Addis Ababa Science and Technology University
Software Engineering Department
Software Component Design
Fake News Detection using Machine Learning

Group members	Id
1. Adugna Benti Itefa	Ets0069/12
2. Rebuma Tadele Debebe	Ets1086/13
3. Nebiyat Bekele Eticha	Ets1052/13
4. Nebiyu Zewge Gedlu	Ets1051/13
5. Sabona Misgana Kebede	Ets1114/13
6. Haileamlak Desalegn	Ets0326/12

Submission date – 30/5/2024

1. Abstract

The rapid proliferation of fake news has become a pressing issue in today's digital era. This project aims to develop a machine learning-based system to identify and classify news articles as real or fake. Leveraging advanced natural language processing (NLP) techniques and machine learning algorithms, this project seeks to provide an effective solution for mitigating the spread of misinformation.

2. Introduction

Fake news refers to fabricated information that mimics the style of legitimate news articles but lacks verifiable facts. The consequences of fake news are far-reaching, influencing public opinion, politics, and even economies. Detecting fake news is a challenging task due to the complexity and subtlety of language.

This project aims to automate fake news detection using machine learning. By training models on labeled datasets, we strive to build a system that can predict the authenticity of news articles with high accuracy.

3. Objectives

- ❖ Develop a machine learning model capable of distinguishing fake news from real news.
- ❖ Use natural language processing techniques to preprocess and analyze textual data.
- ❖ Evaluate the effectiveness of various machine learning algorithms for fake news detection.

4. Methodology

1. **Data Collection**
   - **Datasets:** Publicly available datasets containing labeled fake and real news articles.
   - **Features:** Title, text, subject, date, class.
2. **Data Preprocessing**
   - Tokenization
   - Stopword, punctuation, irrelevant space removal
   - Stemming and lemmatization
   - Shuffling the dataset
3. **Model Selection**
   - **Algorithms Tested:**
     - ➢ Logistic Regression
     - ➢ Decision Tree Classifier

5. Model Training and Evaluation

- Train models on the preprocessed dataset.
- Evaluate performance using metrics such as:
  - ➢ Accuracy_score
  - ➢ Confusion matrix

6. Deployment

- Google colab

7. Tools and Technologies

- **Programming Language:** Python
- **Libraries:**
  - ➢ Pandas, NumPy for data manipulation
  - ➢ Scikit-learn for machine learning
  - ➢ NLTK for NLP
  - ➢ LogisticRegression and Decision Tree Algorithm for learning models
- **Dataset:** News.csv

8. Conclusion

This project demonstrates the potential of machine learning in addressing the fake news epidemic. By combining NLP techniques with advanced algorithms, we can effectively classify news articles and curb misinformation. While challenges remain, the system provides a robust foundation for future enhancements.