Phishing Detection in Emails Using Lightweight Machine Learning Models for Edge Devices

1. Objective:

A phishing email is a common social engineering method that mimics trustworthy email addresses and content to deceive recipients. The objective of this project is to train machine learning models on a curated dataset to predict phishing emails. Both phishing and legitimate emails are collected to form the dataset, and from them, relevant features based on email headers, content, and metadata are extracted. The performance of each model is measured and compared to evaluate their effectiveness in detecting phishing emails.

The required packages for this notebook are imported when needed.

2.Data Collection:

The set of phishing emails are collected from opensource service called Kaggle [1]. This service provides a set of phishing URLs in multiple formats like csv, json etc. that gets updated hourly.

The set of legitimate emails are also obtained from same opensource service. Some of them are also obtained from our personal emails.

3. Data Cleaning and Feature Extraction:

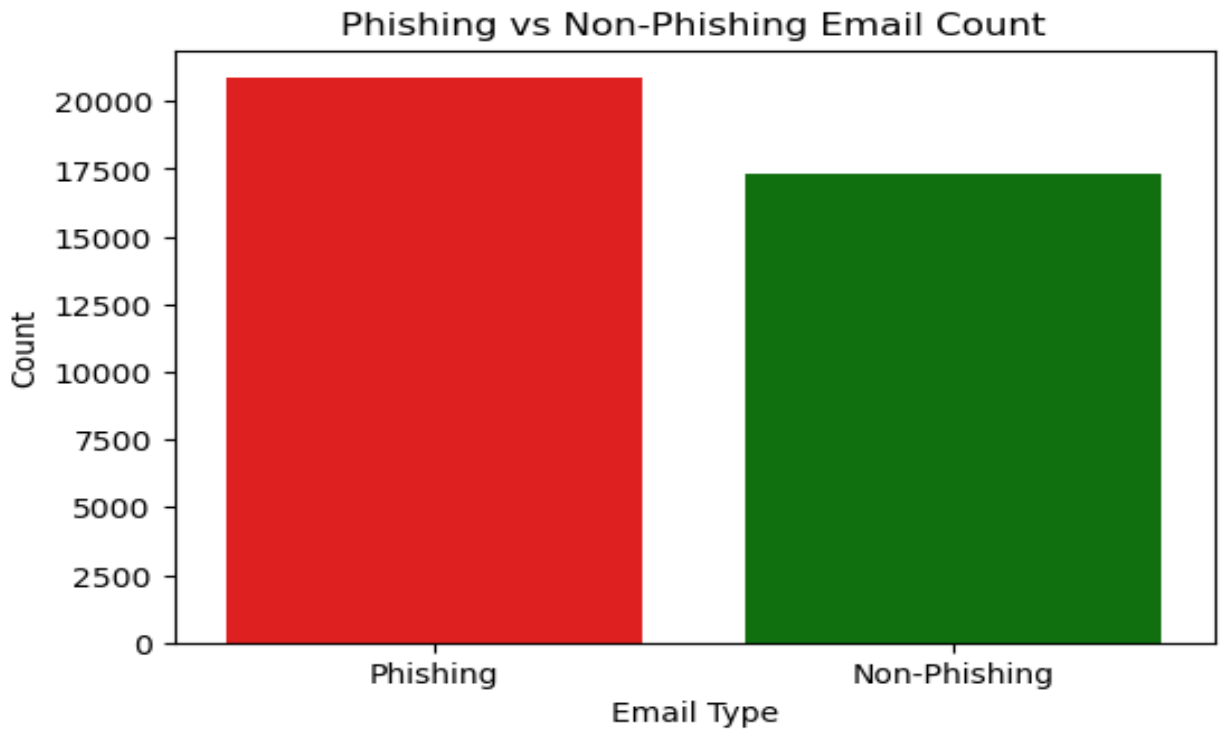
In our project titled "Phishing Detection in Emails Using Lightweight Machine Learning Models for Edge Devices", the data cleaning process was carried out using Pandas within a Jupyter Notebook environment. This phase involved removing duplicate records, handling missing or incomplete data, normalizing textual content, and encoding categorical labels. Unnecessary characters such as HTML tags, URLs, and special symbols were stripped from email texts to reduce noise. Additionally, stop words were removed and text was standardized to lowercase for uniformity. Numerical features were also normalized to ensure consistency across the dataset. This thorough cleaning process helped prepare high-quality input data, which is crucial for training efficient and accurate lightweight machine learning models suited for edge deployment as discussed in [2] and [3].

