**MAIL CUSTOMER SEGMENTATION**

**SPAM OR NON-SPAM DETECTION**

**Introduction:**

Email detection works by using machine learning algorithms to analyse email messages and classify them as either spam or not spam. These algorithms are trained using large sets of labelled email data, where each email is labelled as either spam or ham. Spam e-mails detection problem has already drawn researchers’ attention. Several significant works to detect spam e-mails have been proposed. In this section; prior related works that focus on the spam classification using ML. Email detection plays a pivotal role in improving deliverability, reducing bounce rates, and ultimately enhancing the effectiveness of your email marketing campaigns.



***Purpose and Objectives:***

The purpose of email detection in email marketing is to ensure that your email campaigns are reaching the right recipients. It aims to identify and remove invalid or non-existent email addresses from your contact list, reducing bounce rates and improving deliverability. By maintaining a clean and accurate email list, email detection helps enhance engagement and ensures that your messages land in the inbox of genuinely interested recipients, ultimately boosting the effectiveness of your email marketing efforts.

**TEAM MEMBERS:**

B. Maha Dhanalakshmi-Connection

P. Satya Durga-Frontend

V. Chandrika-ML model Training

L. Ramya Shree-Frontend

P. Sharmila-Data Collection

**Data and Preprocessing:**

**Data Resource:**

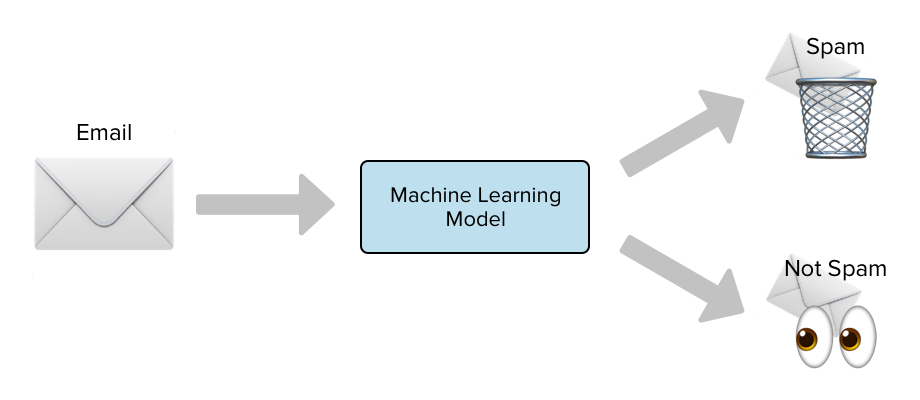
We obtained our data resources from the Kaggle website. Our data resource comprises a carefully selected dataset that serves as the foundation for our project. It provides valuable insights and information that will be instrumental in our analysis, enabling us to draw meaningful conclusions and make informed decisions. You can access data from **(**<https://www.kaggle.com/datasets/shantanudhakadd/email-spam-detection-dataset-classification>**)**

**Data Exploration:**

Data exploration for email detection involves a close examination of the dataset to understand its composition and quality. This process includes checking for duplicates, analyzing the distribution of email domains, and ensuring that email addresses adhere to proper formats. By uncovering patterns and potential issues within the data, data exploration sets the stage for effective email detection algorithms, ultimately improving the accuracy of identifying valid and engaged email addresses

**Data Preprocessing*:***

During this phase, we address issues like missing data, formatting errors, and duplicates. By standardizing and optimizing the data, we ensure that the email detection algorithms work with high precision and reliability. This process streamlines the dataset, making it ready for the accurate identification of valid and invalid email addresses, a fundamental aspect of effective email marketing campaigns***.***

