FRANCESCO SIMBOLA | VR527509

1. ABSTRACT

This project addresses the challenge of email spam detection by implementing a Multinomial Naive Bayes classifier. My approach employs this algorithm, a robust machine
learning method particularly suited for text classification tasks like spam detection.
Emails undergo preprocessing steps such as column cleaning, stopword removal, and feature extraction to refine the dataset and highlight the most relevant patterns associated with spam. The trained Naive Bayes model is then used to classify emails as either spam or not spam, based on the probabilities derived from word occurrences and patterns.
The project evaluates the effectiveness of the model using comprehensive metrics, including the confusion matrix, precision, recall, and F1-score. Visualizations, such as heatmaps and Precision-Recall curves, provide additional insights into the model's performance.

2. MOTIVATION

The Multinomial Naive Bayes algorithm was chosen for spam detection due to its effectiveness with text data represented as word counts, as in the given dataset. It performs well in binary classification tasks and is straightforward to implement.
Its ability to identify the most relevant features enhances its capacity to detect patterns specific to spam. Additionally, it manages imbalanced datasets effectively and provides clear, interpretable results, making it well-suited for this project.

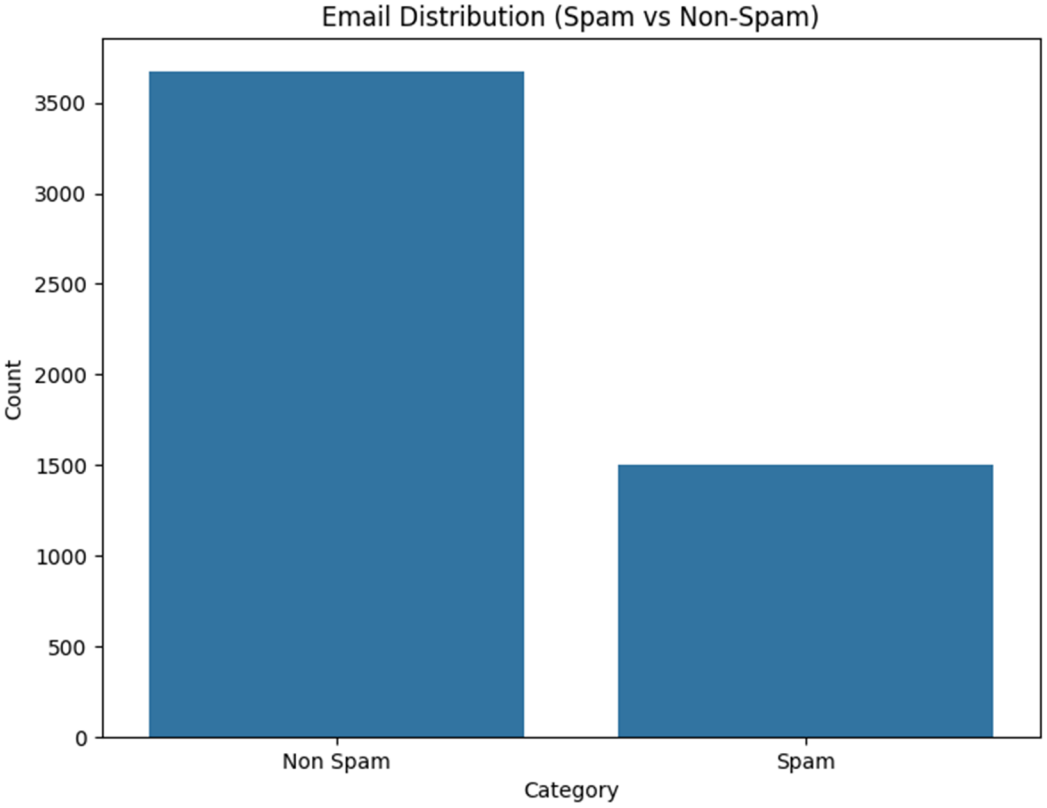
3. DATA PROCESSING

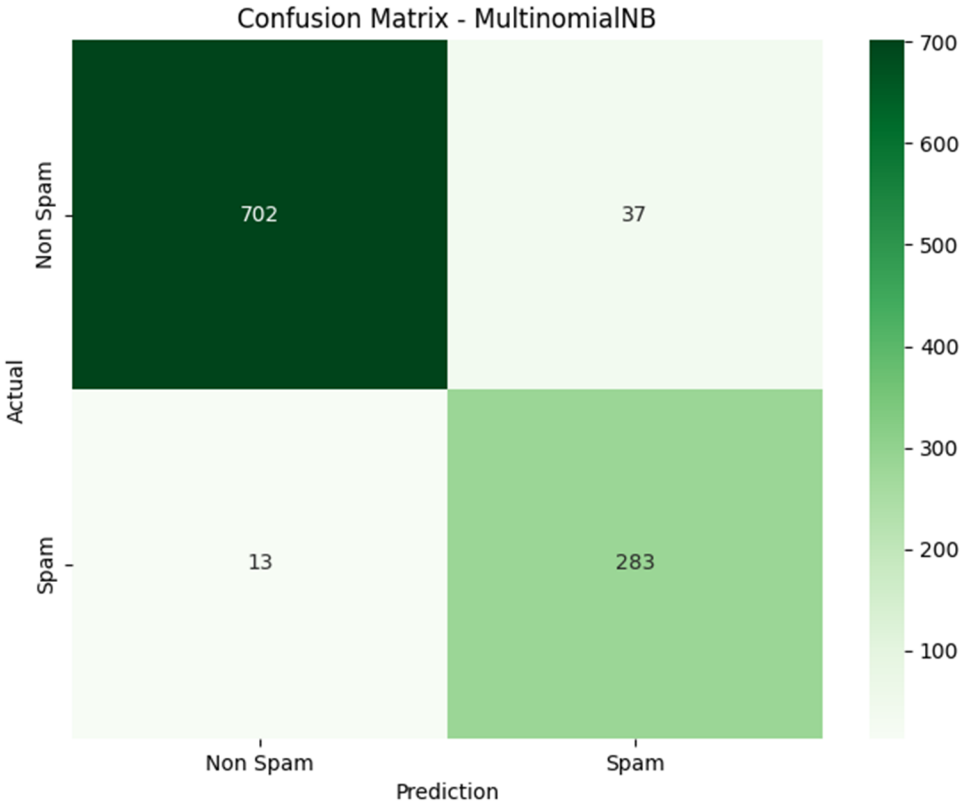
The initial data set of was 5172 rows and 3002 columns, with no values detected. The first column (Email No.) was removed because it is an identifier and does not provide meaningful information for the spam detection model. The initial idea was to remove both stopwords and meaningless words, using an English dictionary, but the result would have led to the removal of 1093 columns and the f-1 score was practically similar. So, after this analysis, I proceeded to remove only the stopwords.

