Fake News Classifier Using ML

Submitted in partial fulfillment of the requirements

of the degree of

Bachelor of Technology

by

Rutik Jayram Torambe

(Reg. No.: -2020BEC014)

Supervisor (s):

Dr. Usha Kamble

Department of Electronics and Telecommunication Engineering,
Shri Guru Gobind Singhji Institute of Engineering & Technology,
Vishnupuri, Nanded, Maharashtra, India, 431606.
2023-24

DECLARATION

I declare that this written submission represents my ideas in my own words and where others' ideas or words have been included, I have adequately cited and referenced the original sources. I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission. I 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.

Rutik Jayram Torambe
Reg. No.:2020BEC014

Date: _____

CERTIFICATE

*This is to certify that the report entitled "***Fake News Classifier Using ML"** *being submitted by* **Rutik Jayram Torambe** (2020BEC014) *to* **Shri Guru Gobind Singhji Institute of Engineering and Technology, Vishnupuri, Nanded (M.S.), India,** *as partial fulfillment for the award of the degree of* **Bachelor of Technology** *in* **Electronics and Telecommunication Engineering**, *is a record of bonafide work carried out by him under our supervision and guidance. The matter contained in this report has not been submitted to any other university for the award of any degree or diploma.*

Dr. Usha Kamble	**Dr. M.V.Bhalerao**
Supervisor	**H.O.D.**
Elect. and Telecom. Engg. Dept.	**Elect. and Telecom. Engg. Dept.**

APPROVAL SHEET

This report entitled "Fake News Classifier Using ML " by Rutik Jayram Torambe is approved for the degree of Bachelor of Technology.

Examiners

_____

_____

_____

Supervisor (s)

_____

_____

_____

Date: _____

Place: _____

ACKNOWLEDGEMENTS

The joy of accomplishments is never truly complete without acknowledging those who helped us get where we are going. Their unwavering guidance and support are the cornerstone of our success.

I am especially thankful to the honorable Chairman, Mr. Sunil Raithatt for providing us with exceptional services and well-equipped laboratories. These resources have proven invaluable in our efforts.

Our esteemed Director, Dr. I am also very grateful to Manesh B. Kokare. His constant support and encouragement added great energy to the whole project.

I met with the head of the department, Dr. M.V. Bhalerao, for unwavering encouragement and invaluable suggestions during my final year project.

Timely instructions and clear deadlines are critical to getting the job done. In this regard, I would like to thank Drs. Usha Kamble provide us guidance.

Finally, we thank the entire Department of Electronics and Telecommunications Engineering, teaching and non-teaching staff, for their cooperation and support throughout the project.

Rutik Jayram Torambe

ABSTRACT

The rapid growth of online social networks in recent years has led to an increase in the number of fake news driven by commercial and political motives, posing serious challenges to the trust and integrity of the Internet. This study examines various theories, techniques and policies, aimed at early detection of fake news accounts, their creators and related content in online social networks, and to measure their effectiveness Especially information dissemination accuracy has emerged as an immediate concern in social media platforms. However, the scope of web-based information provides the task of identifying, testing, and correcting such misleading content, which is often referred to as "fake news" In this paper, we propose a methodology for fake news detection propose and analyze its use on Facebook in particular It is located on an online social media platform. Using the Naive Bayes classification model, our method predicts whether a Facebook post is true or fabricated. Furthermore, we discuss strategies to increase the accuracy of these predictions. Our findings highlight the feasibility of using machine learning to address the challenge of fake news detection.

Keywords: Fake news, social media, Detection algorithms, Information trustworthiness, Machine learning, Online disinformation.

CONTENTS

CHAPTER 1

INTRODUCTION

Fake news is like a big wave of chaos these days. It started out as an obvious lie for the money, but now it’s solid. Ironies, misleading social media posts, and even sneaky government messages make it hard to tell the truth. This has shaken people’s faith in regularity.

In the past, they were clearly "fake news": content that was all made up to make money online. Now people are throwing the word around for whatever fact they don’t like. This "disinformation" has been a big issue in recent elections, showing how disastrous mixed information can be.

To combat it, some want devices to clean up fake news, especially on Facebook. But this "fake news" is hard to beat. You can’t take one side of politics over another; you have to know the reliable news sources. This can be difficult because some fancy websites spread lies, while smaller websites tell the truth.

The first step in dealing with this mess is figuring out what "fake news" really means. Is a funny story with misleading information technically fake? What about an opinion with some wrong facts? These are questions that we need to answer in order to develop better detection systems.

