list of sequences. Each word inside the data has its own sequence. The most frequent sequences are calculated. They are removed from the data. Then LSTM is used which remembers all the past knowledge that the network has seen so

far. It also can forget irrelevant data. After that, the sequence vector of the news content is being trained using fed with sequence vectors of the news content. After training and fitting the model we got an accuracy of more than 98% for a dataset having 22000 news data. In order to test our data the text news must be converted into sequences and after that, the model will predict the news.

In[4] According to the paper at first, all the pre-processing is done on the dataset. Tokenizing and stemming are the main pre-processing steps. Stemming is the process of reducing inflected words to their word stem, base, or root form—generally, a written word form and tokenization is the process of replacing sensitive data with unique identification symbols that retain all the essential information about the data without compromising its security. After all the pre-processing feature selection is done with the help of the TFIDF vectorizer or count vectorizer. Next, the model is trained by logistic regression. It is a machine learning classification algorithm that is used to predict the probability of a categorical dependent variable. In logistic regression, the dependent variable is a binary variable that contains data coded as 1 or 0. The accuracy given by this model is 72% according to the paper.

In[5] According to the paper, the noises like ids, dots, commas, quotations, unwanted columns are removed from the dataset. Then using POS the dataset is being turned into tokens and statistical values. Then unigram and bigram features are extracted using the TFIDF vectorizer function of python sklearn. Then the dataset is divided into two parts: 70% for training and 30% for testing. Then six classifier algorithms are used to produce a classification model. The algorithms are XGboost, Random Forests, Naive Bayes, K-Nearest-Neighbors, Decision Tree and SVM. Then the precision of those models are calculated using the test portion of the dataset and confusion matrix is being produced. Then the accuracy of those models are compared. The model with the highest accuracy is chosen. The highest accuracy can go up to 92%.

DataSet: In **In[1]** one dataset has been used having attributes ‘title’, ‘text’, ‘publishing date’, and ‘label’. The dataset has been taken from Kaggle. The size of the dataset is 25 MB having 20000 rows of data. In **In[2]** one dataset has been used having attributes ‘title’, ‘text’, ‘publishing date’, ‘author’, and ‘label’. The dataset has been taken from Kaggle. The size of the dataset is 96 MB having at least 25000 rows of data. In **In[3]** two datasets have been used. One for the fake news and another for the real news. Both datasets contain ‘title’, ‘text’, ‘subject’, and ‘date’ as attributes. The size of the real news dataset is 61 MB having at least 20000 rows of data and the size of the fake news dataset is 59 MB having at least 19000 rows of data. In **In[4]** three datasets are used. One dataset named as train.csv is used for training purposes which has attributes ‘title’, ‘text’, ‘author’, and ‘label’. The size of the dataset is 96 MB having 25000

rows of data. Another dataset named as test.csv is used for getting accuracy level of the model. The dataset has attributes named 'title', 'text', and 'author'. The size of the dataset is 25 Mb having 5000 rows of data. Another dataset named submit.csv has the mapping of the labels to the dataset test.csv. The size of this file is less than 15 MB.

1.4 Brief Discussion on Problem

The main purpose of this project is to find if a news article is fake or real. Fake news can provide us with misinformation and misconception about a certain topic. So it is important to identify if the news is fake or not. In order to do that, in the beginning we are cleaning and processing the dataset to perform error free training and testing. Then based on the news article or text article authenticity of the news will be determined.

1.5 Planning

There are different methodologies that could be implemented to build fake news detection software. In this case, we divided the project into six phases. The six phases are literature review, data management, model training, model evaluation, model integration, and model deployment.

Literature Review

In this phase various research papers, journals, and articles on fake news detection are being studied thoroughly. This phase mainly deals with studying the pre-implemented works that have been done on fake news detection and studying the tools that are required for implementing fake news detection software.

Data Management

In this phase different news datasets have been collected from different online sources and data pre-processing has been done. The data pre-processing mainly includes removing null data rows from the dataset and removing all kinds of noisy data.

