**Spam prediction using powered AI:**

|  |  |
| --- | --- |
| **Date** | **28-10-2023** |
| **Team ID** | **3921** |
| **Project Name** | **Spam prediction using powered AI** |

**Table contents:**

|  |  |
| --- | --- |
| 1. | Introduction |
| 2. | Machine learning algorithm selection |
| 2.1 | Objective |
| 2.2 | Approach |
| 3. | Model training |
| 3.1 | Objective |
| 3.2 | Approach |
| 4. | Model evaluation |
| 4.1 | Objective |
| 4.2 | Approach |
| 5. | Conclusion |

1**.Introduction:**

This project document outlines the second phase of our Spam Prediction using AI project. In this phase, we will focus on selecting a machine learning algorithm, training the model, and evaluating its performance. This phase is a crucial step in building an effective spam classifier.

**2. Machine Learning Algorithm Selection:**

**2.1 Objective:**

The primary objective in this stage is to select an appropriate machine learning algorithm for our spam classification model. We will consider various algorithms and choose the one that best fits our problem based on factors like accuracy, efficiency, and scalability.

**2.2 Approach:**

Algorithm Research: We will research and evaluate different machine learning algorithms suitable for text classification. Common algorithms to consider include Naïve Bayes, Support Vector Machines (SVM), Decision Trees, Random Forest, and Neural Networks.

Data Compatibility: Assess the compatibility of the chosen algorithms with our dataset, considering factors such as data size, data distribution and text preprocessing requirements.

Algorithm Selection Criteria: Consider factors like accuracy, computational complexity, and ease of implementation. We will also evaluate the algorithm’s ability to handle imbalanced datasets, as spam classification problems often suffer from imbalanced classes.

**3.Model Training:**

**3.1 Objective:**

In this step, we will train the selected machine learning algorithm using a labeled dataset. The trained model will be the core component of our spam classifier.

**3.2 Approach:**

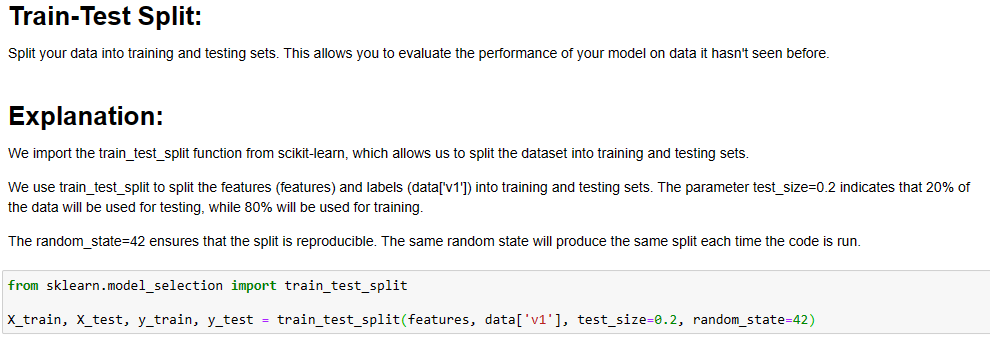
Data Preparation: Preprocess the dataset, including text tokenization, stemming or lemmatization, and handling missing or noisy data.

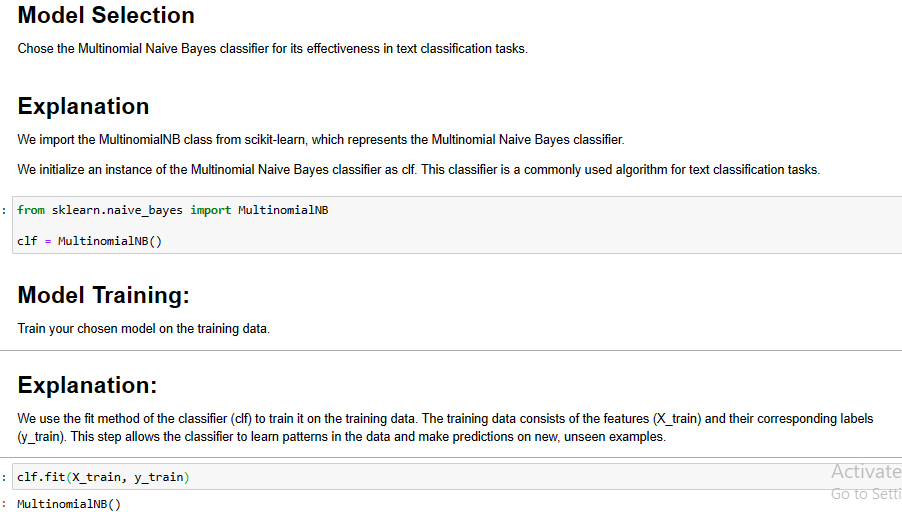
Feature Engineering: Extract relevant features from the text data. This may include bag-of-words, TF-IDF, or word embeddings.

