**Synopsis on**

**Fake News Detection**

**Submitted in the partial fulfilment of the Degree of Bachelor of Technology**

**(Computer Science and Engineering)**

**Submitted by**

**Ammar Adil 02915002721**

**Sanskriti Singh 05415002721**

**Rohit Kumar Kaushal 00515002721**

**Yuvraj Gupta 02815002721**

**Under the supervision of**

**Ms. Jyoti**



**Department of computer Science and Engineering**

**Maharaja Surajmal Institute of Technology Janakpuri, New Delhi.**

**2021-25**

|  |  |  |
| --- | --- | --- |
|  | **INDEX** |  |
| **1. Abstract** |  | **1.** |
| **2. Introduction** |  | **2.** |
| **3. Literature review** |  | **2.** |
| **4. Objective** |  | **3.** |
| **5. Research Methodology** |  | **3.** |
| **6. Gantt chart** |  | **4.** |
| **7. References** |  | **5.** |

# Abstract

# The rapid spread of fake news has become a significant concern in today's digital age, leading to widespread misinformation. This project aims to develop a machine learning-based model to detect fake news by analyzing various features of news articles. The model will be trained on a labeled dataset of real and fake news and will employ algorithms such as Naive Bayes, Support Vector Machines (SVM), and Neural Networks. The effectiveness of the model will be evaluated based on accuracy, precision, recall, and F1-score.

# Introduction

In recent years, the rapid advancement of technology has significantly altered the way information is disseminated and consumed. Mobile phones and other digital platforms have made news and information more accessible than ever. However, this convenience comes with the downside of increased exposure to misinformation and fake news, often driven by biased reporting and political agendas. This problem has led to real-world consequences, such as riots and polarization of voters.

To address this issue, our project focuses on analyzing news data from social media and other sources to classify it as fake or real. By applying machine learning algorithms to text data, we aim to develop a reliable model that can help mitigate the spread of fake news.

# Literature Survey

# Several studies have been conducted on fake news detection using various machine learning techniques. Research by Shu et al. (2017) highlights the effectiveness of feature-based approaches, emphasizing the role of linguistic features, metadata, and network characteristics in identifying fake news. Their work demonstrates that integrating multiple types of features can improve detection performance.

# Recent advancements have introduced the use of deep learning models for text classification tasks, offering significant improvements in accuracy. For instance, Long Short-Term Memory (LSTM) networks have been employed to capture the temporal dependencies in text data, which enhances the model's ability to understand context and semantic meaning (Hochreiter & Schmidhuber, 1997). Additionally, BERT (Bidirectional Encoder Representations from Transformers) has emerged as a powerful tool for natural language understanding. BERT’s bidirectional training allows it to consider the context from both directions, making it highly effective for nuanced text classification tasks (Devlin et al., 2019).

# The use of comprehensive datasets has also been crucial in advancing fake news detection. Datasets such as the Fake News Corpus and the LIAR dataset have been widely utilized to train and evaluate machine learning models. The Fake News Corpus, for example, provides a diverse set of news articles labeled as real or fake, which helps in building robust models. The LIAR dataset offers labeled statements from politicians, which are useful for detecting false claims in political news.

# Despite these advancements, the literature suggests that the complexity of fake news requires continuous innovation in detection methods. Fake news often involves subtle variations in language, context, and intent, making it a challenging problem for existing algorithms. Researchers are exploring new techniques, including multi-modal approaches that combine text with image and video analysis, to enhance detection capabilities further. The ongoing development of more sophisticated models and larger, more diverse datasets is crucial for keeping up with the evolving landscape of misinformation.

# Objective

The primary objective of this project is to develop a machine learning model that can accurately classify news articles as fake or real. This involves:

1. To develop a machine learning model capable of distinguishing between fake and real news articles.
2. To compare the performance of different machine learning algorithms in detecting fake news.
3. To analyse the features that most contribute to the accurate classification of news articles.
4. To create a comprehensive dataset of fake and real news for model training and evaluation.

# Research Methodology

The research methodology for this project involves several key steps:

1. **Data Collection**

Gather a dataset of news articles labelled as fake or real from reputable sources.

1. **Data Preprocessing**

Clean the data by removing irrelevant content, handling missing values, and performing text normalization.

1. **Feature Extraction**

Extract relevant features such as word counts, sentiment scores, and linguistic cues.

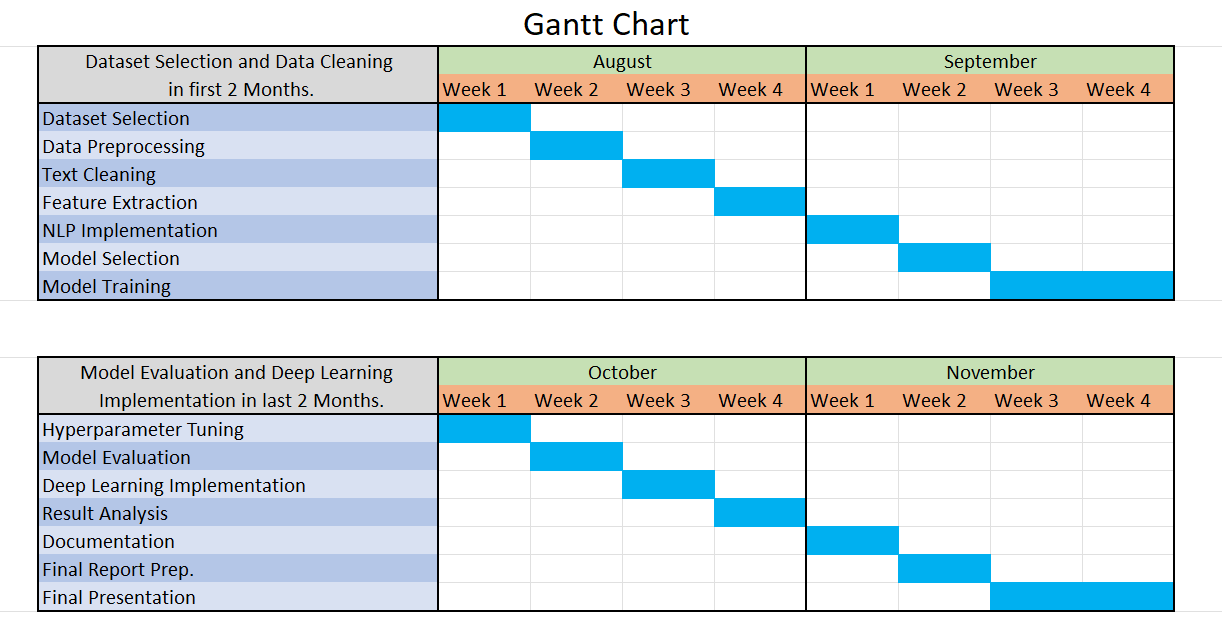
1. **Model Development**

Train various machine learning models (e.g., Naive Bayes, SVM, Neural Networks) on the pre-processed data.

1. **Model Evaluation**

Evaluate the models using metrics like accuracy, precision, recall, and F1-score.

# Gantt chart



**References**

1. Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017). Fake News Detection on Social Media: A Data Mining Perspective. *ACM SIGKDD Explorations Newsletter*, 19(1), 22-36. doi:10.1145/3137597.3137600.
2. Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)* (pp. 4171-4186). doi:10.18653/v1/N19-1423.