**Text Classification using Multinomial Naive Bayes**

**Overview:**

This document provides a guide on how to train a text classification model using Multinomial Naive Bayes and apply it to new text data for label prediction.

**Methodology:**

**1.Loading the Data:**

* Read the training data from 'train.csv' using pandas.
* Explore the dataset to understand its structure.

**2.Text Preprocessing:**

* Clean the text data by removing punctuation, converting to lowercase, and eliminating stopwords.
* Tokenize the text using the NLTK library.
* Utilize CountVectorizer to convert the text into a bag-of-words representation.

**3.Feature Extraction:**

* Transform the bag-of-words representation into TF-IDF (Term Frequency-Inverse Document Frequency) to capture the significance of words.
* Prepare the feature matrix (X) and target variable (Y) for training.

**4.Train-Test Split:**

* Split the data into training and testing sets for model evaluation.

**5.Model Training:**

* Train a Multinomial Naive Bayes classifier using the training data.
* Evaluate the model on the test set and calculate accuracy.

**6.Testing on New Data:**

* Read new text data from 'test.csv.'
* Preprocess the text and transform it into TF-IDF representation using the previously fitted transformers.
* Predict the labels using the trained Naive Bayes model.

**Model Architecture:**

* The model architecture is based on the Multinomial Naive Bayes algorithm, a probabilistic classifier suitable for text classification tasks.
* It leverages the bag-of-words representation and TF-IDF transformation for feature extraction.

**Preprocessing Steps:**

* Load the necessary libraries, including NLTK and scikit-learn.
* Define a text cleaning function to remove punctuation, convert to lowercase, and eliminate stopwords.
* Apply text cleaning to the training and testing datasets.
* Use CountVectorizer to convert text into a bag-of-words representation.
* Transform the bag-of-words into TF-IDF representation for feature extraction.

**Instructions for Prediction:**

**1.Training the Model:**

* Execute the provided code for training the Multinomial Naive Bayes model. Ensure the 'train.csv' file is available.
* The model will be trained on the provided dataset, and accuracy metrics will be displayed.

**2.Testing on New Data:**

* To predict labels for new text entries (testing on new data), execute the code under the "Testing on new data code" section.
* Ensure the 'test.csv' file is available.
* The model will predict labels for the new data, and accuracy metrics will be displayed.

Follow these instructions to successfully train the model and make predictions on new text data.