**FAKE NEWS DETECTION USING NLP**

**PROBLEM STATEMENT:**

Nowadays social networks have become the emerging source of news. Due to easy access to these social networks, the news can be easily manipulated which gives rise to fake news. Fake news can be used for economic as well as political benefits. It can be used as a weapon to spread hate among the community which can harm society. So it is crucial to detect fake news to avoid its consequences. This paper proposes a system that can be used for real-time prediction of news to be real or fake. This system is based on natural language processing to extract features from the data and then these features are used for the training of machine learning classifiers such as Naive Bayes, Support Vector Machine (SVM), Random Forest (RF), Stochastic Gradient Descent (SGD), and Logistic Regression (LR). Each of the classifier performance is evaluated on various parameters. Then the best performing classifier is deployed as a website using flask API for real-time prediction of the news.

**FLOW DIAGRAM FOR PROPOSED SYSTEM:**

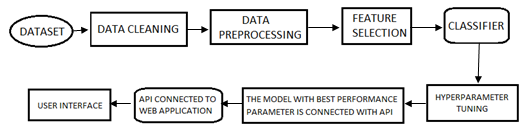


Figure: Flow Diagram For Proposed System

**DATASET:**

There are very few datasets which are available publicly for the detection of fake news. In this paper we have used three different datasets which are available online. The first dataset ISOT Fake News dataset is obtained from a website. The second data that is used in the project is the Fake News Detection dataset from Kaggle. The third dataset used is the Real and Fake News Dataset which is obtained from Kaggle.

**MERGING THE DATASET:**

The first dataset ISOT Fake News dataset is obtained from a website. This dataset was created using data from real world news sources. This dataset consists of two types of articles: fake and real. The dataset consists of two CSV files. First file contains all the news which is true and the second file contains the news which is fake.

Each article contains the following information: article title, text, type and the date the article was published on. The second dataset used in the project is the Fake News Detection dataset from Kaggle. This dataset consists of 4 columns which are the URLs of the news source, the Headline of the news, the Body of the news that is the content of the news and the last column contains the Label of the news which tells whether the news is fake or not. Next, the two datasets are merged together to obtain a single dataset. After the merge we obtained a dataset with 10344 records. Finally, we obtain a master dataset by merging the first dataset with the above merged dataset [dataset with 10344 records], hence the final obtained master dataset consists of 54726 records and three columns, Title, text and Class.