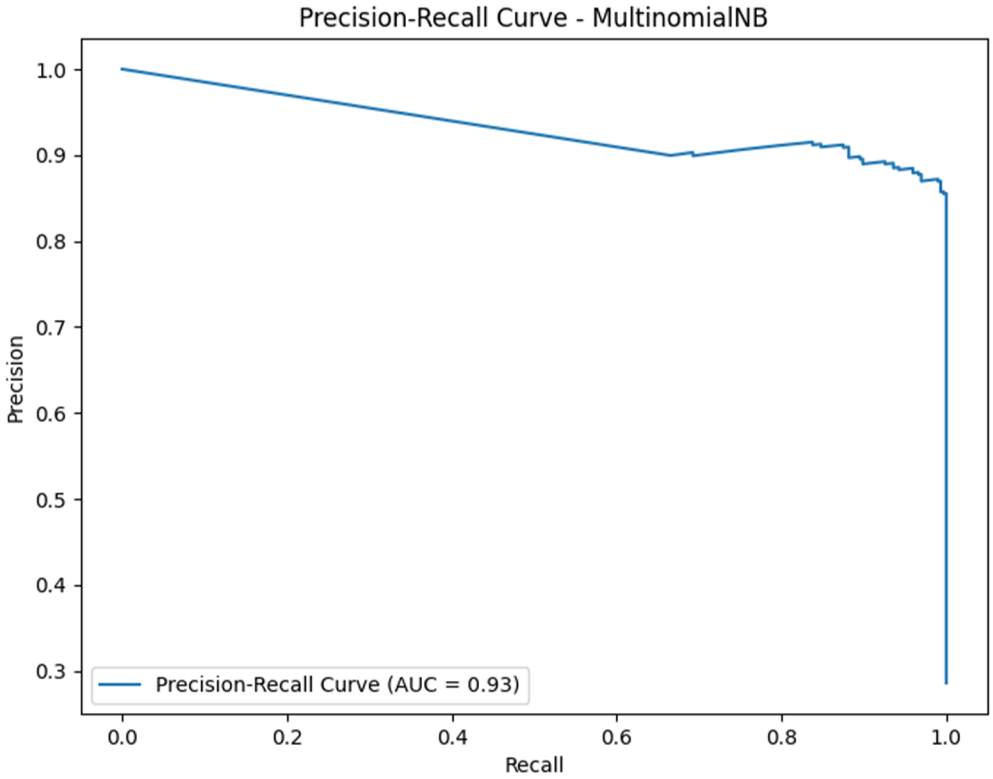
4. RESULT

The results from the Multinomial Naive Bayes model indicate strong performance in identifying spam and non-spam emails, as demonstrated by the metrics derived from the confusion matrix and classification report.
The confusion matrix shows:
702 true negatives (TN): Emails correctly identified as non-spam.
283 true positives (TP): Emails correctly classified as spam.
37 false positives (FP): Non-spam emails misclassified as spam.
13 false negatives (FN): Spam emails that were missed and classified as non-spam.
The classification report provides a deeper understanding of the model's performance:
For non-spam emails (class 0):
Precision of 0.98 means that 98% of emails classified as non-spam were indeed non-spam.
Recall of 0.95 indicates that 95% of all actual non-spam emails were correctly identified.
The F1-score of 0.97 confirms an excellent balance between precision and recall.
For spam emails (class 1):
Precision of 0.88 shows that 88% of emails identified as spam were actually spam.
Recall of 0.96 indicates that 96% of actual spam emails were correctly identified.
The F1-score of 0.92 reflects a strong performance in identifying spam emails.
The overall accuracy of the model is 95%, indicating that the vast majority of emails were classified correctly. The macro average F1-score of 0.94 suggests that the model performs well across both classes, without bias toward either spam or non-spam emails. The weighted average F1-score of 0.95 further confirms the model's robustness, taking into account the class distribution.

Finally, the individual F1-score of 0.9188 highlights the effectiveness of the model in balancing precision and recall.


Email Distribution (Spam vs Non-Spam)
Count
3500
3000
2500
2000
1500
1000
500
0
Non Spam
Spam
Category


Confusion Matrix - MultinomialNB
Actual
Non Spam
Spam
702
37
13
283
Non Spam
Spam
Prediction
700
600
500
400
300
200
100




```
MultinomialNB - Classification Report:
              precision    recall  f1-score   support

           0       0.98      0.95      0.97       739
           1       0.88      0.96      0.92       296

    accuracy                           0.95      1035
   macro avg       0.93      0.95      0.94      1035
weighted avg       0.95      0.95      0.95      1035


MultinomialNB - F1-Score: 0.9188311688311688
```