Social media makes things more complicated. People love sharing on these platforms, which allows inaccurate but interesting information to spread quickly. Friends share unchecked things, and emotional issues travel faster than balanced conversations. This "echo chamber" encourages people to believe what they already think and avoids seeing the opinions of others.

So, what can we do? Learning to judge matters for yourself is key. It means analysing the author, looking for evidence in real media, and debating topics.

Social media platforms also play a role. They can slow the spread of blatant lies while protecting free speech

1.1 Motivation

Fighting fake news involves training machines. We do this with labeled data, tagging news as "real" or "fake." But this data needs cleaning first. We get rid of missing info and turn text into numbers machines can understand. This is like prepping food for a machine's learning process. After cleaning, tools like Matplotlib help us visualize the data with charts. These charts help us see patterns and make the machine learning even better.

1.2 Objective

This project aims to fight fake news! It will examine the spread and destructiveness of fake news. The project uses unique datasets (a collection of fake and true stories) to train machines to recognize fake news. One piece of artificial intelligence, machine learning, will be used to find patterns in data. This will help machines learn to distinguish between real and fake news.

The project also wants to identify machine learning models that work best for different media. More data is better, as it allows machines to run more patterns and better detect fake news. Ultimately this will help people trust the media more and make better decisions based on real information

1.3 Overview of Project

The advanced Internet brings us global news instantly, but allows misinformation and fake news to spread. The frequent appearance of these misleading stories on social media can be very credible, making it difficult for readers to discern the truth from the stories. Efforts are underway to combat this, including developing machine learning algorithms to detect fake news. But the constant use of new and fraudulent information is a race against time, as the fake news creators are constantly changing their tactics.

CHAPTER 2

LITERATURE SURVEY

Social media can be a breeding ground for misinformation, where rumors and misinformation can spread like wildfire. Unfortunately, some people deliberately use deceptive tactics to trick others into making these false statements. These tactics can be covert, such as creating fake statistics that appear credible or using catchy headlines that promise shocking news. The more people share this information, the faster it spreads, making it harder to separate fact from fiction. While there are tools available to help us detect fake news, it remains particularly challenging because of these deceptive tactics.

2.1 Media Rich Fake News Detection: A Survey

Fake news circulates in the wild west of social media, tempting the unsuspecting with catchy headlines like shiny gold in a fool's jar. This new study aims to be a sheriff on this digital frontier, developing a tool to identify these fraudsters with laser-like focus. But instead of a gang and six-shooters, this sheriff is using a powerful algorithm, fueled only by the content and headlines of questionable cases and their results? Amazing 93.3% accuracy in phonies-emission! But is this the ultimate weapon? Can it also reveal the ways in which clickbait headlines are used, even if the content itself isn't composed at all? What about where the story comes from, or how do people react to it? This study is a promising first step, but there are many mysteries waiting to be unearthed in the fight against misinformation. This new tool may be a six-shooter cleaning up the social media saloon, but it could be an entire arsenal waiting to be built.

2.1.1 Weakly Supervised Learning for Fake News Detection on Twitter

This study addresses the challenge of detecting fake news on social media, especially Twitter. Although it may seem like a simple classification (true or false) problem, it is difficult to collect sufficient training data. Handwritten tweets are expensive and time-consuming.

The authors propose a less conservative approach. They collect a huge amount of tweets data themselves and label it based on the source (trusted or unreliable). This is a noisy approach because not all tweets from bad sources are fake, and vice versa.

Despite the incorrect data, the researchers were able to train the classifier with an impressive F1 score of 0.9 for false positives. This means that even with noisy data, this method can be effective in detecting fake news on social media.

2.2 Fake News Detection in Social Media

The Internet age may have transformed communication, but it has also opened the gates to a dangerous phenomenon: fake news. Defined as intentional content designed to mislead readers, fake news can be spread for a variety of reasons from increased web traffic to influencing public opinion social media platforms have become a haven for misinformation if such an emphasis on user-generated content and rapid sharing of information

This rise in fake news poses a serious challenge. Often covered with quirky headlines and eye-catching graphics, this fabricated content can mislead users into believing lies. The consequences can be far-reaching, affecting everything from personal decision-making to social and political life.

Researchers are actively developing solutions to combat the spread of fake news. One approach focuses on content analysis, carefully examining the text and structure of the newspaper itself. This method thoroughly analyzes post titles and content with an accuracy of up to 93.3% to identify potential red flags of fraud.

Another approach goes beyond content, and delves into the social context of how information is distributed. This method analyzes how users engage with posts – who likes them, who shares them, and how quickly they get traffic. By combining content analysis with social network data, the researchers developed a strategy to achieve 81.7% accuracy in the Facebook Messenger chatbot. This suggests that the behaviors used can be valuable clues in detecting fake news.