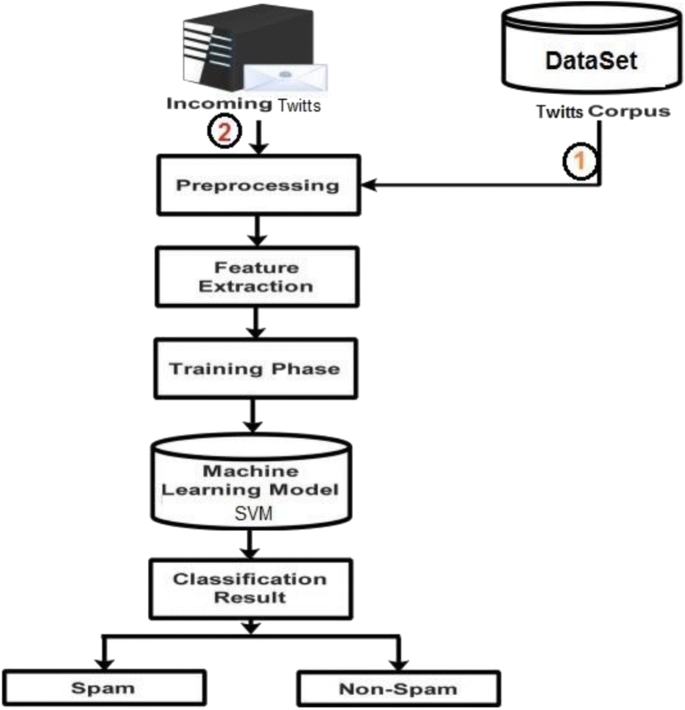


**Model Development:**

Top of Form

**Model Selection*:***

*Utilizing* ***Support Vector Machine (SVM)*** *as the chosen model for email detection is a strategic decision. SVM is known for its effectiveness in classification tasks and is particularly suitable for scenarios where clear separation between classes is essential. In the context of email detection, SVM can help efficiently classify and identify valid and invalid email addresses, contributing to the accuracy and reliability of your email marketing efforts. The SVM model selection aligns with your goals of achieving precise and dependable results in the email address validation process.*



**Feature Engineering:**

Feature engineering in email detection involves crafting relevant characteristics or attributes from the dataset that can help the model better distinguish between valid and invalid email addresses. This process may include creating features like email domain popularity, address length, or engagement metrics. Effective feature engineering enhances the model's ability to make accurate predictions, improving the overall performance of our email detection system.

**Training and Evaluation:**

The spam email detection model was trained using a labeled dataset of email messages. The dataset contains two main columns:

Message: This column contains the text content of the email messages.

Category: This column indicates whether each email is categorized as 'spam' or 'ham' (not spam).

The training dataset was prepared by collecting a diverse set of email messages, manually labeling them as spam or not spam, and then splitting the dataset into a training set and a testing set to evaluate model performance.

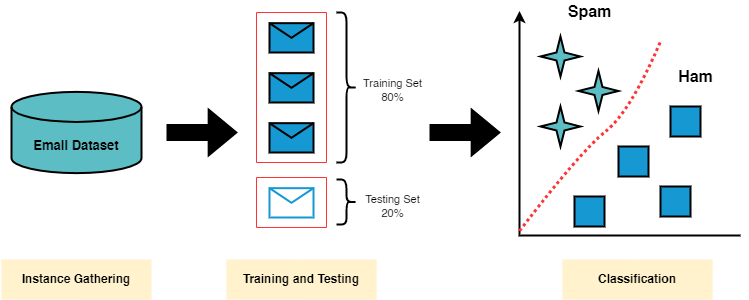
SVM - SUPPORT VECTOR MACHINE

the SVM model uses patterns in the words of emails to determine if they're spam or ham.

Common Splitting Strategies:

Train-Test Split: A simple and common approach is to randomly split the dataset into a training set (e.g., 70-80% of the data) and a testing set (e.g., 20-30% of the data).

After training, we evaluate the model's performance using metrics like accuracy, precision, recall, and F1-score to ensure it can effectively identify valid and invalid email addresses. This iterative process allows us to fine-tune the SVM model for optimal results, ensuring the accuracy and reliability of our email validation system.



