Spam Email Detection Using Logistic Regression Approach

Abstract:

Spam email detection is a critical problem in the field of machine learning. In this approach,
we use logistic regression to build a model that can predict whether an email is spam or not. The logistic regression algorithm is well-suited for binary classification problems and can be
trained on a labeled dataset of spam and ham emails. The approach involves collecting a dataset of labeled emails, preprocessing the text data using techniques such as Bag of Words or TF-IDF, splitting the preprocessed data into
training and testing sets, training a logistic regression model on the training set, evaluating
the performance of the model on the testing set using metrics such as accuracy, precision,
recall, and F1 score, tuning the model's hyperparameters, and deployingthe trained model
to detect spam emails in real-time. The logistic regression approach is aneffective and efficient way to build a spam email detection system that can help prevent email spam from cluttering up ourinboxes. Our
approach achieved very high which accuracy 96%.