

Name: OM M PATEL

Roll No.: 21BCP094

Group: G3

Division: D2

PROJECT

Title: Spam Detection using Naive Bayes Classifier

Problem Statement :

The goal of this project is to classify text messages as either "spam" (unwanted promotional or malicious messages) or "ham" (regular, non-spam messages). Effective spam detection improves user experience by filtering out undesirable content.

Dataset :

Source: The dataset consists of predefined messages stored in a text.txt file.

Format: Each message is labeled as either "spam" or "ham".

Sample:

- spam Win a free vacation to Bali now
- ham Hey, Let's catup tonight?

Methodology

1. **Data Preprocessing:** Text messages are preprocessed to standardize input (tokenization, lowercase conversion, and removal of punctuation).
2. **Model Choice:** Naive Bayes Classifier, a probabilistic algorithm well-suited for text classification.
3. **Implementation:**
 - **Training:** The classifier learns from the labeled dataset to identify common word frequencies in spam and ham messages.
 - **Testing:** Predefined messages and user input are classified based on the trained model.
4. **Evaluation:** test messages are classified, and user input is also allowed to validate the model's performance in real-time.

Results/Output :

```

• (base) om-college@MacBook-Air MLProject % javac *.java
○ (base) om-college@MacBook-Air MLProject % java Main

Classifying predefined test messages:
Message: "Win a new Iphone" - Classified as: spam
Message: "Hey, Can we met for coffee" - Classified as: ham
Message: "Cheap watches for sale" - Classified as: spam
Message: "How have you been?" - Classified as: ham

Enter your own message to classify as 'spam' or 'ham' (type 'exit' to quit):
Enter message: You have won $500 gift card to Target,click here to Claim reward !!
Your message was classified as: spam
Enter message: You have won $500 gift card son, I am very happy for you
Your message was classified as: ham
Enter message: My friend you have won $500 gift card, Let's party
Your message was classified as: ham
Enter message: My friend you have won $500 gift card, let's party and shop via tha gift card
Your message was classified as: ham
Enter message: My friend you have won $500 gift card, let's party and shop via tha gift card on Zomato
Your message was classified as: ham
Enter message: My friend you have won $500 gift card, let's party and shop via tha gift card onApple
Your message was classified as: ham
Enter message: Your Wells Fargo account has been locked for suspicious activity. Please log in here and verify your account
Your message was classified as: spam
Enter message: Your IRS tax refund is pending acceptance. Must accept within 24 hours
Your message was classified as: spam
Enter message: Hello, let's meet near your place
Your message was classified as: ham
Enter message: Hello let's eat with the coupon you won.
Your message was classified as: ham
Enter message: Hello, Wanna use the Coupon to eat? Then spend 20 dollars to claim it
Your message was classified as: ham
Enter message: Hello, Wanna use the Coupon to eat? Then spend 20 dollars on your account and verify.
Your message was classified as: spam
Enter message: █

```

Insights

- **High Accuracy:** Naive Bayes performed well, especially on short text-based data with distinct spam indicators.
- **Frequent Words:** Words like “redeem”, “claim”, “win”, “free,” and “click” were strong indicators of spam.
- **User Input :** Allowing user input provided a way to test the model interactively and see classification accuracy.

Challenges Faced :

- **Data Sparsity:** As my dataset is quite small, so it may not cover all potential spam/ham message formats, impacting generalizability.
- **User Input Handling:** Ensuring clean input from the console required avoiding multi-line quotes and handling exit conditions properly.