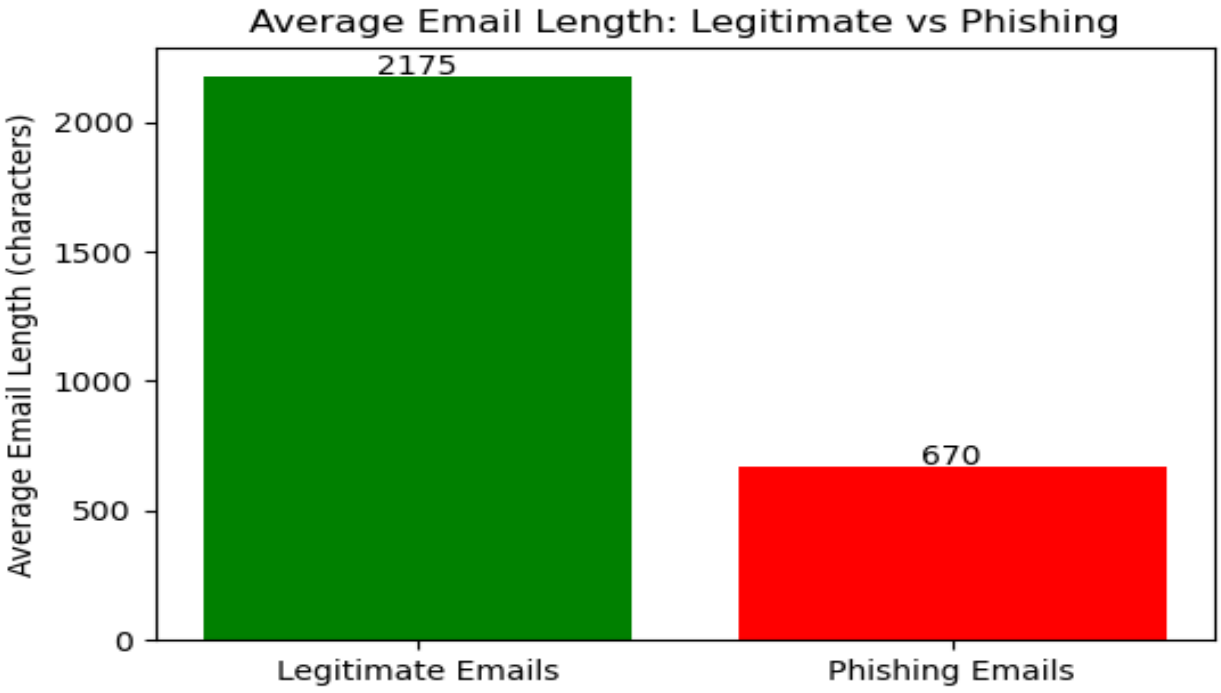
The below Mentioned Category of features are extracted and used:

1. URL
2. Sender Name
3. Sender Domain
4. Text (Subject + Body)
5. Label (Phish / legit)
6. Hour
7. Day of Weeks

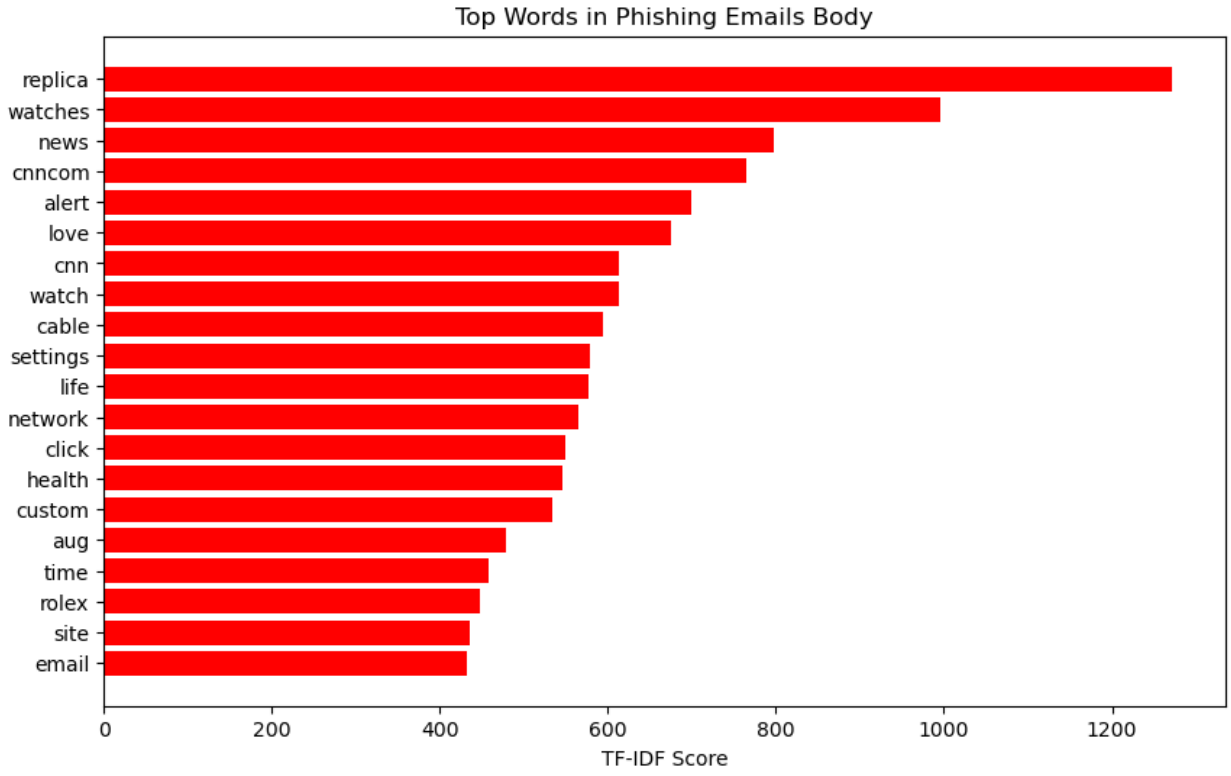
4. Data Analysis:

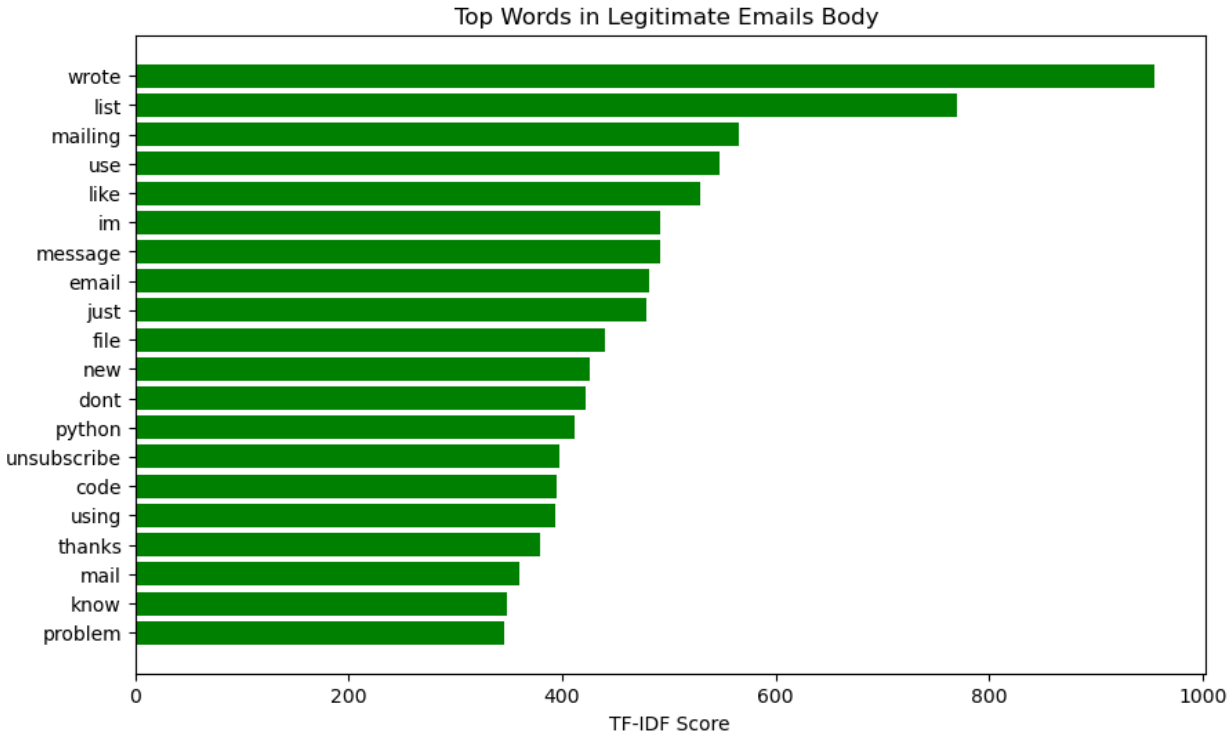
The Analysis is done on behalf of the extracted features.