The ease and speed with which information travels through social media makes it difficult for users to assess credibility in real time. Automatic recognition systems using content and society. The

Internet age may have transformed communication, but it has also opened the floodgates to a dangerous phenomenon: fake news. Defined as intentional content designed to mislead readers, fake news can be spread for a variety of reasons from increased web traffic to influencing public opinion social media platforms have become a haven for misinformation if such an emphasis on user-generated content and rapid sharing of information

This rise in fake news poses a serious challenge. Often covered with quirky headlines and eye-catching graphics, this fabricated content can mislead users into believing lies. The consequences can be far-reaching, affecting everything from personal decision-making to social and political life.

Researchers are actively developing solutions to combat the spread of fake news. One approach focuses on content analysis, carefully examining the text and structure of the newspaper itself. This method thoroughly analyzes post titles and content with an accuracy of up to 93.3% to identify potential red flags of fraud.

Another approach goes beyond content, and delves into the social context of how information is distributed. This method analyzes how users engage with posts – who likes them, who shares them, and how quickly they get traffic. By combining content analysis with social network data, the researchers developed a strategy to achieve 81.7% accuracy in the Facebook Messenger chatbot. This suggests that the behaviors used can be valuable clues in detecting fake news.

The ease and speed with which information travels through social media makes it difficult for users to assess credibility in real time. Automatic recognition systems using content and society

2.3 The Spread of Fake News by Social Bots

Fake news has become a major concern, especially during elections. To understand the spread, researchers analyzed millions of messages on Twitter during the 2016 U.S. election. the election. The findings revealed a surprising culprit: automated accounting, or "bots." These bots act like fakers, but are programmed to spread information quickly. Influencers are often targeted on Twitter, giving fake news a wider audience. Unfortunately, these bots mislead regular people, sharing much of the fake news without their knowledge. This study suggests that tackling the fake

news problem may require eliminating these dangerous bots, potentially reducing the spread of misinformation on the Internet

2.4 Misleading Online Content

Internet content can be a double-edged sword. Notorious for their over-the-top headlines and scare tactics, clickbait articles take advantage of this by getting users to click. This information spreads misinformation and poses a serious threat.

This paper proposes the use of an automated clickbait detection system to address this issue. This program will scrutinize online resources to look beyond text alone. It would analyze the overall medium, and identify the manipulative strategies employed by clickbait writers.

These solutions are powered by two powerful tools in data science: big data and deep learning. Big data refers to the large amount of information collected by governments and businesses. Although these stories are large and unsophisticated, they are powerful. Imagine analyzing information about national security, business practices, or even medical research! Companies like Google use big data to understand user behavior and make informed decisions.

Deep learning, driven by the human brain, excels at analyzing unlabeled data (data that has not yet been classified). This makes it the perfect partner for big data by filtering unstructured information and revealing hidden patterns.

The researchers envision that deep learning will solve other big data challenges. One is the ever-changing nature of information, which is especially evident on social media. How can we analyze ever-changing information? Another challenge is "high-dimensional data," which combines text and images. There is a need to develop models to deal with such complexity The paper suggests future research directions. It is important to choose the most appropriate case studies, such as the availability of gold in a river. Researchers are also investigating how to adapt deep learning models

CHAPTER 3

PRE-REQUIREMENTS OF PROJECTS

Before diving into fake news identification programs using machine learning (ML), it is important to establish a solid foundation. This first phase, called project prerequisites, sets the stage for the entire project life cycle. This includes defining project objectives, understanding its scope, and aligning stakeholder expectations. By engaging stakeholders early and clarifying project objectives, project managers foster collaboration to ensure clear direction. These foundational steps guide subsequent activities, such as data collection, model development, and analysis. A well-defined project scope and purpose are critical for the successful ML model implementation of the fake news detection.

3.1 Hardware Requirements

Identify and list the essential hardware components necessary for the project's execution, ensuring compatibility with the fake news detection system's computational needs.

- ✓ CPU TYPE: Intel i3, i5, i7 or AMD
- ✓ RAM Size: Min 2 GB
- ✓ Hard Disk Capacity: Min 4 GB
- ✓ Graphic Card: Optional (Not Required)

This System can Work in any given system as Laptop, iPad, Phone or any System that possesses Internet.

3.2 Software Requirements

Specify the necessary software tools and platforms required to develop and deploy the fake news identification system, ensuring compatibility with the chosen ML algorithms and frameworks.

