Spam Email Detection

Session: 2022 – 2026

Submitted by:

Mohammad Saad Akmal 2022-CS-148

Submitted to:

Sir Nauman Shafi

Department of Computer Science
University of Engineering and Technology
Lahore, Pakistan

Contents

1 Objective

The objective of this project is to build a predictive model using machine learning techniques that can detect unwanted and spam emails. This helps improve email security, reduces the risk of phishing attacks, enhances user experience by keeping inboxes clean

2 Background and Motivation

Spam emails have become a big problem for people and organizations. These are unwanted messages that fill our inboxes with ads, scams, and other junk. As email remains one of the most popular ways to communicate, dealing with spam has become crucial. There are several reasons to work on this project, such as security and improved communication.

3 Proposed Methodology

We will approach the project with the following steps:

3.1 Data Collection

We will collect spam detection data from publicly available datasets, such as those found on the UCI Machine Learning Repository or Kaggle.

3.2 Algorithms/Models

We plan to implement decision trees, logistic regression, KNN, and Naive Bayes. We will experiment with different algorithms and select the one that provides the best performance.

3.3 Tools and Technologies

We will use Python with libraries such as `scikit-learn`, `TensorFlow`, `spaCy`, `NumPy`, and `Pandas` for data preprocessing, model building, and evaluation.

3.4 Evaluation Metrics

The success of the models will be evaluated using metrics such as accuracy, precision, recall, F1-score, and ROC-AUC curve, with a focus on balancing recall and precision for imbalanced data.

4 Expected Outcomes

We expect to develop a predictive model with an accuracy of at least 75% in identifying spam emails. One potential challenge is handling imbalanced data, as spam emails are often less frequent compared to legitimate emails.

5 References

- Email Spam Detection Using Machine Learning Algorithms, IEEE `https://ieeexplore.ieee.org/abstract/document/9183098`

6 Conclusion

By the end of this project, we want to create a working spam email detection model that can accurately identify spam emails in real-time. This model will help reduce the amount of spam that gets into users' inboxes and ensure that important emails are not missed.