

Building a Smarter AI-Powered Spam Classifier

Abstract :

This research project focuses on the development of a smarter AI-powered spam classifier, which aims to enhance the accuracy and efficiency of spam detection in digital communication platforms. Spam emails, messages, and comments have become increasingly sophisticated, making traditional spam filters less effective. To address this challenge, we propose a multi-module approach that combines various machine learning techniques, natural language processing (NLP) algorithms, and user behavior analysis.

Module

Module 1: Data Preprocessing

In this module, we preprocess and clean the incoming data, transforming it into a structured format suitable for analysis. This includes text normalization, removal of HTML tags, and handling of special characters. We also extract relevant features such as sender information, message content, and timestamps.

Module 2: Content Analysis

Using advanced NLP techniques, this module analyzes the content of messages to identify spam patterns. We employ tokenization, sentiment analysis, and topic modeling to detect suspicious language and topics commonly associated with spam. Additionally, deep learning models are used to identify subtle linguistic cues.

Module 3: User Behavior Analysis

Spammers often exhibit distinct behavioral patterns. This module profiles user interactions, considering factors like click-through rates, response times, and historical behavior. By analyzing user engagement data, we can identify anomalies that indicate spammy activities.

Module 4: Machine Learning Models

We employ a diverse set of machine learning algorithms, including decision trees, random forests, and neural networks, to classify messages based on the insights gained from the previous modules. These models are continuously trained and updated to adapt to evolving spam tactics.

Module 5: Feedback Loop

To create a self-improving system, we implement a feedback loop that allows users to report false positives and false negatives. This feedback is used to fine-tune our models and improve overall accuracy.

The proposed multi-module approach offers a comprehensive solution to the spam classification problem. By combining data preprocessing, content analysis, user behavior profiling, and machine learning, we aim to build a smarter AI-powered spam classifier capable of adapting to evolving spam tactics and achieving higher accuracy rates in identifying and mitigating spam across various digital communication channels.