

**The Superior University**

**Project Title**

**SPAM EMAIL DETECTOR**

**Project Details**

1. Course: Data Structures And Algorithm
2. Instructor: (SIR Rasikh Ali)
3. Semester: (3rd)
4. Section: (AI-3C)
5. Submission Date: (10/12/2024)
6. Group Members:

|  |  |  |  |
| --- | --- | --- | --- |
| **Name** | **Roll No** | **Email** | **Contact** |
| ZAID MEHMOOD | SU92-BSAIM-128 | su92-bsaim-f23-128@superior.edu.pk | 0320-7280623 |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

**Table of Contents**

1. **Introduction**
2. **Objectives**
3. **Setup and Tools**
4. **Libraries and Technologies**
5. **Dataset Overview**
6. **Model Development**
7. **Application Workflow**
8. **Key Code Sections**
9. **Screenshots**
10. **Challenges and Solutions**
11. **Conclusion**

**1. Introduction**

**This project is a spam email detection system designed to classify emails as spam or not spam based on their content. It leverages machine learning to analyze textual features for classification.**

**2. Objectives**

* **To develop a system that accurately identifies spam emails.**
* **To preprocess textual data effectively for classification.**
* **To implement and evaluate a machine learning model for spam detection.**
* **To ensure a user-friendly application workflow.**

**3. Setup and Tools**

**Tools Used:**

* **Python: Programming language for development.**
* **Jupyter Notebook: For writing and testing the machine learning code.**
* **Pandas and Numpy: For data manipulation.**
* **Scikit-learn: For machine learning model development.**

**4. Libraries and Technologies**

**Library Purpose:**

* **Pandas: Data manipulation and analysis.**
* **TfidfVectorizer: For converting text to numerical features.**
* **MultinomialNB: Naive Bayes classifier for text classification.**
* **NLTK: For text preprocessing, including stemming and stopword removal.**

**5. Dataset Overview**

* **Dataset Name: Email\_Data.csv**
* **Description: Contains email text and corresponding labels (spam or not spam).**
* **Shape: (Enter number of rows and columns).**

**6. Model Development**

* **Model Used: Multinomial Naive Bayes.**
* **Steps:**
  1. **Data Cleaning: Removed unnecessary text and performed stemming.**
  2. **Feature Extraction: Used TfidfVectorizer for text transformation.**
  3. **Model Training: Split data into training and testing sets, achieving a high accuracy.**
* **Metrics: Evaluated using accuracy, precision, recall, and F1 score.**

