Ideation Phase

Defining the Problem Statements

Date	**26-09-2023**
Team ID	**1025**
Project Name	**6112-Building a Smarter AI-Powered Spam Classifier**

AI-POWERED SPAM CLASSIFIER

Problem Definition and Design Thinking

Introduction

The problem is to build an AI-powered spam classifier that can accurately distinguish between spam and non-spam in emails or text messages. The goals is to reduce the number of false positives (classifying legitimate messages as spam) and false negative (missing actual spam messages) while achieving a high level of accuracy

Problem Statement

Objective: Develop an AI-Powered spam classifier using natural language processing (NLP) and Machine learning techniques to accurately distinguish between spam and non-spam messages in email or text messages.

Data: The SMS Spam Collection is a set of SMS tagged messages that have been collected for SMS Spam research. It contains one set of SMS messages in English of 5,574 messages, tagged according being ham (legitimate) or spam.

Key Challenges:

1. Data Quality: Ensuring the dataset is clean, complete, and free of errors.

2. Feature Selection: Identifying the most relevant features for detecting spam messages.

3. Model Selection: Choosing the appropriate machine learning algorithm(s) for the task.

4. Deployment: Creating a user-friendly interface or API for end-users to detect spam message

Design Thinking Approach

Data collection:

We will need a data set containing labelled examples of spam and non-spam messages. We can use a Kaggle data set for this purpose

Data preprocessing:

The text data needs to be cleaned and pre-processed. This involves removing special characters, converting text to lowercase, and tokenizing the text into individual words.

Feature extraction:

We will convert the tokenized words into numerical feature using techniques like TF-IDF(Term Frequency-Inverse Document frequency).

Model Selection:

We can experiment with various machine learning algorithm such as navie bayes, support vector machines and more advanced techniques like learning using neural network.

Evaluation:

We will measure the models performance using metrics like accuracy, precision, recall and f1-score.

Iterative Improvements:

We will fine-tune the model and experiment with hyper parameters to improve it’s accuracy

Prototype

Create a prototype of the machine learning model and the user interface for separating spam and non-spam messages.

Actions:

- Develop a Jupyter Notebook or Python script for data pre-processing, model training, and evaluation.

- Create a simple web interface using tools like Flask or Django to allow users to separate spam and non-spam messages.

- Test the prototype with a subset of the dataset to ensure it meets performance objectives.

Test

Evaluate the model's performance using appropriate dataset and gather feedback from users.

Actions:

- Split the dataset into training and testing sets.

- Train the model on the training set and evaluate it on the testing set.

- Collect user feedback on the web interface for usability and accuracy.

Implement

Once the prototype meets the defined objectives and receives positive feedback, proceed with full implementation.

Actions:

- Train the final machine learning model on the entire dataset.

- Deploy the model as part of a production-ready web application.

- Conduct thorough testing to ensure the application is robust and user-friendly.

Iterate Continuous improvement is essential. Gather user feedback and iterate on the model and interface to enhance accuracy and usability.

Actions:

- Monitor the model's performance and retrain it periodically with updated data.

- Address user feedback and make necessary improvements to the web interface.

Conclusion

In conclusion, the spam classifier system is a vital tool in the ongoing battle against unwanted and potentially harmful email communication. Through the application of machine learning algorithms and natural language processing techniques, this system can accurately identify and filter out spam messages, thus enhancing the user experience and reducing security risks associated with phishing and malicious content. As email continues to be a primary means of communication, the effectiveness of spam classifiers is paramount in maintaining the integrity and security of digital communication. Ongoing

research and development in this field are essential to stay ahead of evolving spamming techniques and ensure that users can enjoy a safer and more productive online experience.