

SPECIAL PROJECT

**TITLE: NEWS VERACITY ANALYSIS USING MACHINE LEARNING
ON SOCIAL MEDIA DATA**

NAME: R .Raj Kumar

ENROLL.NO: 22STUCHH010390

ABSTRACT

In the digital age, the rapid dissemination of information has made the detection of fake news a critical challenge. This project focuses on developing a robust fake news detection system that leverages machine learning models to identify and classify misleading or false news articles. The system Analyse news content, such as headlines and text, and determines its credibility based on various linguistic and statistical features. The project utilizes Python programming language, employing several key libraries such as **NumPy**, **Pandas**, **Seaborn**, and **Matplotlib**. NumPy and Pandas are essential for data preprocessing, manipulation, and cleaning, allowing for efficient handling of large datasets. Seaborn and Matplotlib provide powerful visualization tools to explore patterns and relationships in the data, facilitating a better understanding of feature distributions. For classification, multiple machine learning models are implemented and evaluated to determine the best performance. These models include **Logistic Regression**, **Decision Tree Classifier**, **Gradient Boosting Classifier**, and **Random Forest Classifier**. Logistic Regression is used as a baseline model for binary classification, while Decision Tree Classifier helps in understanding feature importance. Random Forest and Gradient Boosting Classifiers are leveraged for their ability to handle complex data patterns and boost overall model accuracy through ensemble learning techniques. The performance of each model is assessed using metrics such as accuracy, precision, recall, and F1-score. By comparing these models, the project aims to identify the most effective approach for detecting fake news with high accuracy and reliability.

KEYWORDS: logistic regression, decision tree classifier, gradient boosting classifier, and random forest classifier.

About This Project

This project focuses on **Fake News Detection** using machine learning techniques. With the growing amount of information shared online, especially through social media, identifying false or misleading news has become a major challenge. Our project aims to develop a system that can automatically detect whether a given news article is **real** or **fake** based on its textual content.

1.Data Preprocessing:

We worked with two datasets: one containing real news articles and the other containing fake ones.

In this stage:

- We cleaned the text data by removing URLs, HTML tags, special characters, punctuation, and digits.
- We standardized the text by converting it to lowercase and removing unnecessary spaces.
- The cleaned text was transformed into numerical features using **TF-IDF vectorization**, allowing machine learning models to understand and process the text.

2. Splitting the Dataset:

- The dataset is split into training and test sets. This is done to ensure that we have data to train the model and separate data to evaluate its performance.

3.Model Training:

- logistic regression, decision tree classifier, gradient boosting classifier, and random forest classifier.