Introduction:

- **Who I Am:** Koray Aman Arabzadeh, aspiring to innovate in cybersecurity from Mid Sweden University.
- **Project focus:** Combating phishing and social engineering using ML and NLP.
- **Project Goal:** To challenge and innovate in the fight against phishing using ML and NLP.
- **Why It Matters:** Highlighting the importance of evolving cybersecurity to match evolving threats.

Date: 2024-04-01

Problem Background and Motivation

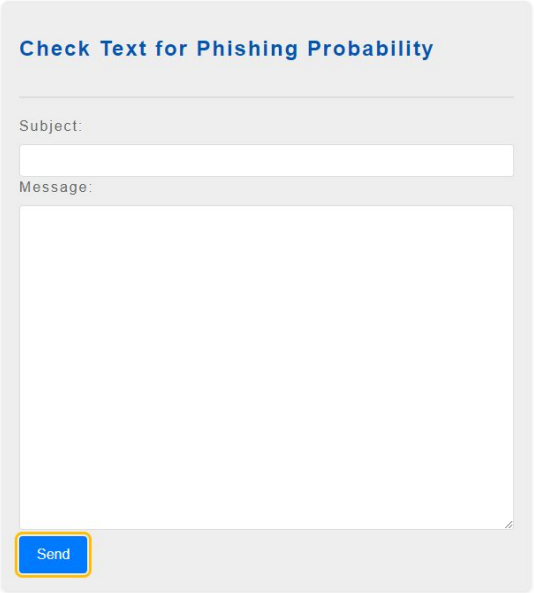
- **Urgency:** Address sophisticated phishing and social engineering threats.
- **Innovation:** Test and integrate ML and NLP to evaluate email threats.
- **Objective:** Enhance digital communication security and predictability.

Theoretical Framework

- **Strategy:** Combine ML's analytical power with NLP's linguistic analysis.
- **Purpose:** Form a proactive defense against future cyber threats.
- **Vision:** Establish new benchmarks in cybersecurity tactics.

Methodology Overview

- **Process:** Curate balanced dataset with ChatGPT 3.5 and Gemini Google AI.
- **Implementation:** Preprocess data, train on advanced algorithms.
- **Outcome:** Develop a Flask-based web application for real-time detection.


Check Text for Phishing Probability
Subject:
Message:
Send

Results

- **Overview:** Promising advancements in email security demonstrated.
- **User testing:** High accuracy with Logistic Regression in detecting phishing.
- **Importance:** Balanced dataset and robust models enhance system potential.

Testing Insights

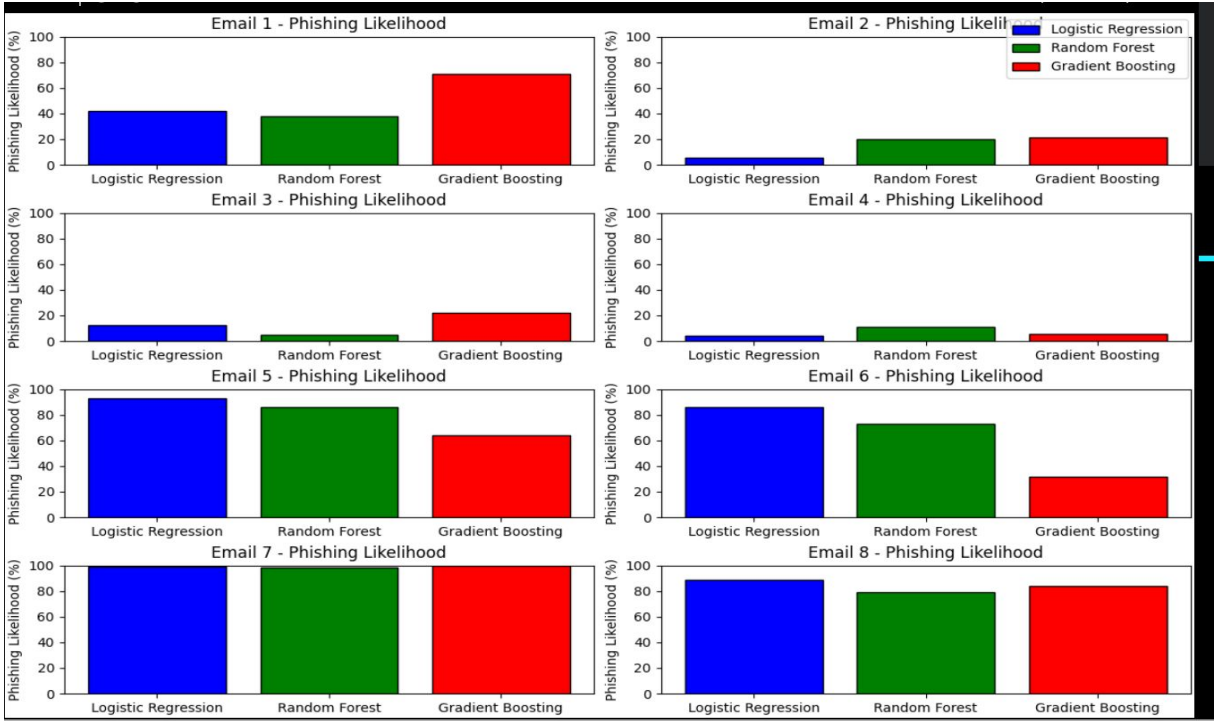
- Real-world email samples tested by two independent testers.
  - Test 1: Legitimate Emails
    - Tester 1: 17.45% phishing probability.
    - Tester 2: 8.19% phishing probability.
  - Test 2: Phishing Attempts
    - Tester 1: 99.04% phishing probability, showcasing high detection accuracy.
    - Tester 2: 49% phishing probability, indicating room for improvement, needs larger datasets to train on.

Feedback Highlights:

- **Tester insights:** Advanced algorithms and balanced datasets are key.
- **Importance:** Refining models based on user feedback for future enhancements.

Comparative Analysis

- **Observation:** Logistic Regression and Random Forest consistent; GradientBoostingClassifier sensitive.
- **Issue:** Potential overfitting in GradientBoostingClassifier.
- **Solution:** Expand testing and dataset for greater model generalization



Future Directions & Conclusion

- **Reflection:** Necessity for adaptable cybersecurity tools highlighted.
- **Future focus:** Expand dataset, refine models, enhance public cybersecurity awareness.
- **Commitment:** Develop accessible, community-focused cybersecurity tools.

Closing

Thank You & Questions