

Title: Spam Detection Model Results and Explanation

1. Dataset Overview

The dataset contains SMS messages labeled as either 'ham' (not spam) or 'spam'. Each row contains a label and the SMS content.

Example rows from the dataset:

label	message
ham	Go until jurong point, crazy.. Available only ...
ham	Ok lar... Joking wif u oni...
spam	Free entry in 2 a wkly comp to win FA Cup fina...

2. Missing Values

No missing values were found in the dataset.

3. Model Performance Metrics

Confusion Matrix:

True Negatives (TN): 966 - The model correctly predicted 'ham' messages.

False Positives (FP): 0 - The model incorrectly predicted spam for 'ham' messages.

False Negatives (FN): 23 - The model incorrectly predicted 'ham' for 'spam' messages.

True Positives (TP): 126 - The model correctly predicted 'spam' messages.

Classification Report:

Precision (ham): 0.98, Recall (ham): 1.00, F1 Score (ham): 0.99

Precision (spam): 1.00, Recall (spam): 0.85, F1 Score (spam): 0.92

Overall Accuracy: 0.98 (98%)

The model has high precision and recall, indicating good performance.

4. Model Accuracy

The overall accuracy of the model is 97.9%, which means that it made correct predictions for nearly all of the SMS messages in the test set.

5. Sample Message Prediction

A sample spam message ('Congratulations! You've won a \$1000 Walmart gift card. Call now to claim your prize.') was tested with the model. The model correctly classified it as 'Spam'.

6. Summary

The Naive Bayes model used for spam detection has performed very well, achieving an accuracy of 97.9%. It correctly identified most spam and non-spam messages, with only a few false negatives for spam.

This project demonstrates the effectiveness of text classification using machine learning, particularly for spam detection.