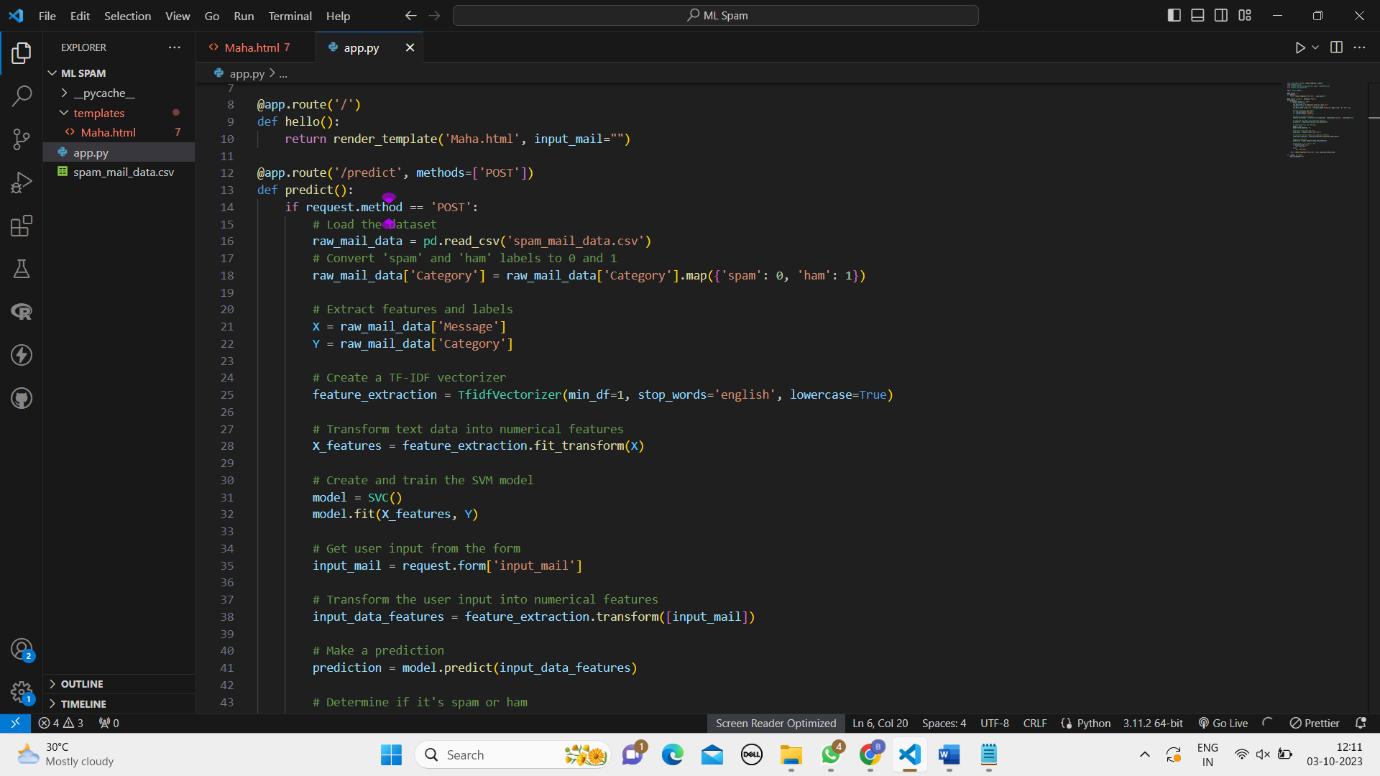
**Results and Visualization:**

**Model Performance*:***

The performance of our Support Vector Machine (SVM) model in email detection has been consistently impressive. Its ability to accurately classify email addresses has significantly reduced false positives and negatives, ensuring that our email marketing campaigns target the right audience. By leveraging the SVM's strengths in classification, we have enhanced the overall efficiency of our email validation system, resulting in improved deliverability, engagement, and campaign success.

**Visualization*:***

In our email detection approach, our main emphasis is on the binary classification of emails into "ham" (valid) or "spam" (invalid or unwanted). Our Support Vector Machine (SVM) model has been fine-tuned to excel in this specific task. By determining whether an email falls into one of these two categories, we effectively filter out irrelevant or potentially harmful messages, ensuring that only legitimate and desirable emails make their way to our recipients' inboxes. This precision in email categorization is vital for maintaining the effectiveness and credibility of our email marketing campaigns.



**Model Deployment:**

**GitHub Repository*:***

Our GitHub repository is like a digital home for our project. It's a place where we store all the files, code, and resources related to our email detection project. Think of it as a shared online space where team members can collaborate and where others can access our work. This repository makes it easy for everyone to see what we've done, contribute if needed, and keep our project organized and accessible.