Model Training: Train the selected machine learning algorithm using the preprocessed data. We will use a portion of the dataset for training and reserve another portion for validation and testing.

Hyperparameter Tuning: Fine-tune the hyperparameters of the model to optimize its performance.

**Code:**





**4. Model Evaluation:**

**4.1 Objective :**

The model’s performance needs to be evaluated to ensure that it meets the project’s requirements and objectives. We will use various evaluation metrics to measure the effectiveness of the spam classifier.

**4.2 Approach :**

Evaluation Metrics: We will use common evaluation metrics such as accuracy, precision, recall, F1-score, and ROC-AUC to assess the model’s performance.

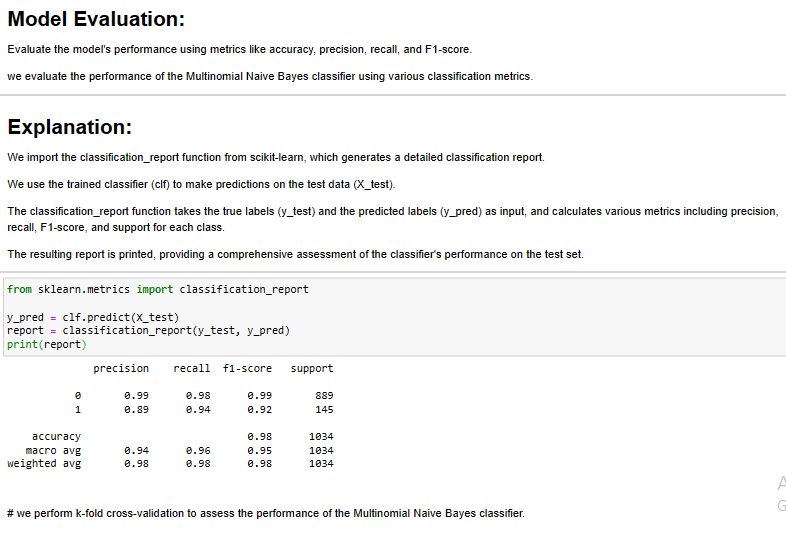
Cross-Validation: Implement cross-validation to obtain a robust estimate of the model’s performance.

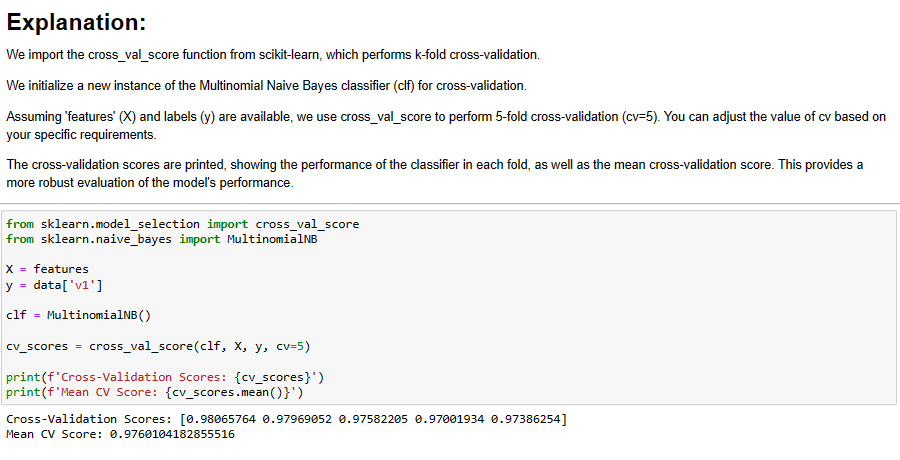
Confusion Matrix Analysis: Analyze the confusion matrix to understand the model’s performance on false positives and false negatives.

Threshold Optimization: Fine-tune the decision threshold for classification, balancing precision and recall based on the specific requirements.

Performance Visualization: Generate visualizations (e.g., ROC curves) to present the model’s performance comprehensively.

**Code**

****

****

**5.Conclusion :**

In this phase, we will be focusing on selecting the most suitable machine learning algorithm, training the model, and evaluating its performance for our spam prediction project. The successful completion of this phase will enable us to move forward with model deployment and integration into the final spam prediction system.