

**Aim:** Implement Multi-Class Classifier for Text Categorization from various news articles or websites to categorize based on news title.

**Objective:** To understand about how text categorization is carried out using different Classifier algorithms.

**Pre-requisites:** Knowledge of the Classification algorithm to be used for categorization.

**Theory:**

- Multi-Class Classifier for text Categorization.

→ Text Classification is one of the important applications of Natural language processing. It is the process of categorizing text into a group of words by using NLP. Text Classification can automatically analyze text & then assign a set of predefined tags or categories based on its context. There are variety of ways in which machine learning to classify texts.

(i) Rule-based System - In the rule-based approach, texts are separated into an organized group using a set of handcrafted linguistic rules. Those handcrafted linguistic rules contain users to define a list of words.

(ii) Machine System - Machine based Classifier learns to make a classification based on past observation from the data sets. User data is prelabelled as train & test data.

(iii) Hybrid System - Hybrid approach usage combines a rule based & machine based approach. It is an user-based to tag creation & use machine learning to train & create rule.

## • Uses of Text Categorization in NLP Application

→ Text Classification is becoming an increasingly important part of business as it allows to easily get insights from data & automate business processes. For Eg.

(i) Sentiment Analysis - The process of understanding if a given text is talking positively or negatively about a given subject.

(ii) Topic Detection - The task of identifying the theme or topic of a piece of text [Eg. Customer Support or Pricing in Customer Feedback]

(iii) Language Detection - The procedure of detecting the language of a given text [Eg. Know if an incoming Support text is written in English or Spanish]

## \* Conclusion:

→ We implemented Classification techniques for text Categorization over the Subjected data.