**7. Application Workflow**

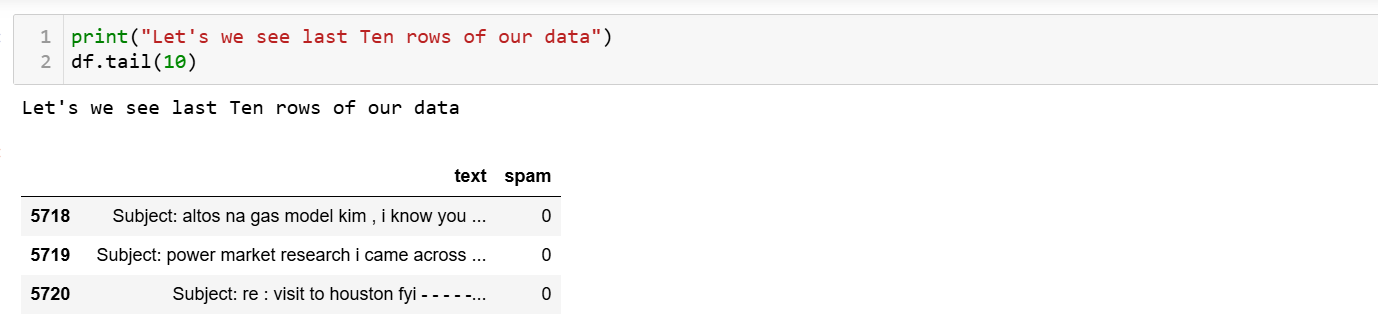
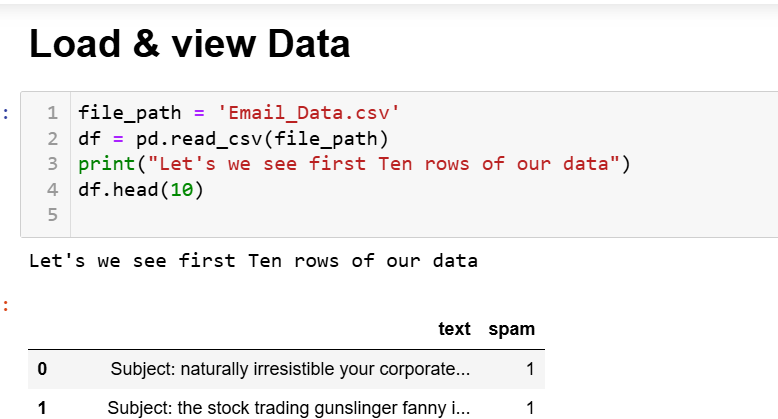
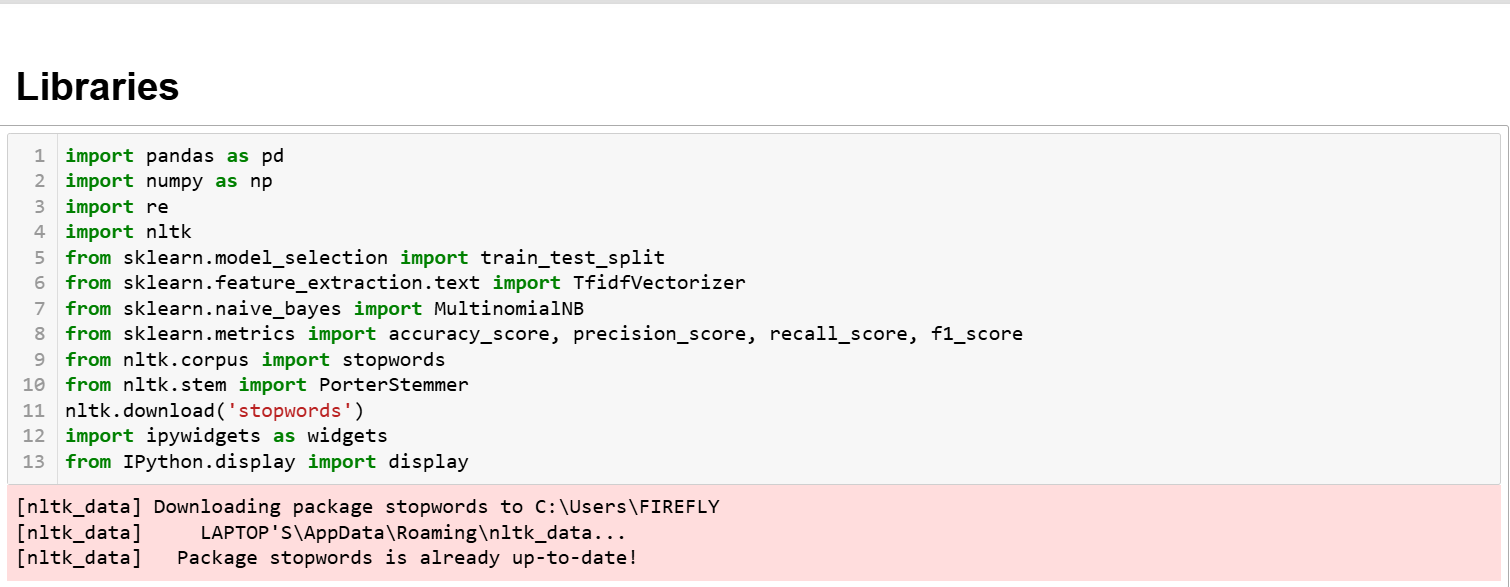
1. **User Inputs: Email content is provided by the user.**
2. **Data Preprocessing: Cleans and transforms the text data.**
3. **Model Prediction: Predicts whether the email is spam or not.**
4. **Output Display: Displays results to the user.**

