# Spam Email Detection Using Machine Learning

## Abstract

Email communication has become an essential part of personal and professional life. However, spam emails pose a significant threat to users by promoting scams, phishing attacks, and malware distribution. This research focuses on developing a Spam Email Classification System using machine learning techniques. We employ the Naïve Bayes algorithm as our final model and utilize the TF-IDF vectorizer for feature extraction. The model achieves an accuracy of 97.22%, demonstrating effective classification of spam and non-spam emails.

## 1. Introduction

With the rise of digital communication, the volume of spam emails has increased dramatically. These unwanted emails not only clutter inboxes but also pose security risks. Traditional spam filters rely on rule-based approaches, which often fail to adapt to new spam techniques. Machine learning provides a robust solution by learning patterns from data and classifying emails automatically. This study presents a machine learning-based spam detection system using the Naïve Bayes algorithm.

## 2. Related Work

Several studies have explored spam detection using machine learning algorithms such as Decision Trees, Random Forest, and Support Vector Machines. While these models provide good performance, Naïve Bayes is widely preferred due to its efficiency in handling textual data. Additionally, feature extraction techniques like Term Frequency-Inverse Document Frequency (TF-IDF) enhance classification performance.

## 3. Methodology

### 3.1 Dataset

Our dataset comprises labeled spam and non-spam emails. The data consists of legitimate messages such as bank notifications and personal communications, while spam messages include promotional content and phishing attempts.

### 3.2 Data Preprocessing

The raw email text undergoes preprocessing, including:  
- Tokenization  
- Removal of stopwords  
- Lemmatization  
- TF-IDF vectorization

### 3.3 Model Selection and Training

We evaluated multiple models, including Random Forest and Logistic Regression. However, the Naïve Bayes algorithm was selected as the final classifier due to its superior performance in handling text-based classification.

### 3.4 Deployment

The trained model is integrated into a web-based application using Streamlit, allowing users to input email text and receive real-time classification results.

## 4. Results and Discussion

The model achieves high accuracy in distinguishing spam from legitimate emails. The evaluation metrics confirm the effectiveness of the classifier:

Model Accuracy: 97.22%

### Classification Report:

precision recall f1-score support  
 0 0.97 1.00 0.98 965  
 1 1.00 0.79 0.88 150  
 accuracy 0.97 1115  
 macro avg 0.98 0.90 0.93 1115  
weighted avg 0.97 0.97 0.97 1115

### Confusion Matrix:

[[965 0]  
 [ 31 119]]

## 5. Conclusion

This research demonstrates that machine learning-based spam detection provides a reliable solution to email filtering. The Naïve Bayes model, combined with TF-IDF vectorization, effectively classifies emails with high accuracy. Future work may include deep learning techniques and real-time adaptive filtering to enhance detection capabilities.

## References

1. J. Schmidhuber, "Deep Learning in Neural Networks: An Overview," Neural Networks, 2015.  
2. W. S. Noble, "What is a Support Vector Machine?" Nature Biotechnology, 2006.  
3. T. Joachims, "Text Categorization with Support Vector Machines: Learning with Many Relevant Features," ECML, 1998.