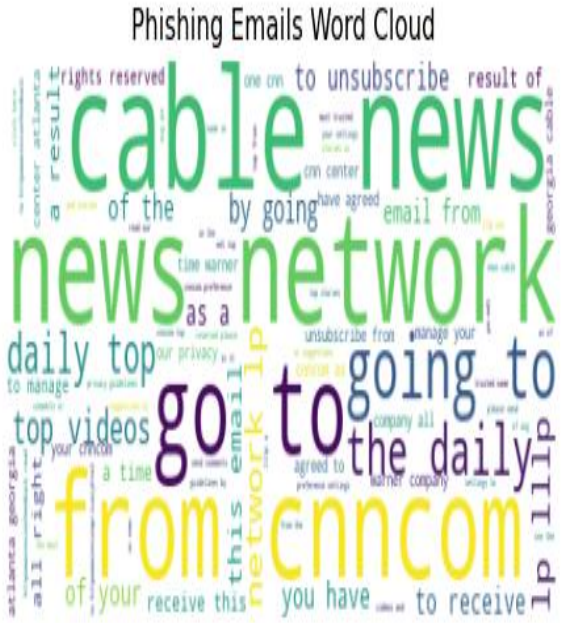


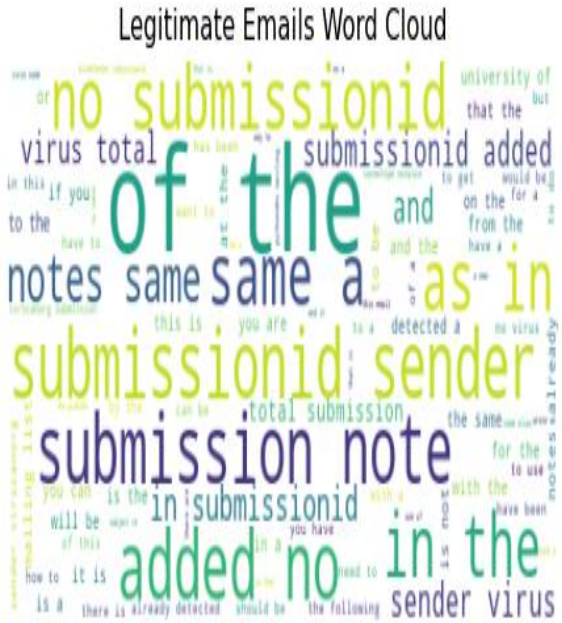
4. Data Analysis:

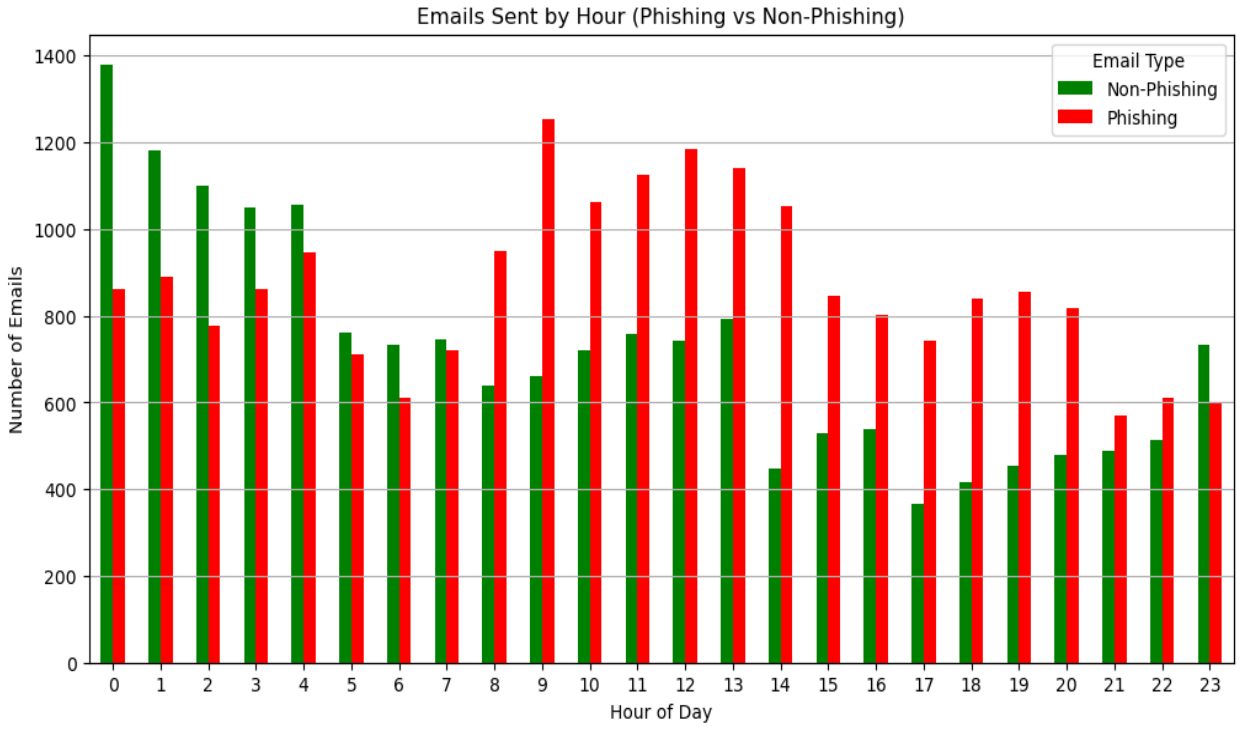




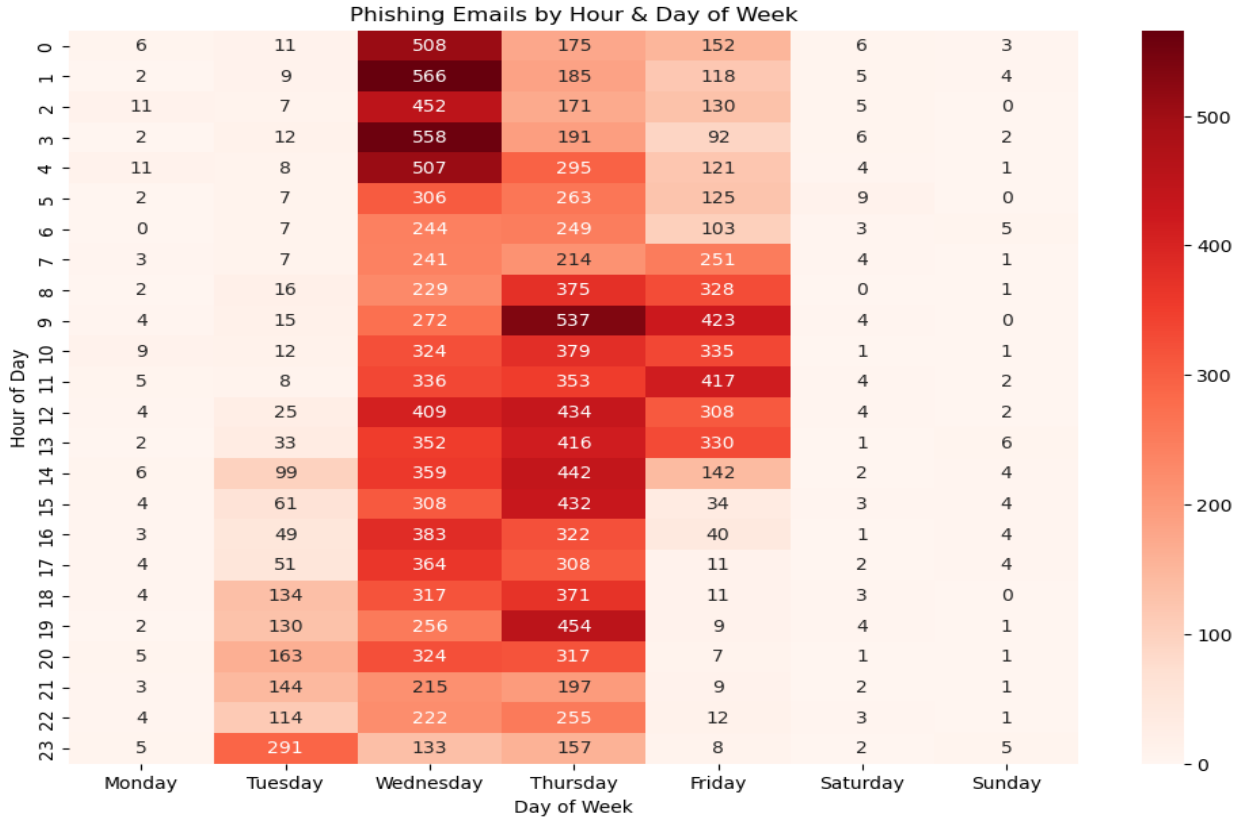
4. Data Analysis:

