Project - 117: Build a Model to Classify Emails as Spam or Non-Spam

Abstract:

This project aims to create a model that can automatically identify and sort emails as either spam (unwanted) or non-spam (legitimate). This will help clean up users' inboxes and make email management easier.

Outcome of the Project:

The goal is to develop a system that accurately sorts emails into spam or non-spam categories, improving how users handle their emails.

Project Prerequisites:

1. Python 3.7.4

2. IDE: VS CODE

Required Tools:

1. Python: For coding

2. numpy: For data handling

3. pandas: For managing data

4. nltk: For processing text

5. scikit-learn: For building the model

Project Steps:

1. Data Collection and Preparation:

- Gather a set of emails that are already labeled as spam or non-spam.

- Clean and prepare the text by breaking it into words, removing common words, and making everything lowercase.

2. Model Training:

- Create a model using techniques like Naive Bayes or Support Vector Machine (SVM) to tell spam from non-spam.

- Train the model with the prepared email data.

3. Model Evaluation:

- Test the model to see how well it identifies spam and non-spam emails.

- Improve the model by adjusting settings and trying different methods to get better results.

4. Integration:

- Set up the model to work with an email system.
- Ensure it can automatically sort incoming emails as spam or non-spam.

5. Testing and Deployment:

- Test the model with real emails to make sure it works correctly.
- Deploy the system to start sorting emails in real-time.

This project will help users manage their emails more effectively by automatically filtering out spam.

Team members:

Sathvik-2320030235

Manaswi-2320030245

Karthik-232003006