**Analysing social media data through data mining**

**OBJECTIVE:**

The primary goal of this project is to analyse social media data using data mining to determine whether social media text input can be classified as positive, negative, or neutral using Decision Tree, Random Forest, Naive Bias, and NLP machine learning approaches.

**ABSTRACT**

Nowadays, people from all around the world use social media sites to share information. Twitter for example is a platform in which users send, read posts known as ‘tweets’ and interact with different communities. Users share their daily lives, post their opinions on everything such as brands and places. Companies can benefit from this massive platform by collecting data related to opinions on them. The aim of this paper is to present a model that can perform sentiment analysis of real data collected from Twitter. Data in Twitter is highly unstructured which makes it difficult to analyse. However, our proposed model is different from prior work in this field because it combined the use of supervised and unsupervised machine learning algorithms. The process of performing sentiment analysis as follows: Tweet extracted directly from Twitter API, then cleaning and discovery of data performed. After that the data were fed into several models for the purpose of training. Each tweet extracted classified based on its sentiment whether it is a positive, negative or neutral. Data were collected on two subjects McDonalds and KFC to show which restaurant has more popularity. Different machine learning algorithms were used. The result from these models were tested using various testing metrics like cross validation and f-score. Moreover, our model demonstrates strong performance on mining texts extracted directly from Twitter.

**Keywords:** Decision Tree, Random Forest, Naïve Bias.

**EXISTING METHOD**

The increasing growth of machine learning, computer techniques divided into traditional methods and machine learning methods. This section describes the related works of classification of a Analysing social media data through data mining Using Machine Learning Model Detection and how machine learning methods are better than traditional methods. The existing method in this project have a certain flow is used for model development Decision Tree and Naïve bias are used algorithms in existing system. But it requires large memory and result is not accurate.

**Disadvantages:**

1. Accuracy low

2. Requires more time

3. Difficult to handle

**PROPOSED SYSTEM**

Many machine learning algorithms are available for prediction and diagnosis a Analysing social media data through data mining Some of the machine learning algorithm are Decision Tree, Random Forest. We used proposed Analysing social media data through data mining In this stage we have first implement Random Forest Classifier algorithm on these dataset and the implement algorithm individual then we are implement Voting Ensemble algorithm for combine these results and an compute the final accuracy.

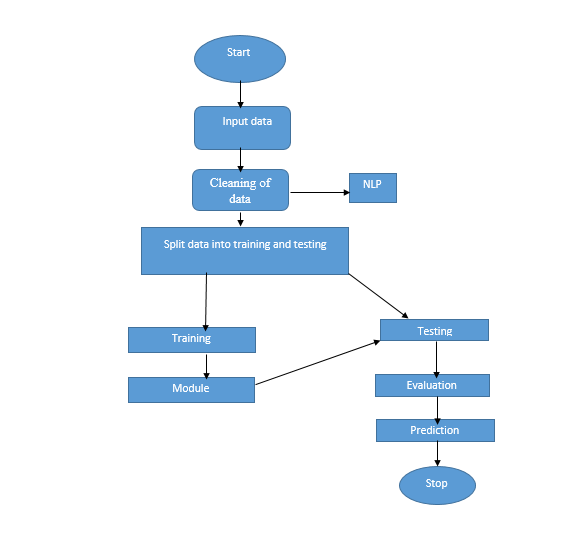
**Advantages:**

1. Requires less time

2. Good Accuracy

3. Easy to handle

**Block Diagram:**

****

**Fig 1. Block Diagram of Proposed System**

**HARDAWARE AND SOFTWARE REQUERMENTS**

**H/W Configuration:**

• Processor - I3/Intel Processor

• Hard Disk -160 GB

• RAM - 8 GB

**S/W Configuration:**

• Operating System : Windows 7/8/10 .

• Server side Script : HTML, CSS & JS.

• IDE : Pycharm.

• Libraries Used : Numpy, IO, OS, Django, keras.

• Technology : Python 3.6+.

**LEARNING OUTCOMES:**

* About Classification in machine learning
* About Preprocessing Techniques
* About Decision Tree
* About Random Forest
* About Naïve Bias,
* Knowledge on PyCharm Editor