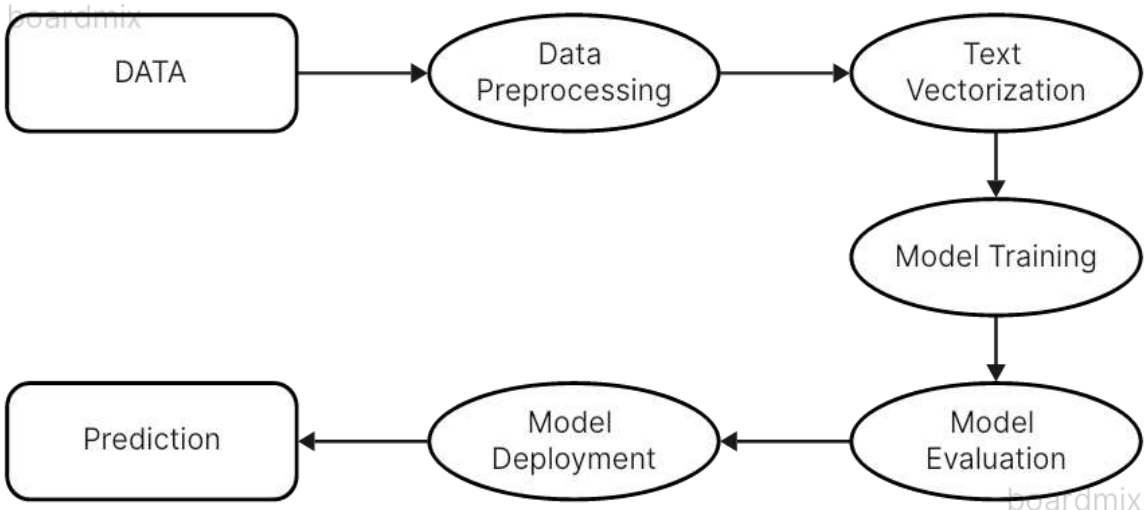
- ✓ Operating system: Windows / Linux / iOS
- ✓ IDE: Jupyter Notebook / PyCharm / Anaconda / VS Code

- ✓ Cloud IDE: Google Collab
- ✓ Programming Language: Python
- ✓ Python Libraries: scikit-learn, seaborn, pandas, matplotlib, wordcloud, pickle, flask
- ✓ Other: HTML, CSS, JS.

CHAPTER 4

PROPOSED METHOD

The proposed method aims to identify fake news by combining data preprocessing and ML techniques. By cleaning, tokenizing, and vectorizing text data, we prepare it for training with a Passive Aggressive Classifier. This model learns patterns in the data to classify news articles as fake or real. The implementation integrates this trained model into a Flask web application for real-time prediction.



4.1 Data Collection and Preprocessing

Obtaining a high-quality and diverse dataset is the foundation for building an effective machine-learning model. The data collection process involves sourcing relevant datasets, while preprocessing steps like cleaning, normalization, and text vectorization transform the raw data into a format suitable for model training. Careful consideration of dataset characteristics and appropriate preprocessing techniques is crucial for accurate model performance.

4.1.1 Data Collection

Obtaining a high-quality and diverse dataset is crucial for building an effective fake news detection system. In this case, the dataset used is a CSV file named 'news.csv,' which contains a collection of news articles along with their corresponding labels ('FAKE' or 'REAL'). This dataset was likely sourced from publicly available repositories or platforms that specialize in compiling and labeling news articles based on their authenticity.

When selecting a dataset for this task, several factors should be considered:

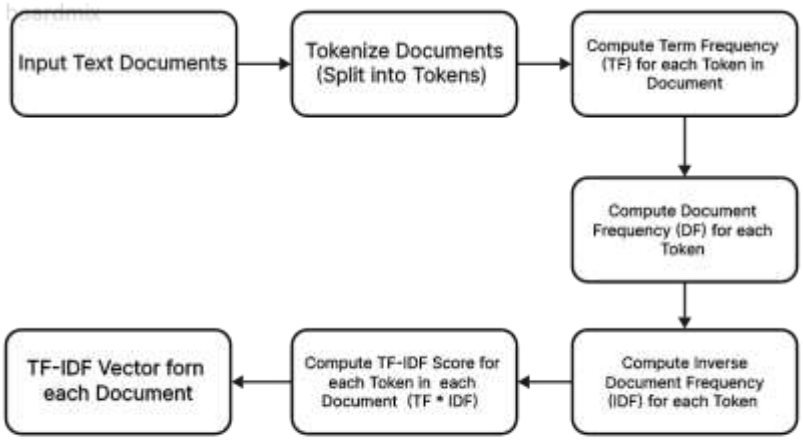
- **Dataset Size:** A larger dataset can potentially lead to a more generalizable model, as it exposes the model to a wider range of examples during training. However, it's important to strike a balance between dataset size and computational resources required for training.
- **Diversity of News Sources:** The dataset should include news articles from a diverse range of sources, encompassing different writing styles, viewpoints, and topics. This diversity helps the model learn to distinguish fake news from real news across various sources and contexts.
- **Labeling Criteria:** Understanding the criteria used to label news articles as "FAKE" or "REAL" is crucial. Some datasets may label satire or opinion pieces as fake news, while others may focus solely on intentionally misleading or fabricated articles. Aligning the labeling criteria with the desired definition of fake news is essential for interpreting the model's predictions accurately.
- **Data Quality:** The dataset should be carefully curated to ensure the accuracy of the labels and the relevance of the news articles. Noisy or irrelevant data can adversely affect the model's performance.

