**System Implementation**

**MODULES:**

* **Login Module** :

In the Login module, the users who have already registered with the application can login using their credentials such as userid and password. The userid and password will be checked with the data stored already in the dataset. If it is valid, the users will be logged in to the application.

* **Registration** :

In the Registration module, the users can register their credentials and details such as their name, userid, password, phone number, email and gender when they are using the application for the first time. The details are stored in the database with an automatic generated id. They can use the userid and password to login to the application next time.

* **Upload Module :**

The Upload module is divided into two sub modules called Upload Dataset and Upload News. Upload Dataset module is to upload the entire database and Upload News is to upload the news separately which will be appended into the dataset as soon as it is submitted. In the Upload News sub module, news details such as newsid, title, text and label are uploaded.

* **Analysis Module :**

The Analysis module is used to perform analysis using various Machine Learning Algorithms such as TFIDF Vectorizer, Count Vectorizer, Passive Aggressive Classifier and Hashing Vectorizer.

* **Graphical Analysis Module :**

Graphical analysis based on NB (Naïve Bayes) algorithm is used to analyse the fake and real news.

**3.2 PROPOSED SYSTEM**

In this project a model was built based on the TFIDF Vectorizer, Count Vectorizer( i.e ) word tallies relatives to how often they are used in other artices in your dataset ) can help . Since this problem is a kind of text classification, Implementing a Naive Bayes classifier will be best as this is standard for text-based  processing. The actual goal is in developing a model which was the text transformation (count vectorizer vs tfidf vectorizer) and choosing which type of text to use (headlines vs full text). Now the next step is to extract the most optimal features for countvectorizer or tfidf-vectorizer, this is done by using a n-number of the most used words, and/or phrases, lower casing or not, mainly removing the stop words which are common words such as “the”, “when”, and “there” and only using those words that appear at least a given number of times in a given text dataset.

In this project, a model was built based on some of the NLP classification models such as TFIDF Vectorizer, Count Vectorizer, Hash Vectorizer and Passive Aggressive Vectorizer. The model analyses the dataset by extracting the most optimal features of text. It was done by using n-number of the most used words, and/or phrases, lower case or not, mainly removing the stop words which are the common words such as “the”, “is”, “when”, and “there” and only using those words that appear at least a given number of times in the uploaded dataset.