Experiment No.1
Study various applications of NLP and Formulate the Problem Statement for Mini Project based on chosen real world NLP applications
Date of Performance:
Date of Submission:

Aim:
The aim of this project is to develop an efficient and user-friendly news text classification system that automates the categorization of news articles into relevant topics, alleviating the challenges posed by the vast volume of digital news content.

Objective:
- Develop a robust news text classification model capable of categorizing news articles into multiple relevant topics or classes.
- Gather and preprocess a diverse dataset of news articles to ensure data quality and consistency.
- Train and evaluate machine learning models, including deep learning, to achieve accurate news classification.

Abstract:
In the digital age, the overwhelming volume of news articles poses a challenge for users seeking relevant information. This project introduces a news text classification system that employs natural language processing and machine learning to automate the categorization of news articles into pertinent topics. The system's primary objective is to enhance the accessibility and organization of news content. By using Python libraries for NLP, machine learning, and web development, and by making sound hardware and software choices, this project aims to provide an efficient, scalable, and user-friendly solution.

Methodology:
1. **Data Collection:** Collect a diverse dataset of news articles from reputable sources, encompassing a broad spectrum of topics and genres.
2. **Text Preprocessing:** Clean and prepare the text data through tasks like tokenization, stopword removal, stemming, and special character removal.
3. **Text Vectorization:** Convert the preprocessed text into numerical form using techniques like TF-IDF (Term Frequency-Inverse Document Frequency) or word embeddings.
4. **Feature Extraction:** Extract relevant features from the text vectors to create a feature matrix that will serve as input for the machine learning model.
5. **ML Model Training**: Choose an appropriate machine learning model, train it using the feature matrix, and fine-tune hyperparameters for optimal performance.
6. **Deployment:** Implement the model into a user-friendly interface, facilitating real-time news article classification, and ensure the system can handle scalability and variations in news content sources.