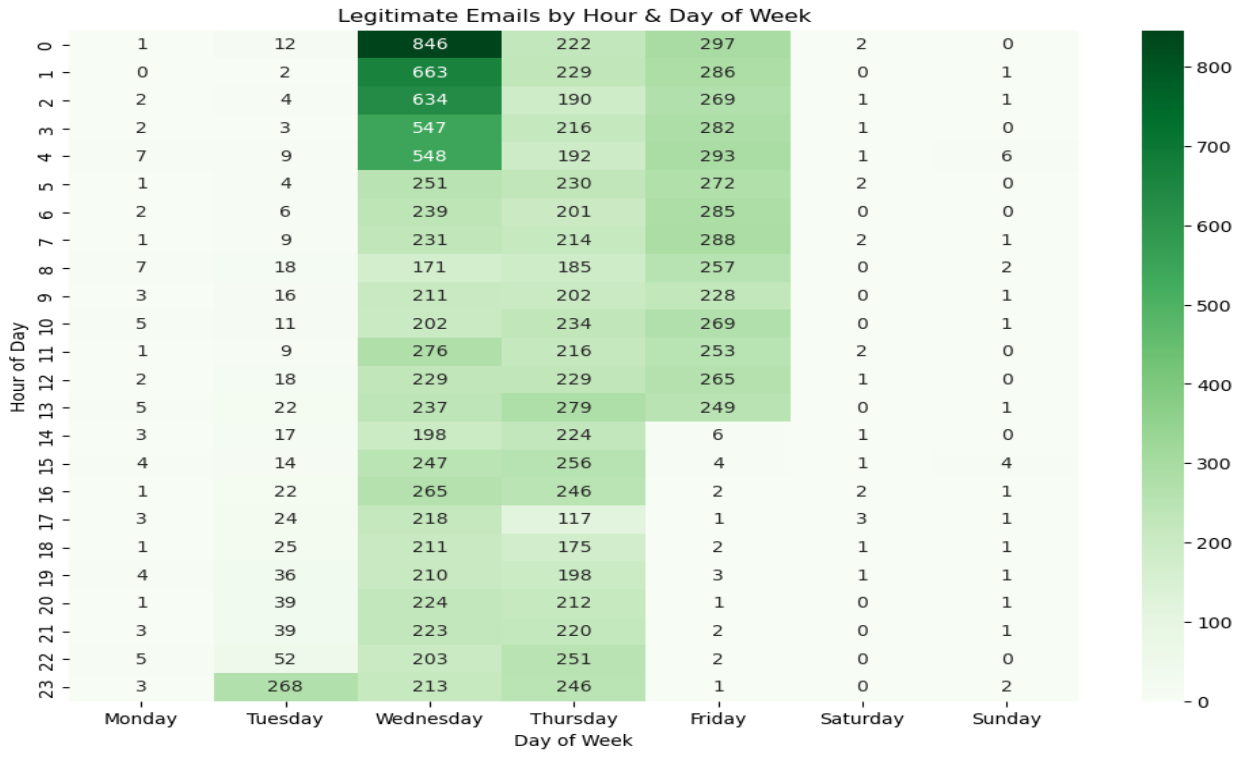




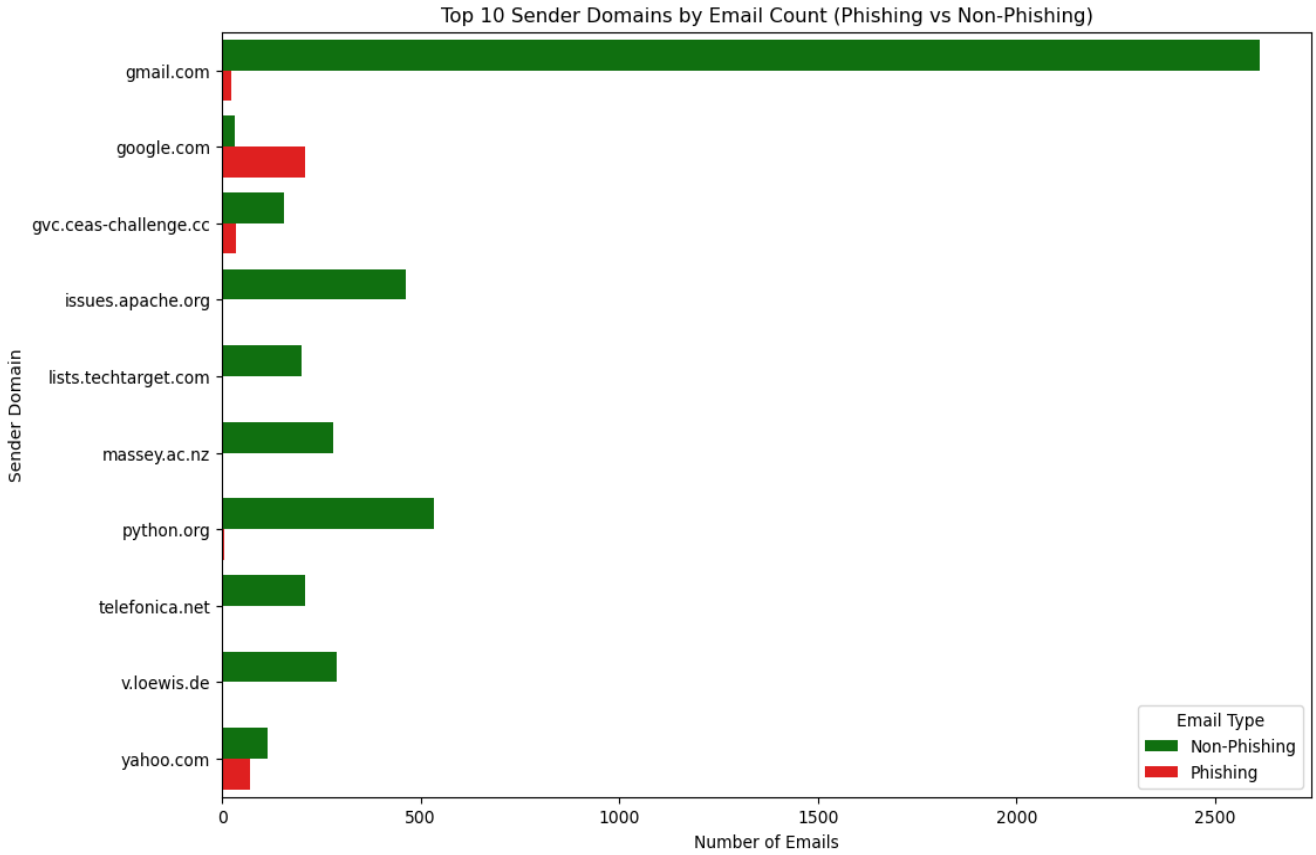


