

NEIL GOGTE INSTITUTE OF TECHNOLOGY

Department of Computer Science and Engineering

MINI PROJECT

Abstract

Spam Detection

Introduction:

In today's digital era, spam messages pose a significant threat to online communication by cluttering inboxes and delivering malicious content. Effective spam detection is essential for filtering out irrelevant or harmful messages while ensuring that legitimate communications are not mistakenly classified as spam. This project aims to develop a machine learning-based spam detection system that can accurately classify messages as spam or non-spam.

Methodology:

The project employs a dataset of spam and non-spam messages to train a classification model. Techniques such as Natural Language Processing (NLP) are used to preprocess the text data, including tokenization, stemming, and removing stopwords. Feature extraction methods like TF-IDF (Term Frequency-Inverse Document Frequency) and word embeddings help in understanding the text content. Various machine learning algorithms, including Naïve Bayes, Logistic Regression, and Support Vector Machines (SVM), are tested to determine the most accurate model for spam detection. The model's performance is evaluated using accuracy, precision, recall, and F1-score metrics.

Expected Results:

The expected outcome of this project is an efficient and accurate spam detection system that can classify messages with high precision. The model should be able to differentiate between spam and non-spam messages effectively, reducing the likelihood of false positives and false negatives. This system can be integrated into email services or messaging platforms to enhance user experience by minimizing spam-related disruptions.

Conclusion:

Spam detection is a crucial aspect of digital security and user convenience. By leveraging machine learning techniques, this project aims to develop a robust spam classification system. The findings from this project can serve as a foundation for further improvements in automated spam filtering technologies.

Keywords:

Spam Detection, Machine Learning, Natural Language Processing, Text Classification, Email Filtering

Section: CSE-B**Batch No.: A1****Names of the Students**

S. Nithin, 245322733112

Adarsh, 24532273370

Name of the Guide

Dr. T. R. Srinivas