Abdullah Rehman

F2021266327

Spam Dataset

This proposal describes the initiative to look at the "Spam Emails" dataset. The goals of the project are to develop a machine learning model for spam detection, analyze the characteristics of spam emails, and evaluate the efficacy of the model.

The dataset utilized in this study came from Kaggle's "Spam Emails" collection. Resources to complete this project successfully.

Kaggle Dataset: "**Spam Email**"

Programming language: "**Python**"

 Libraries: "**Pandas, Scikit-Learn and Matplotlib**"

Computing resources: "**PC with sufficient RAM and processing power**"

The project will begin with a thorough examination of the dataset. This means looking at the distribution of elements, identifying missing values, and showing the relationships between different elements. Feature engineering techniques will be used to extract more features that can improve the performance of the model.

Model choice and training: Several machines learning techniques, including Random Forest, Naive Bayes, and Support Vector Machines (SVM), will be considered for spam identification.

The final model will be chosen based on a variety of factors, such as interpretability, processing efficiency, and accuracy. The model will be trained on the dataset using appropriate techniques, including cross-validation, to avoid overfitting.

The results of the analysis will be evaluated in order to understand the components that go into spam emails.

Through the analysis of Kaggle's "Spam" data, this project aims to understand the characteristics of spam emails. In this research, we aim to create an efficient spam detection model to strengthen user protection against spam emails and improve spam filtering systems.