Model Training

In this phase the pre-processed data have been trained with the algorithm related to the problem. In this case, the algorithms used are the logistic regression algorithm, passive-aggressive classifier algorithm and TensorFlow algorithm.

Model Evaluation

In this phase the main purpose is to check the performance level of different algorithms that have been applied on the same dataset. The accuracy is mainly measured by F1 score, precision, recall and accuracy rate which is based on the confusion matrix report. In this case, the logistic regression model gave an accuracy of 97% and TensorFlow model gave an accuracy of 98%.

Model Integration

In this phase, the model is integrated into a web portal or android application to allow entry of news content. Till now this phase has not been started in this project.

Model Deployment

In the last stage, the model is applied to new or fresh data and outcomes are monitored. In this phase, every result that is created is taken into close observation and recorded. The model still can be improved if necessary.

Project Timeline

A project timeline is a visual list of tasks or activities placed in chronological order, which lets project managers view the entirety of the project plan in one place. A project timeline provides an in-depth overview of the entire project from start to finish.

For our project we have used a Gantt chart to illustrate the activities done so far and the future activities needed to be done for the project. The timeline is divided into different phases with all the activities allocated to each phase corresponding to the time required.

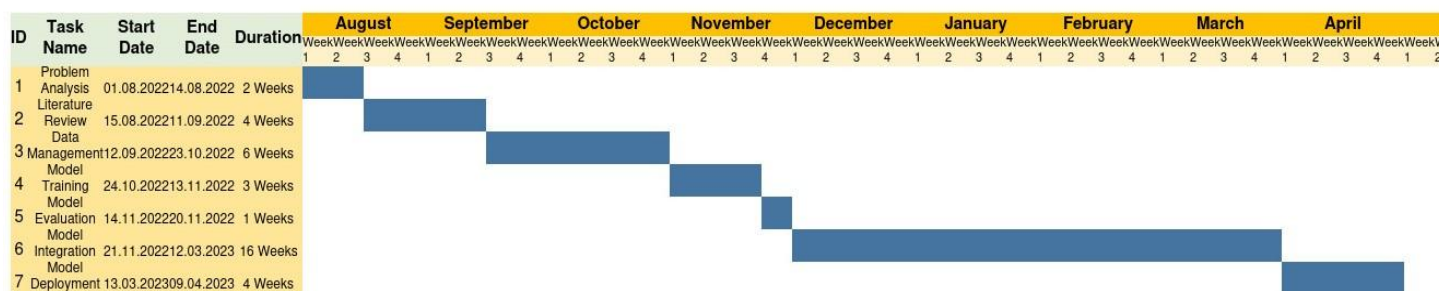


Fig.1 Gantt Chart to monitor Activity and Planning

1.6 Organization

The project report is organized as follows. Chapter 2 contains the basic concepts used and problem analysis. Chapter 2 is mainly divided into three sub-chapters, the first sub-chapter 2.1 tells a brief description of the problem, the sub-chapter 2.2 tells about the basic concepts used in this project to solve the problem and the sub-chapter 2.3 tells about the algorithms used to solve the problem. After that, chapter 3 contains the conclusion of the project. It tells about what is the actual output we got after applying the chapter 2.2 algorithms. Chapter 4 contains all the references that we used till now to solve the problem partially.

2. Concepts And Problem Analysis

2.1 Problem Analysis

The term fake news means news articles that are intentionally and verifiably false designed to manipulate people's perceptions of real facts, events, and statements. It's about information presented as news that is known by its promoter to be false based on facts that are demonstrably incorrect or statements or events that verifiably did not happen. Fake news provides us with misinformation, and disinformation which can mislead people into a wrong path. The project tries to identify which news article is fake or not and help the society to remove misconceptions about the news.

2.2 Concepts

The main concept that has been used in solving the problem of fake news detection is logistic regression, recurrent neural network, and passive-aggressive classifier. While applying these concepts we have also applied different python modules. Here is a brief description of every concept that has been used in this project.

