**Phishing Email Detection System**

Phishing email detection involves identifying fraudulent emails that attempt to trick recipients into revealing sensitive information or taking harmful actions.

* **Senders’ Email Address Analysis**: One of the first steps in phishing detection is to verify the legitimacy of the sender's email address this is done by checking if the domain of the sender's email address matches the claimed identity of the organization or individual because phishing emails often use addresses that closely mimic legitimate ones but with subtle variations or from suspicious domains
* **URL / Link Verification**: Phishing emails often include links that redirect users to fake sites designed to capture sensitive information. This involves checking the domain of the link and ensuring it matches the expected domain of the purported sender.
* **Urgent or Threatening Language**: Phishing emails often use urgent or threatening language to create a sense of urgency, prompting recipients to act quickly without due consideration. Eg: "Immediate action required," "Account suspended," or "Verify your account"
* **Requests for Personal Information**: Be wary of emails that ask for sensitive personal information, such as login credentials financial details. Legitimate organizations typically do not request such information via email.
* **Grammar and Spelling**: Many phishing emails contain noticeable grammar and spelling mistakes such as mismatched fonts, unusual formatting, or awkward phrasing, can also indicate a phishing attempt.

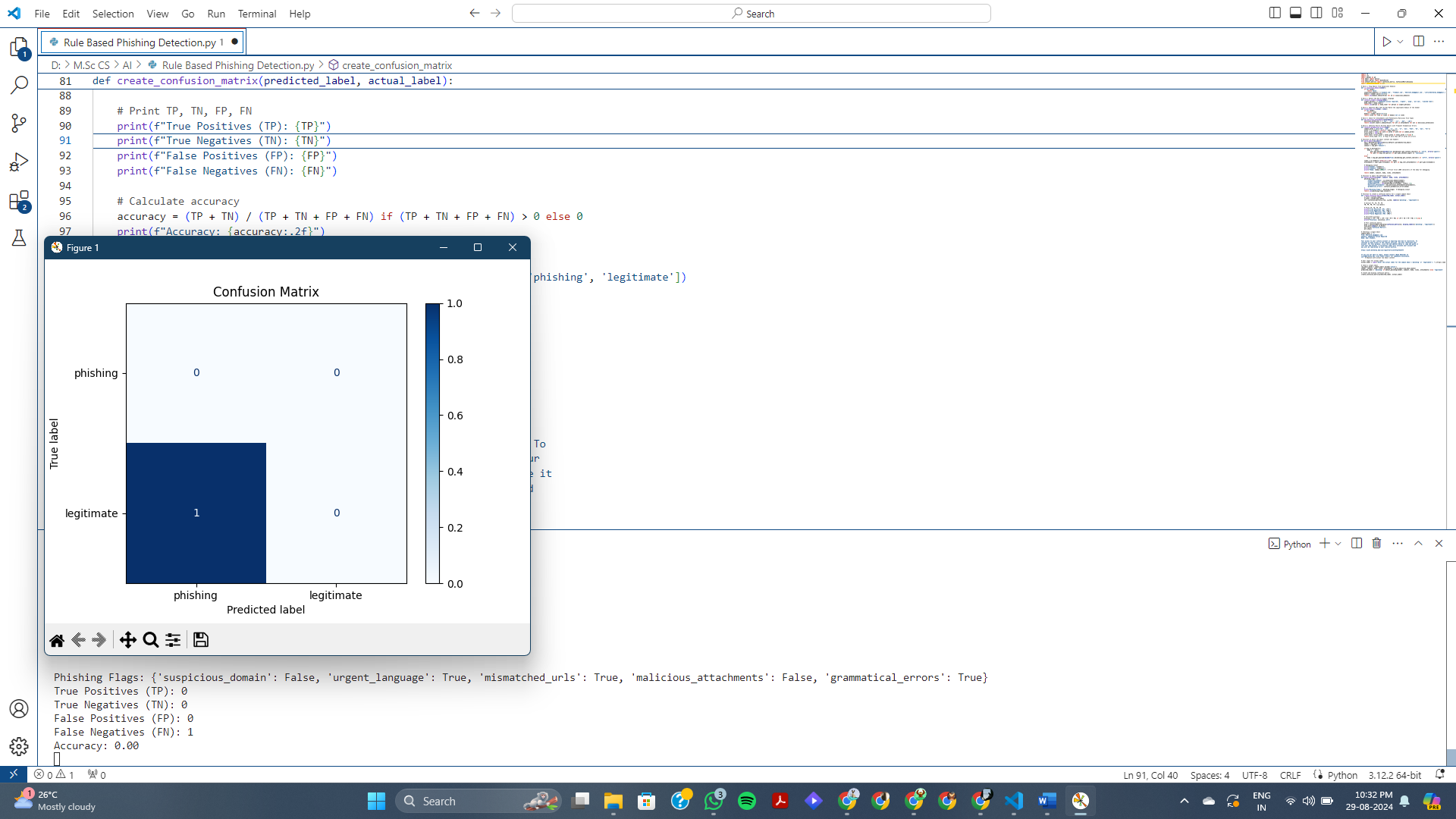
**Rules Implemented in the Phishing Email Detection System**

* Rule 1: Flag emails from suspicious domains (e.g., domains with typos)
  + Phishing emails often uses domain names that closely mimic legitimate ones but with subtle variations or use suspicious domains
* Rule 2: Detect the use of urgent language (e.g., immediate action required, account expiration)
  + Phishing emails often use urgent or threatening language to create a sense of urgency, prompting recipients to act quickly without due consideration.
* Rule 3: Identify URLs that do not match the legitimate domain of the sender
  + Phishing emails use redirects to mask the true destination of a link.
* Rule 4: Check for attachments with potentially malicious file types (e.g., .exe, .zip)
  + Check the file extensions of attachments to determine if they are known to be associated with malware.
* Rule 5: Identify poorly written emails with frequent grammatical errors
  + Many phishing emails contain noticeable grammar and spelling mistakes such as mismatched fonts, unusual formatting, or awkward phrasing, can also indicate a phishing attempt.

**Evaluation Results with Explanations**

* True Positive Rate (TPR): An instance for which both predicted and actual values are positive
* False Positive Rate (FPR): An instance for which predicted value is positive but actual value is negative
* True Negatives (TNR): An instance for which predicted value is negative but actual value is positive
* False Negatives (FNR): An instance for which both predicted and actual values are negative
* Accuracy: It gives the percentage of correct classifications

(TP+TN)/(TP+FP+FN+TN)



**Suggestions for Improving the System**

* Use of larger dataset for phishing email detection system
* Use of ML model for better detection