# Spam Detection Using Naive Bayes Classifier

Avas Bajracharya

Department of Computing, East Tennessee State University

CSCI 5037 – Natural Language Processing & Text  
Analysis

Dr. Ahmad Al Doulat

Assignment Due Date: 9/24/2024

# 1. Overview of the Project

## a. Project Type

This main objective of this project is to take a spam collection dataset and apply the natural language processing concepts that we learned in the classroom, then feed the data into a Naïve Bayes classifier model to classify the messages as either spam or legitimate.

## b. Problem to Address

The main problem of this project will be to filter out data such as stop words, punctuations, URL and links etc., to clean and improve the model performance. Besides, this we also need to build a classifier that can accurately predict spam and legitimate messages.

## c. Motivation

With the vast amount of email traffic, spam continues to be a significant issue, both in terms of personal inconvenience and for cybersecurity concerns. Reducing spam helps improve user experience and mitigates the risk of phishing or malicious attacks. This project explores machine learning methods to enhance the accuracy of spam detection systems.

## d. Literature Review

Previous studies have shown that machine learning is highly effective in spam detection tasks. Techniques like decision trees, support vector machines, and deep learning have been explored, but traditional classifiers remain popular due to their simplicity and effectiveness. The ‘A Comprehensive Survey of Spam Detection Techniques’ highlights various approaches used in the field, emphasizing the significance of feature extraction and selection in improving spam detection accuracy.

# 2. Overview of the Dataset

The dataset for this project will be sourced from the SpamAssassin public dataset, which is commonly used in spam detection research. This dataset includes a collection of labeled emails, marked as either spam or ham (legitimate), with various metadata like email headers, subject lines, and body text. The dataset is well-structured for supervised learning and is ideal for training the classifier.

# 3. Rough Tentative Approach to Solving the Problem

The project will proceed with the following steps:  
1. Data Preprocessing: Clean the dataset by removing unnecessary metadata and normalizing text data (lowercasing, removing punctuation, etc.). Tokenization and feature extraction will be applied using techniques like Term Frequency-Inverse Document Frequency (TF-IDF) to convert the text into numerical features.  
2. Model Selection: Implement a suitable machine learning classifier, as text classification requires reliable algorithms for feature learning.  
3. Training and Testing: Split the dataset into training and testing sets, and use cross-validation to evaluate performance.  
4. Evaluation Metrics: Measure accuracy, precision, recall, and F1-score to assess the effectiveness of the model.  
5. Optimization: Explore hyperparameter tuning to improve performance.

# 4. Team Members & Roles

- Alice (Data Scientist): Responsible for data preprocessing and feature extraction.  
- Bob (Machine Learning Engineer): Focused on model implementation and evaluation.  
- Charlie (Project Manager): Manages timeline, resources, and ensures the project meets its milestones.

# 5. Questions the Project Will Answer

- How effective is the chosen machine learning classifier in detecting spam?  
- What are the key features that help differentiate spam from legitimate messages?  
- How can the model be optimized for better accuracy and reduced false positives?

# 6. Things to Learn from the Project

- An understanding of text classification using machine learning techniques.  
- Insights into feature extraction methods that are most suitable for spam detection.  
- How spam detection models can be deployed and tested in real-world applications.

# 7. Conclusion

This project aims to build a reliable spam detection system using machine learning techniques. By leveraging real-world email datasets, we expect to develop a model that accurately classifies spam while maintaining high efficiency. The project will provide valuable insights into the performance of different algorithms in text classification and will help us learn about various optimization techniques to improve the model’s accuracy.