Logistic Regression

Logistic regression estimates the probability of an event occurring, such as voted or didn't vote, based on a given dataset of independent variables. Since the outcome is a probability, the dependent variable is bounded between 0 and 1. In logistic regression, the value is converted into a probability with the help of the sigmoid function. The class label which has the higher probability is selected as the predicted class.

Recurrent Neural Network

Recurrent Neural Network (RNN) is a type of Neural Network where the output from the previous step is fed as input to the current step. In traditional neural networks, all the inputs and outputs are independent of each other, but in cases like when it is required to predict the next word of a sentence, the previous words are required and hence there is a need to remember the previous words. Thus RNN came into existence, which solved this issue with the help of a Hidden Layer. The main and most important feature of RNN is the Hidden state, which remembers some information about a sequence. An RNN remembers each and every information through time. It is useful in time series prediction only because of the feature to remember previous inputs as well. This is called Long Short Term Memory(LSTM).

Passive Aggressive Classifier

Passive Aggressive Classifiers belong to the category of online learning algorithms. In online learning, a machine learning model is trained and deployed in production in a way that continues to learn as new data sets arrive. So we can say that an algorithm like Passive Aggressive Classifier is best for systems that receive data in a continuous stream. Passive-Aggressive algorithms are generally used for large-scale learning. Passive-Aggressive algorithms are somewhat similar to a Perceptron model, they do not require a learning rate.

Passive-Aggressive Algorithms Work as below:

- **Passive:** If the prediction is correct, keep the model and do not make any changes. i.e., the data in the example is not enough to cause any changes in the model.
- **Aggressive:** If the prediction is incorrect, make changes to the model. i.e., some change to the model may correct it.

TFIDF Vectorization

TFIDF works by proportionally increasing the number of times a word appears in the document but is counterbalanced by the number of documents in which it is present. Hence, words like 'this', 'are' etc., that are commonly present in all the documents are not given a very high rank. However, a word that is present too many times in a few of the documents will be given a higher rank as it might be indicative of the context of the document.

Term Frequency: Term frequency is defined as the number of times a word (i) appears in a document (j) divided by the total number of words in the document.

Inverse Document Frequency: Inverse document frequency refers to the log of the total number of documents divided by the number of documents that contain the word. The logarithm is added to dampen the importance of a very high value of IDF.

TFIDF is computed by multiplying the term frequency with the inverse document frequency.

Python Module And Library Used

- **Scikit-learn(sklearn) in Python**
 - The sklearn library contains a lot of efficient tools for machine learning and statistical modeling including classification, regression, clustering and dimensionality reduction.
- **NumPy in Python**
 - NumPy is a general-purpose array-processing package. It provides a high-performance multidimensional array object, and tools for working with these arrays. It is the fundamental package for scientific computing with Python. It is open-source software.
- **Pandas in Python**
 - Pandas is an open source library in Python. It provides ready to use high-performance data structures and data analysis tools.
- **NLTK (Natural Language Toolkit)**
 - Natural language processing (NLP) is a field that focuses on making natural human language usable by computer programs. NLTK, or Natural Language Toolkit, is a Python package that can be used for NLP.
- **Matplotlib in Python**
 - Matplotlib is a python library used to create 2D graphs and plots by using python scripts.
- **Gensim in Python**
 - Gensim is a Python library for topic modelling, document indexing and similarity retrieval with large corpora.

2.3 Proposed Algorithms

Proposed Algorithm For Fake News Detection using RNN Model

Step 1: At first step two datasets i.e. ‘Fake’ and ‘Real’ are read.

Step 2: Then they are divided according to their publisher names. Rows with no publishers are set as “unknown.”.

Step 3: Then the 'title' and 'text' column of each dataset is merged together into a single 'text' column. And a 'class' column is added(1 for real and 0 for fake).

Step 4: The 'text' and 'class' columns of two datasets are merged together into a single one.

Step 5: The characters of the 'text' part are converted into lowercase and all the special characters are removed.