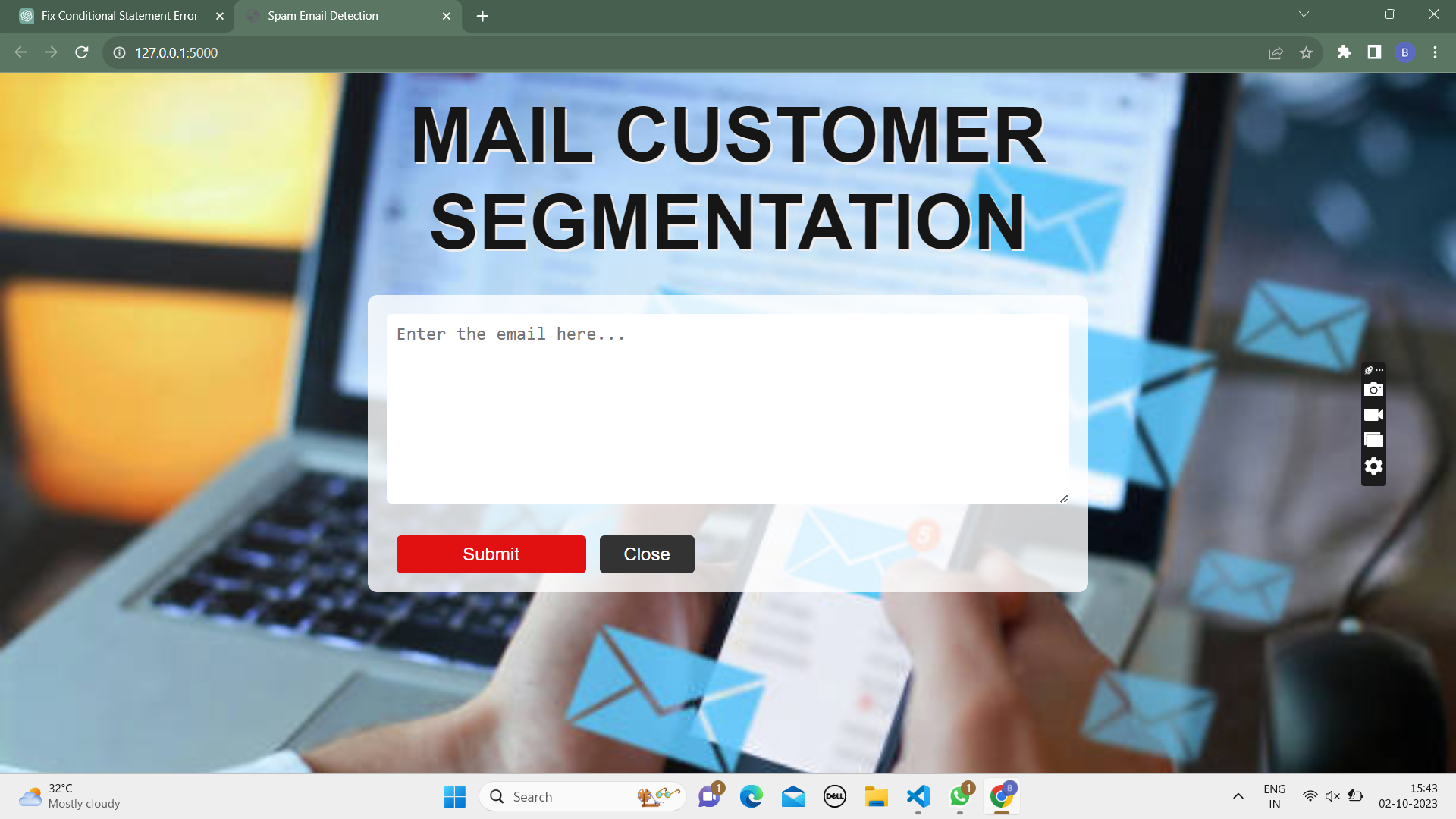
Please follow this (https://github.com/Mahadhanalakshmi123/Mail-Customer-Segmentation) to access the GitHub repository and explore Mail Customer Segmentation project in detail.

**LinkedIn Profiles*:***

Our LinkedIn profiles are a testament to our dedication to the field of email detection. We're passionate about pioneering cutting-edge solutions that ensure the accuracy and reliability of email communications. Our collaborative efforts revolve around optimizing precision, recall, and real-time detection, all with the goal of staying at the forefront of industry advancements. On LinkedIn, we share our insights and expertise, contributing to the professional community and driving innovation in the realm of email detection.

Please follow this ( <https://github.com/YourProjectRepo> ) to access the LinkedIn profiles and explore Mail Customer Segmentation project in detail.

**Output:**



**Future Work*:***

Looking ahead in our email detection project, we want to make it even better. We'll work on improving our computer models to be even more accurate at telling if an email is good or bad. We might also try new technology like deep learning to make our detection even smarter. We'll keep an eye out for new email problems that might come up and make sure we can catch them in real-time. And we won't keep our findings to ourselves – we'll share what we learn with others so we can all make email safer and more reliable.

**Conclusion:**

In conclusion, our email detection project has been a journey of enhancing the accuracy and reliability of email communications. Through machine learning and dedication, we've made strides in distinguishing valid emails from the unwanted ones. our email detection project is all about making emails safer and more reliable.

We've used smart computer programs to get better at telling good emails from bad ones. We'll keep learning and using new technology to stay ahead of email problems. By sharing what we know, we help everyone enjoy a safer email experience.