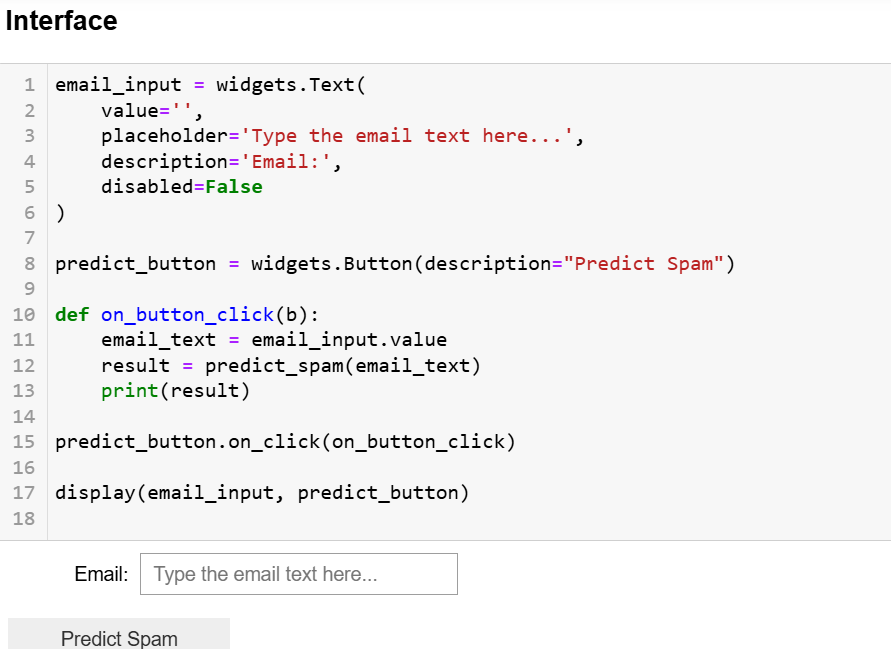
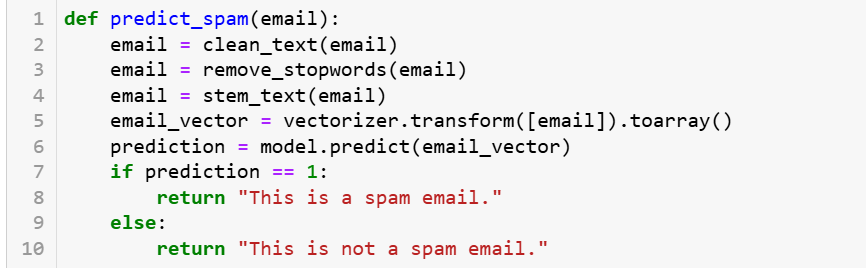
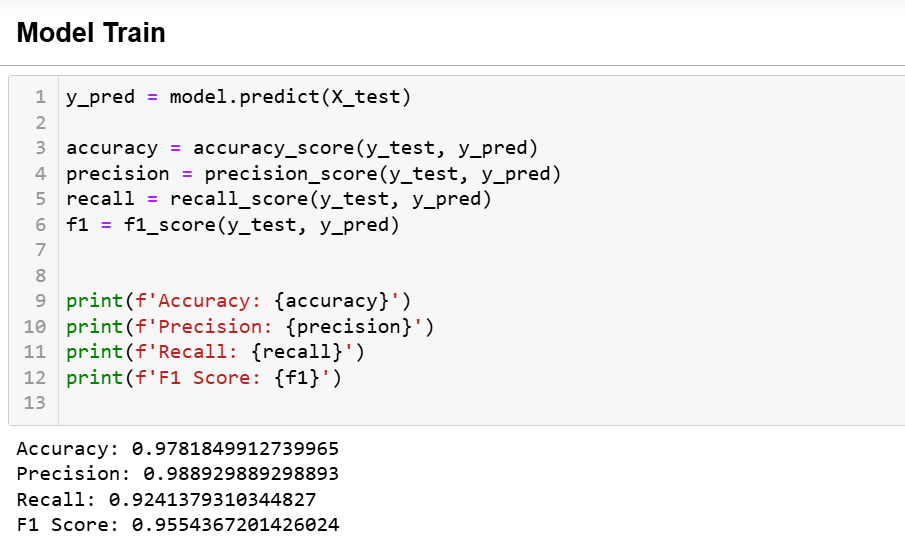
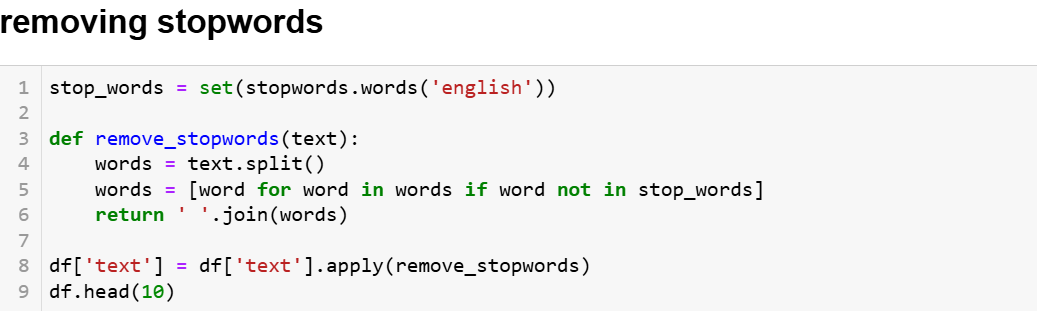
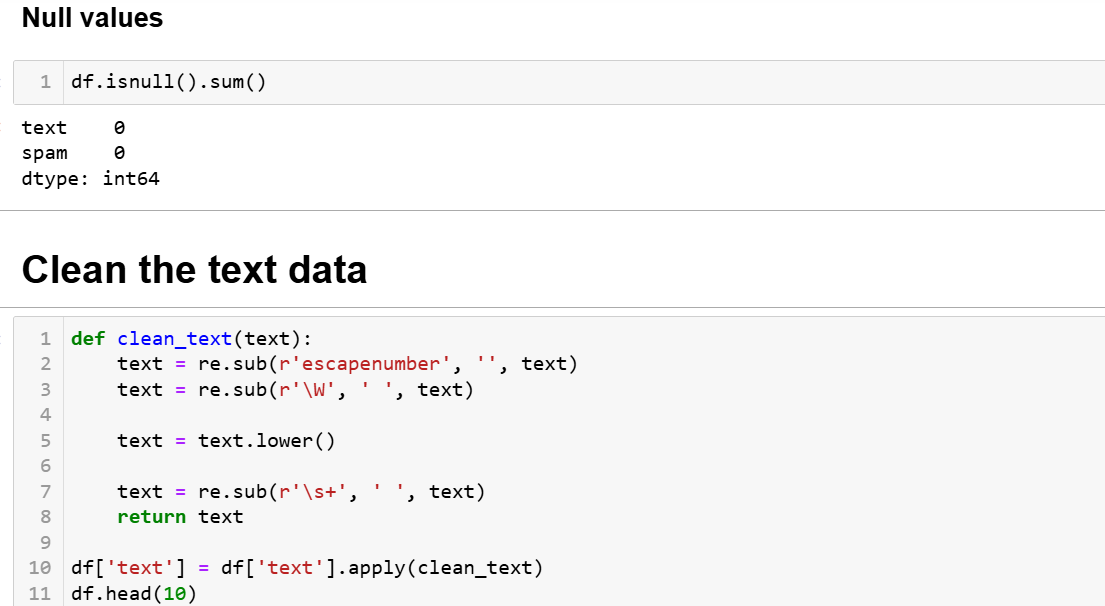
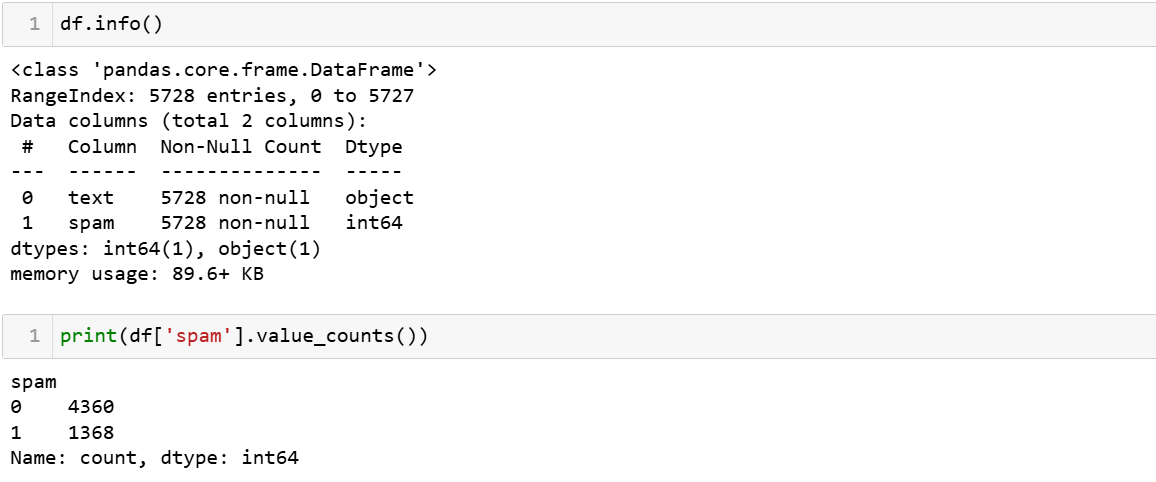
**8. Key Code Sections**

**Highlights:**

* **Data loading and inspection.**
* **Text preprocessing (stemming, removing stopwords).**
* **Model training and evaluation.**
* **Example input and output cases.**

**9. Screenshots**

****

****

**10. Challenges and Solutions**

**Challenge Solution:**

* **Handling inconsistent data formats: Applied extensive cleaning steps.**
* **Improving accuracy: Tuned model parameters and tested various configurations.**

**11. Conclusion**

**This project successfully demonstrates the application of machine learning in spam email detection. The developed system is accurate, efficient, and user-friendly.**