Step 6: Each word inside the 'text' column is turned into a list and the words are converted into vectors.

Step 7: The words are then turned into the form of sequences.

Step 8: Then the most frequent words are padded into default value 0.

Step 9: Weight matrix is created using the vectors.

Step 10: A model is then created using the sequential function and the weight matrix is added.

Step 11: LSTM and Dense layers are added into the model.

Step 12: 70% of the data is used for training the model and the remaining 30% is used for testing.

Step 13: Then the accuracy level is calculated.

Step 14: Then a text data is taken as input. The text data is converted into sequences and padded. Then the result is predicted.

Proposed Algorithm For Fake News Detection using Logistic Regression Model

Step 1: The training dataset is read having columns author, title, text, and labels.

Step 2: All the missing data are replaced with "empty string".

Step 3: A new column is created combining the author and title of the news and naming the column as authorAndTitle.

Step 4: Removing all kinds of punctuation from the data in column authorAndTitle and also changing all the alphabets into lowercase.

Step 5: Removing all kinds of stopwords and applying stemming to the column authorAndTitle.

Step 6: The data is then converted into numerical form and then converted to feature form.

Step 7: 70% of the data from column authorAndTitle is selected randomly which is used to train the logistic regression model.

Step 8: Rest of the 30% of the data is taken for testing purposes and the accuracy of the model is calculated.

In this project, the optimal model will be integrated with a web application or android application, so that it can be used by common people to understand how much to rely on a given news article.

3. Conclusion

In this project till now we implemented logistic regression and recurrent neural network (RNN) to find the authenticity of any news article. In our training dataset, the logistic regression model has an accuracy of 97% and the model using tensorflow and RNN has an accuracy level of 98%. But the model using tensorflow and RNN takes much time as compared to logistic regression.

4. Reference

4.1 Book Reference

[1] Kudari, Jayashree M., V. Varsha, B. G. Monica, and R. Archana. "Fake news detection using Passive aggressive and TF-IDF vectorizer." 2020 International Research Journal of Engineering and Technology (2020).

[2] Muhammad Syahmi Mokhtar, Yusmadi Yah Jusoh, NoviaAdmodisastro, NorainiChe Pa, AmruYusrinAmruddin, "Fakebuster: Fake News Detection System Using Logistic Regression Technique In Machine Learning ", International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249-8958 (Online), Volume-9 Issue-1, October, 2019.

[3] Sheng How Kong, Li Mei Tan, Keng Hoon Gan, Nur Hana Samsudin, "Fake News Detection using Deep Learning", IEEE Xplore, June 06,2020.

[4] Fathima Nada, Bariya Firdous Khan, Aroofa Maryam, Nooruz-Zuha, Zameer Ahmed, "FAKE NEWS DETECTION USING LOGISTIC REGRESSION", International Research Journal of Engineering and Technology (IRJET), May 2019

[5] Z Khanam , B N Alwasel , H Sirafi and M Rashid, “Fake News Detection Using Machine Learning Approaches”, IOP Conference Series: Materials Science and Engineering

[6] Dr.M. RAJESWARI1, A.SRINIRANJANI, “FAKE NEWS DETECTION USING LOGISTIC REGRESSION ALGORITHM WITH MACHINE LEARNING”, 2022 JETIR June 2022, Volume 9, Issue 6

[7] Dr.Rajiv Chopra, “Machine Learning”, 2nd edition, Learning Algorithms, Page No: 141, Khanna Book Publishing Co LTD, 2022

4.2 Web Reference

1. <https://thecleverprogrammer.com/2021/02/10/passive-aggressive-classifier-in-machine-learning/>
2. <https://towardsdatascience.com/covid-fake-news-detection-with-a-very-simple-logistic-regression-34c63502e33b>
3. <https://stackoverflow.com/>
4. <https://askubuntu.com/>
5. <https://scikit-learn.org/stable/>
6. <https://www.w3schools.com/python/pandas/default.asp>
7. https://www.w3schools.com/python/numpy/numpy_intro.asp