**IMPLEMENTATION:**

**MODULES:**

* User
* Admin
* Data Preprocessing
* Machine Learning

**MODULES DESCRIPTION:**

**User:**

The User can register the first. While registering he required a valid user email and mobile for further communications. Once the user register then admin can activate the user. Once admin activated the user then user can login into our system. User can upload the dataset based on our dataset column matched. For algorithm execution data must be in float format. Here we took jruvika dataset and real or fake dataset. User can also add the new data for existing dataset based on our Django application. User can click the Classification in the web page so that the data calculated Accuracy and macro avg and weighted avg based on the algorithms. User can display the algorithm and prediction results. After that user can logout.

**Admin:**

Admin can login with his login details. Admin can activate the registered users. Once he activate then only the user can login into our system. Admin can view the overall data in the browser. Admin can click the Results in the web page so calculated Accuracy and macro avg and weighted avg based on the algorithms. All algorithms execution complete then admin can see the overall accuracy in web page. After that admin can logout.

**Data Preprocessing:**

The preliminary steps of preparing the data for training the classifiers are pre-processing, vectorization and feature extraction. The datasets contained a lot of extra metadata so only the required columns of headline/title, text/body and label/type are extracted. Then the complete corpus is changed into lowercase; white spaces, punctuations and stop words are removed, stemming is applied and finally, the headline and body of the article are concatenated together to make it a single instance of the corpus. To convert the text into a vector form count vectorizer is used.

**Machine learning**:

Machine Learned classifiers are supervised methods in which fully labeled data is being used for training the classifiers using the explicit features extracted from training samples. This work deals with only the headline and body of the news article and tries to distinguish between the fake and real news based on features extracted from these two fields. Five supervised machine-learned classifiers are being independently trained and tested for two different datasets according to the architecture described in given basepaper.

Algorithm 1: Working steps in Machine Learning framework

1: Import required python libraries 2: Read the data file

3: Perform data cleaning

4: Convert text into vector form using Count Vectorizer 5: Split the dataset into train and test samples

6: Perform tf-idf feature extraction

7: Train ML classifiers (SVM, LR, RF, NB, KNN) on extracted features

8: Test the trained models using test samples.

The classifier which bags up the highest accuracy could be determined as the best classifier