4.1.2 Data Preprocessing

Once the 'news.csv' dataset is acquired, it undergoes several preprocessing steps to clean and transform the data into a format suitable for machine learning algorithms:

- **Cleaning:** This step involves removing irrelevant characters, URLs, and extra spaces from the text data. Cleaning the data helps to reduce noise and improve the model's ability to learn meaningful patterns from the news articles.

- **Normalization:** Converting all text to lowercase ensures consistency and prevents the model from treating uppercase and lowercase versions of the same word as different features. This step is important for accurate feature extraction and representation.
- **Stop Word Removal:** Common words such as "the," "a," "an," "is," "and," etc., are eliminated from the text data. These stop words don't contribute significantly to the meaning of the text and can add noise to the data, making it computationally more efficient for the model to learn.
- **Text Vectorization:** The textual data is transformed into numerical representations using techniques like TF-IDF (Term Frequency-Inverse Document Frequency) vectorization. TF-IDF is a statistical measure that assigns higher weights to words that are frequent in a specific document but rare across the entire dataset. This approach helps to identify the keywords and phrases that are most characteristic of real or fake news articles. For example, words like "hoax," "conspiracy," "unbelievable," or "sensational" might receive higher weights in fake news articles, while terms like "official," "investigate," "confirmed," or "credible" could be more heavily weighted in real news articles.



TF-IDF Vectorization

- **Data Splitting:** After preprocessing, the dataset is typically split into training and testing sets. The training set is used to train the machine learning models, while the testing set is held out to evaluate the models' performance on unseen data.

By performing these preprocessing steps, the data is cleaned, transformed, and prepared for effective model training, ensuring that the machine learning algorithms can learn meaningful patterns and features from the news articles to distinguish between real and fake news accurately.

4.2 Model Training and Evaluation

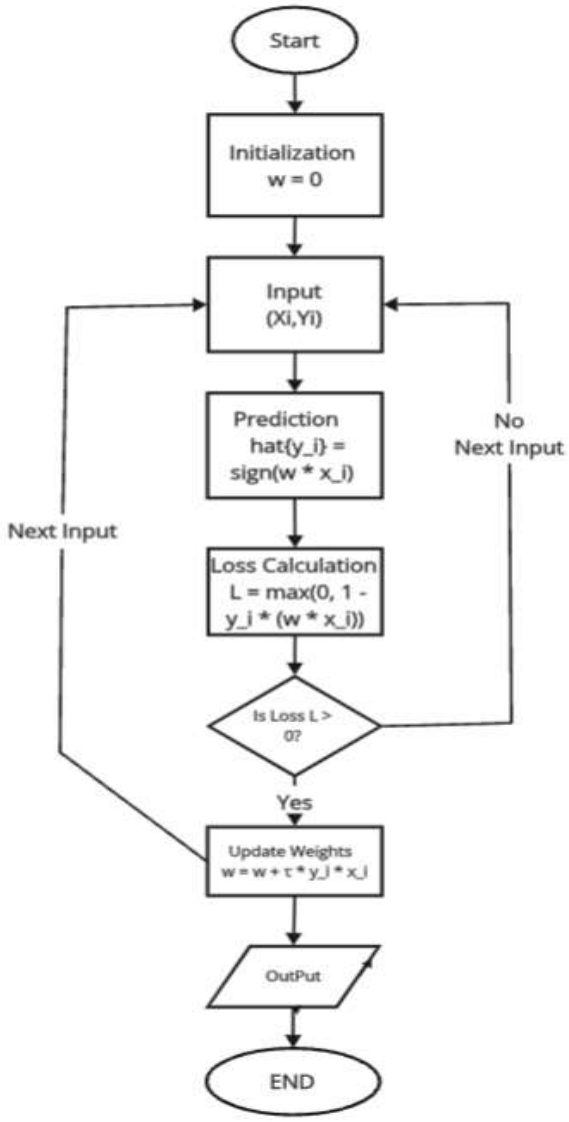
After the data has been preprocessed and transformed into a suitable format, the next step is to train and evaluate various machine learning models for the task of fake news detection. The code employs three different algorithms: Passive Aggressive Classifier, Random Forest Classifier, and Naive Bayes Classifier.

