

---

# TEXT CLASSIFICATION USING NAIVE BAYES AND SENTIMENT ANALYSIS ON BLOG POSTS

---

## Overview

In this assignment, you will work on the "blogs\_categories.csv" dataset, which contains blog posts categorized into various themes. Your task will be to build a text classification model using the Naive Bayes algorithm to categorize the blog posts accurately. Furthermore, you will perform sentiment analysis to understand the general sentiment (positive, negative, neutral) expressed in these posts. This assignment will enhance your understanding of text classification, sentiment analysis, and the practical application of the Naive Bayes algorithm in Natural Language Processing (NLP).

## Dataset

The provided dataset, "blogs\_categories.csv", consists of blog posts along with their associated categories. Each row represents a blog post with the following columns:

- **Text:** The content of the blog post. Column name: Data
- **Category:** The category to which the blog post belongs. Column name: Labels

## Tasks

### 1. Data Exploration and Preprocessing

- Load the "blogs\_categories.csv" dataset and perform an exploratory data analysis to understand its structure and content.
- Preprocess the data by cleaning the text (removing punctuation, converting to lowercase, etc.), tokenizing, and removing stopwords.
- Perform feature extraction to convert text data into a format that can be used by the Naive Bayes model, using techniques such as TF-IDF.

### 2. Naive Bayes Model for Text Classification

- Split the data into training and test sets.

- Implement a Naive Bayes classifier to categorize the blog posts into their respective categories. You can use libraries like scikit-learn for this purpose.
- Train the model on the training set and make predictions on the test set.

### **3. Sentiment Analysis**

- Choose a suitable library or method for performing sentiment analysis on the blog post texts.
- Analyze the sentiments expressed in the blog posts and categorize them as positive, negative, or neutral. Consider only the Data column and get the sentiment for each blog.
- Examine the distribution of sentiments across different categories and summarize your findings.

### **4. Evaluation**

- Evaluate the performance of your Naive Bayes classifier using metrics such as accuracy, precision, recall, and F1-score.
- Discuss the performance of the model and any challenges encountered during the classification process.
- Reflect on the sentiment analysis results and their implications regarding the content of the blog posts.

### **Submission Guidelines**

- Your submission should include a comprehensive report and the complete codebase.
- Your code should be well-documented and include comments explaining the major steps.

### **Evaluation Criteria**

- Correct implementation of data preprocessing and feature extraction.
- Accuracy and robustness of the Naive Bayes classification model.
- Depth and insightfulness of the sentiment analysis.
- Clarity and thoroughness of the evaluation and discussion sections.
- Overall quality and organization of the report and code.

Good luck, and we look forward to your insightful analysis of the blog posts dataset!