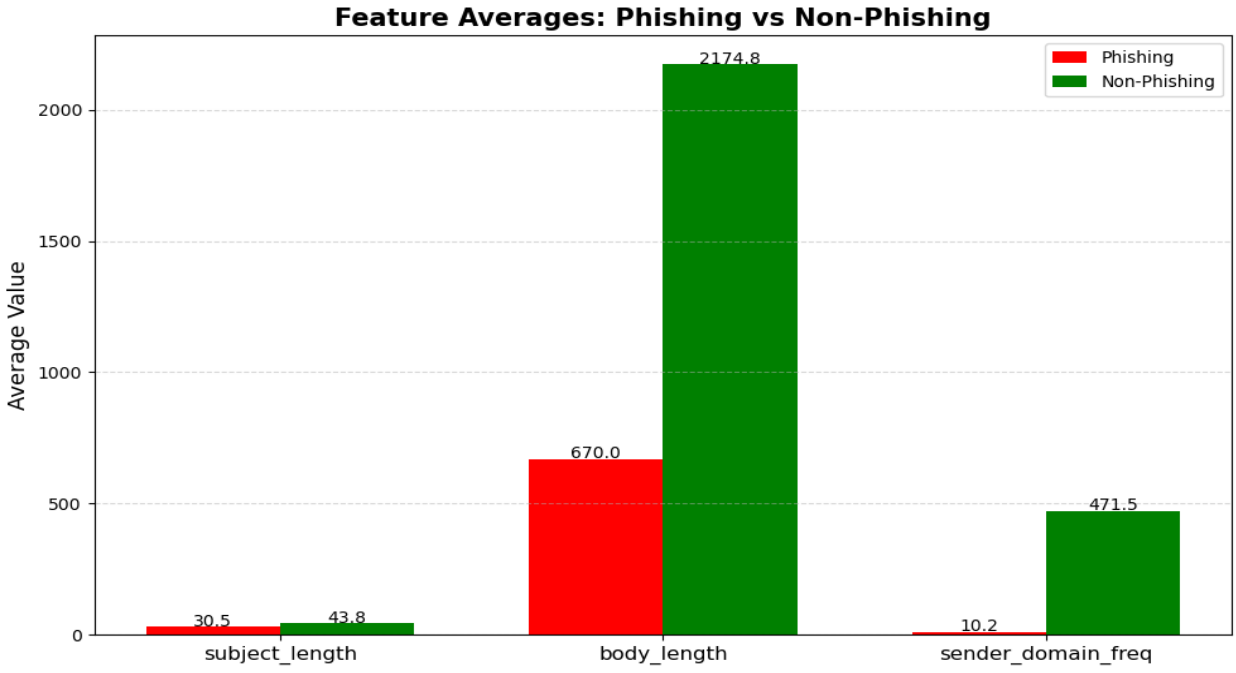
4. Data Analysis:



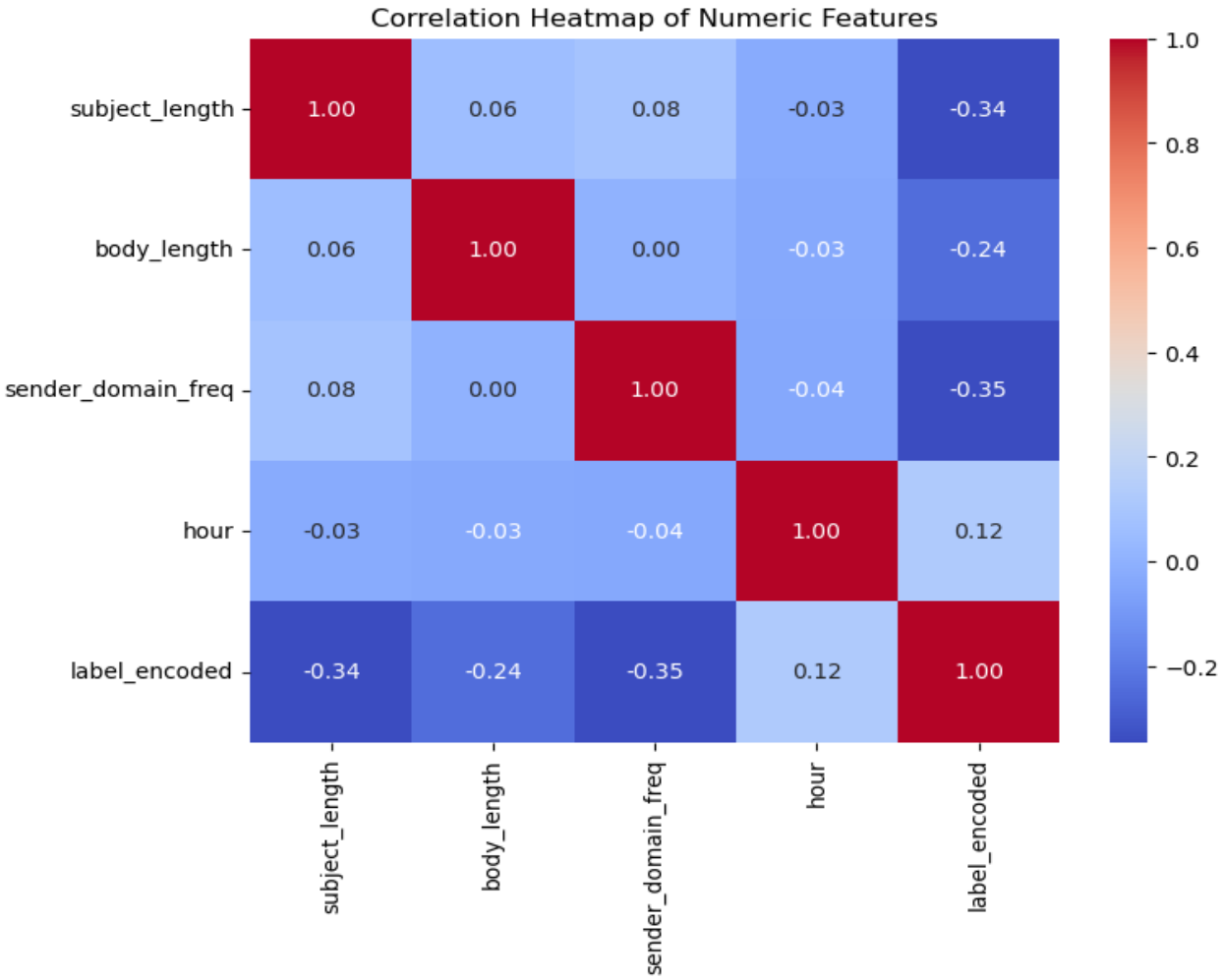


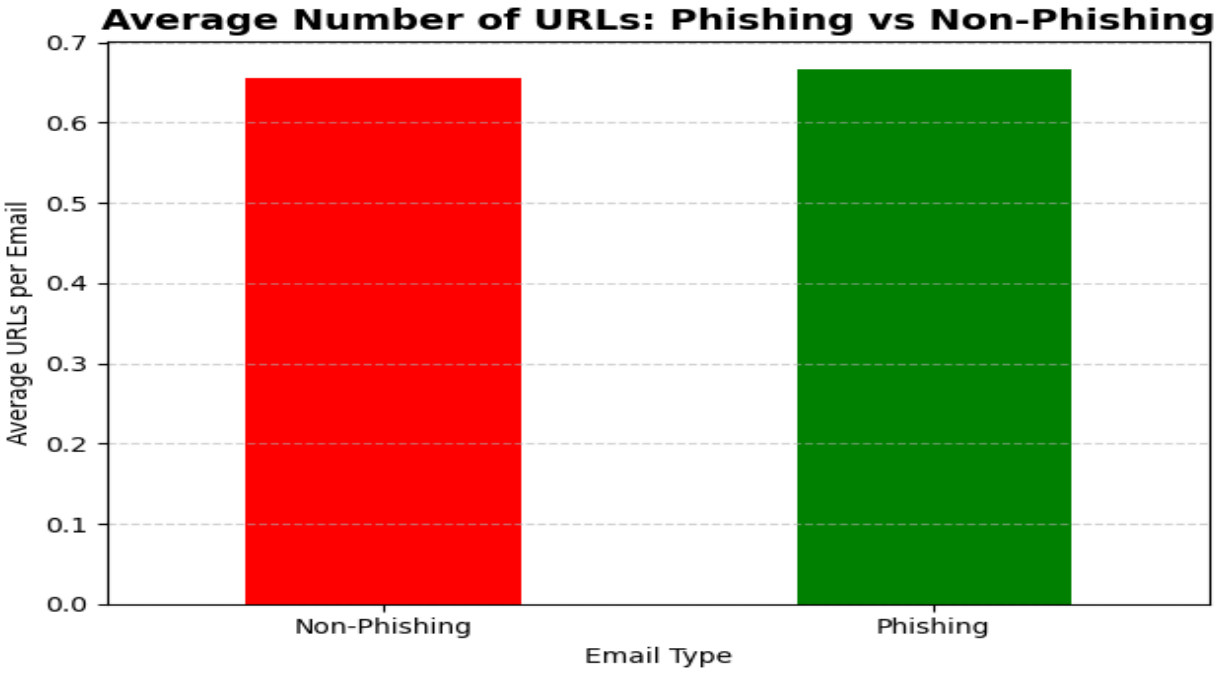
4. Data Analysis:





4. Data Analysis:





5. Conclusion:

The above visualizations provide a clear comparison between phishing and legitimate emails, highlighting key differences in their textual content and feature distributions. This analysis helps in understanding the patterns and characteristics commonly associated with phishing emails, which is crucial for building effective detection models.

6. References:

[1] Kaggle, “Phishing Email Detection Dataset,” : https://www.kaggle.com/datasets/naserabdullahalam/phishing-email-dataset/data

[2] S. Gopal, “Shreya Gopal Sundari”: https://github.com/shreyagopal/Phishing-Website-Detection-by-Machine-Learning-Techniques

[3] OpenAI, “ChatGPT: Language Model for Conversational AI,” https://chat.openai.com/