4.2.1Passive Aggressive Classifier

The Passive Aggressive Classifier is an online learning algorithm that belongs to the family of linear classifiers. It is designed to handle large-scale data scenarios and dynamic environments where data distributions may change over time. The key idea behind this algorithm is its adaptive behavior: if a prediction is correct, no changes are made to the model, but if a prediction is incorrect, the model aggressively updates its parameters to correct the mistake.

The code initializes a Passive Aggressive Classifier, fits it to the vectorized training data, and then makes predictions on the testing data. The performance of the model is evaluated using a classification report, which provides metrics such as precision, recall, and F1-score for each class ('FAKE' and 'REAL').

Accuracy of the model is 93.3%



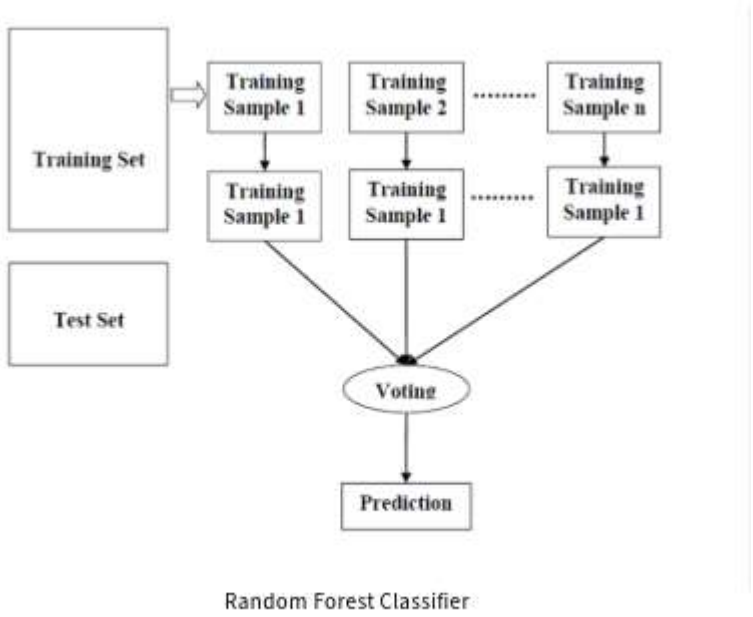
Passive Aggressive Classifier

4.2.2 Random Forest Classifier

The Random Forest Classifier is an ensemble learning technique that combines multiple decision trees to make predictions. It uses a technique called Bootstrap Aggregating (bagging) to create diverse subsets of data for each tree, reducing the risk of overfitting and improving the model's generalization ability.

During the training process, the code initializes a Random Forest Classifier and fits it to the vectorized data using a Count Vectorizer (another text vectorization technique). The model's performance is then evaluated on the testing data using a classification report.

Accuracy of the model is 90.2%



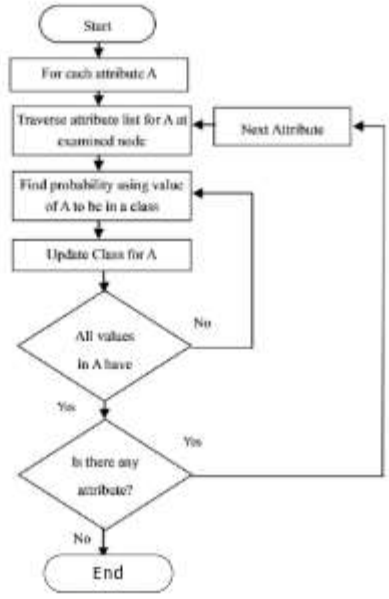
Random Forest Classifier

4.2.3 Naive Bayes Classifier

The Naive Bayes Classifier is a probabilistic classifier based on Bayes' Theorem, which calculates the probability of a hypothesis given the evidence. Despite its simplistic assumption of feature independence, it is widely used due to its simplicity and efficiency in machine learning tasks, particularly in text classification problems.

The code initializes a Multinomial Naive Bayes Classifier, trains it on the vectorized data using a Count Vectorizer, and evaluates its performance on the testing data using a classification report.

