BEC Email Classifier Prototype - Report

1. Introduction

Business Email Compromise (BEC) is a growing threat in the cybersecurity landscape. With five major categories- CEO fraud, account compromise, false invoice schemes, attorney impersonation, and data theft- these attacks use seemingly benign emails to infiltrate organizations. This project focuses on developing a machine learning-based email classifier to detect and prevent such attacks.

BEC Email Classifier Prototype - Report

2. Project Objectives

The aim of this project is to build a machine learning model using scikit-learn to classify emails as either benign or malicious. The system will focus on textual analysis of emails and apply Logistic Regression to identify suspicious patterns.

BEC Email Classifier Prototype - Report

3. Data Processing

- Data Collection: Emails are loaded from directories for spam and ham (non-spam) emails.

- Preprocessing: Emails are converted into a structured format (DataFrame), then tokenized and vectorized using CountVectorizer, which converts text into numerical data for model training.

BEC Email Classifier Prototype - Report

4. Model Development

- Model: Logistic Regression is chosen for its simplicity and effectiveness in binary classification.

- Training: The model is trained on 80% of the data and evaluated on the remaining 20%.

BEC Email Classifier Prototype - Report

5. Model Evaluation

- Accuracy: The model achieved an accuracy score of X% (replace with actual score after running).

- Confusion Matrix: Provides insights into how many emails were correctly classified as benign or malicious.

- Classification Report: Further details precision, recall, and F1-score for the model's performance.

6. Feature Importance

After training, the model's coefficients are analyzed to determine the most influential words in classifying an email as spam or ham. This provides valuable insights into how the system makes its decisions.

7. Conclusion and Future Work

While this model is a basic prototype, it demonstrates the potential for machine learning to assist in email security. Future work could involve expanding the model to consider more complex features, such as headers, hyperlinks, and attachments, or exploring more advanced models like neural networks.