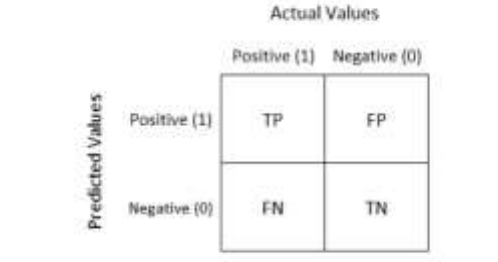
Accuracy of the model is 88.1%



Naive Bayes Classifier

4.2.4 Confusion Matrix

In addition to the classification reports, the code generates and plots a confusion matrix for the predictions made by the Naive Bayes Classifier. A confusion matrix is a table that summarizes the performance of a classification model by showing the number of true positives, true negatives, false positives, and false negatives. It provides a visual representation of the model's performance and helps identify potential biases or misclassifications.



Confusion Matrix [Image 2] (Image courtesy: My Photoshopped Collection)

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

By training and evaluating multiple algorithms, the code aims to find the most accurate and robust model for fake news detection. The performance metrics, such as precision, recall, and F1-score, can be used to compare the models and select the best-performing one for deployment or further fine-tuning. In our case the most accurate classification algorithm is Passive Aggressive Classifier with accuracy of 93%.

4.3 Web Application Development

To make the fake news detection system more accessible and user-friendly, a web application is developed using the Flask framework in Python. Flask is a lightweight and flexible web framework that allows for rapid development of web applications, making it an ideal choice for this project.

Flask Application Setup:

The code begins by importing the required libraries and modules, including Flask for creating the web application, scikit-learn for machine learning algorithms, pickle for loading the trained model, and pandas for data manipulation.

The Flask application is initialized using `app = Flask(__name__)`, which creates an instance of the Flask class. This instance will be used to define the routes and handle the application logic.

Loading TF-IDF Vectorizer and Trained Model:

The code then loads the TF-IDF vectorizer (`tfvect`) and the trained Passive Aggressive Classifier model (`loaded_model`) from the 'model.pkl' file using pickle. The TF-IDF vectorizer is responsible for transforming the user's input text into a numerical representation that the model can understand, while the trained model is used to make predictions on the vectorized data.

Fake News Detection Function:

The `fake_news_det` function is defined, which takes a news article as input, vectorizes it using the TF-IDF vectorizer, and predicts whether the news is 'FAKE' or 'REAL' using the loaded model.

Flask Routes:

The code defines two Flask routes:

1. `@app.route('/')`: This route defines the home page of the web application, which renders the 'index.html' template. The 'index.html' file should contain a form with an input field or textarea
2. where the user can enter the news article text.
3. `@app.route ('/predict', methods=['POST']) `: This route handles the POST request from the form on the 'index.html' template. When the user submits the form, the '/predict' route is called, which retrieves the user's input (news article) from the form, calls the `fake_news_det` function to predict whether the news is 'FAKE' or 'REAL', and renders the 'index.html' template with the prediction displayed.

Running the Flask App:

The code checks if the script is being run directly (not imported as a module), and if so, it starts the Flask application in debug mode using `app.run(debug=True)`. This allows for easier development and debugging, as the server will automatically reload when changes are made to the code.

User Interface and Interaction:

To use the web application, the user needs to navigate to the URL where the Flask app is running (e.g., `http://localhost:5000/` if running locally). The 'index.html' template will be rendered, displaying a form where the user can enter the news article text.

After entering the text and submitting the form, the '/predict' route is called, and the prediction ('FAKE' or 'REAL') is rendered on the same page. The user can then interpret the prediction and determine the authenticity of the news article.

Overall, the web application provides a user-friendly interface for interacting with the fake news detection system, making it accessible to a wider audience without requiring technical expertise in machine learning or data processing.

CHAPTER 5

EXPERIMENTAL RESULTS

This section presents the performance evaluation of three machine learning models Passive Aggressive Classifier, Random Forest Classifier, and Naive Bayes Classifier, for the task of fake news detection. The models were trained and tested on the 'news.csv' dataset, and their performance was assessed using several evaluation metrics.

5.1 Model Performance Results

Performance evaluation of each machine learning model on the parameter like accuracy, precision, recall and F1 score.

5.1.1 Passive Aggressive Classifier

The Passive Aggressive Classifier is an online learning algorithm that updates its parameters aggressively when a prediction is incorrect, making it suitable for large-scale data scenarios. It exhibited excellent performance across all evaluation metrics, correctly classifying 93% of the news articles and maintaining a balanced precision, recall, and F1-score for both 'FAKE' and 'REAL' classes.

- Accuracy: 93%
- Precision (FAKE): 0.93
- Precision (REAL): 0.93
- Recall (FAKE): 0.93
- Recall (REAL): 0.94
- F1-score (FAKE): 0.93
- F1-score (REAL): 0.94

5.1.2 Random Forest Classifier

The Random Forest Classifier is an ensemble learning technique that builds multiple decision trees and combines their predictions, leading to more accurate and robust results. It achieved an overall accuracy of 90% and performed exceptionally well in identifying 'REAL' news articles, with a high recall of 0.92 and balanced precision and F1-score.

- Accuracy: 90%
- Precision (FAKE): 0.91
- Precision (REAL): 0.90
- Recall (FAKE): 0.89
- Recall (REAL): 0.92
- F1-score (FAKE): 0.90
- F1-score (REAL): 0.90

5.1.3 Naive Bayes Classifier:

The Naive Bayes Classifier is a probabilistic classifier based on Bayes' Theorem, which assumes feature independence. While it had the lowest overall accuracy of 88%, it demonstrated high precision for the 'FAKE' class (0.92) and high recall for the 'REAL' class (0.93), indicating its strength in identifying specific aspects of fake and real news articles, respectively.

- Accuracy: 88%
- Precision (FAKE): 0.92
- Precision (REAL): 0.85
- Recall (FAKE): 0.83
- Recall (REAL): 0.93
- F1-score (FAKE): 0.87
- F1-score (REAL): 0.89

5.2 Model Selection

The Passive Aggressive Classifier emerged as the top-performing model, achieving the highest accuracy of 93% and demonstrating balanced performance across precision, recall, and F1-score for both 'FAKE' and 'REAL' classes. The Random Forest Classifier also performed well, with an accuracy of 90% and balanced metrics for the 'REAL' class. While the Naive Bayes Classifier had the lowest accuracy, it exhibited high precision for 'FAKE' news and high recall for 'REAL' news. Based on these results, the Passive Aggressive Classifier is recommended for deployment in the fake news detection system, although further model optimization may enhance overall performance.

CHAPTER 6

CONCLUSIONS AND FUTURE SCOPE

6.1 Conclusions

New fake news detection algorithms using machine learning algorithms show promising potential in addressing the widespread problem of fake news on social media. Using Twitter's advanced analytics and classification algorithms, the model correctly predicts the likelihood that news will actually be fake, providing robust solutions to deal with misinformation Achievable research happens confirm if the proposed system is economically viable, independent of open source software Use available technologies, such as open source software and publicly available data sets for benefit it is in the lower cost range Customized features were minimal, ensuring that the system stayed within budget constraints and did not require significant capital investment . Furthermore, the model's reliance on readily available resources means that it can be implemented without significant financial burden, making it an attractive option for organizations with limited budgets This work shows for example it can be easily integrated into existing services

6.2 Future Scope

Future work includes several enhancements and extensions to further improve the effectiveness of the Fake news detection algorithm

Incorporating additional data sources: Expanding the model to include data from other social media platforms and news websites could improve accuracy and allow for more detailed analysis.

Real-time detection: Having the ability to detect fake news in real time will make the system more efficient and useful for users who want to verify news immediately.

Using feedback: Providing feedback mechanisms where users can report inaccuracies or confirm findings will help improve model accuracy over time in by continuous learning.

Scalability and Performance Optimization: Ensuring that the system can efficiently handle large amounts of data and be scaled to meet the growing number of users is important to maximize adoption

Multilingual support: Extending the model to support multiple languages would provide access to a global audience, and handle fake news in a variety of language contexts.

REFERENCES

1. Project Source code: Github.com
2. Parikh, S. B., & Atrey, P. K. Media-Rich Fake News Detection: A Survey IEEE Paper (2018, April).
3. Automatic deception detection: Methods for finding fake news IEEE Paper (2015, November).
4. Weekly supervised learning for fake news detection on Twitter (X.com) IEEE Paper (2018, August)
5. Automatic Online Fake News Detection Combining Content and Social Signals IEEE Paper (2018, May)
6. GFG Article on Passive Aggressive Classifiers by geeksforgeeks.org/user/alokesh985/ https://www.geeksforgeeks.org/passive-aggressive-classifiers/
7. Article on Random Forest Algorithm by geeksforgeeks.org/user/susmit_sekhar_bhakta/ https://www.geeksforgeeks.org/random-forest-algorithm-in-machine-learning/
8. GFG Article on Naive Bayes Classifiers by geeksforgeeks.org /https://www.geeksforgeeks.org/naive-bayes-classifiers/
9. Python 3.12.3 documentation