Phishing Patterns through Machine Learning

1. Introduction

Phishing is one of the most prevalent cyber threats, manipulating users into compromising sensitive information. This project aims to analyze the structure and content of phishing emails to extract meaningful patterns using machine learning techniques. By doing so, we hope to contribute to better detection mechanisms and increase awareness about phishing tactics.

2. Objectives

- Analyze textual content from phishing email subjects and bodies.
- Identify recurring features and linguistic patterns common in phishing attempts.
- Develop a basic machine learning classifier to flag potential phishing emails.
- Generate visual and statistical insights to support security awareness.

3. Literature Review

3.1 Traditional ML Techniques

- Naive Bayes: Often used for spam detection due to its simplicity and effectiveness on text data.
- Support Vector Machines (SVM): Performs well on high-dimensional feature spaces like TF-IDF.
- Decision Trees and Random Forests: Offer interpretability, useful in security contexts.

3.2 Recent Deep Learning Approaches

- Recurrent Neural Networks (RNNs): Capture sequence structure in text but struggle with long dependencies.
- Long Short-Term Memory (LSTM): Improves over RNNs, often used in text classification.
- BERT and Transformers: State-of-the-art models in NLP. Capable of contextual understanding, outperforming traditional techniques in many text-based classification tasks.

4. Methodology

4.1 Data Preprocessing

- Removal of HTML tags, URLs, punctuation.
- Tokenization, lemmatization.

- Vectorization using TF-IDF or word embeddings.

4.2 Exploratory Data Analysis

- Average subject line length: [ ]
- Common keywords in phishing emails: [ ]
- Frequency distribution of spam trigger words.
- Visualizations: Word clouds, bar charts, heatmaps.

5. Model Development

5.1 Feature Extraction
- Text vectorization using [e.g., CountVectorizer, TF-IDF].
- Optional: Feature selection to reduce dimensionality.

5.2 Classifier Training
- Model Used: [e.g., Logistic Regression, Naive Bayes]
- Training/Validation Split: 80/20
- Evaluation Metrics: Accuracy, Precision, Recall, F1 Score

6. Findings & Insights

- Phishing emails often use [e.g., "urgent", "action required", "account suspended"] in subject lines.
- Linguistic tone is often threatening or manipulative.
- Visual patterns:

7. Recommendations

- Enhance spam filters using keyword and ML-based detection.
- Implement awareness training highlighting common phishing patterns.
- Adopt ML models that evolve with new data over time.

8. Conclusion

This project outlines a foundational approach to understanding phishing patterns through machine learning. With further refinement and real-time integration, these methods could form the backbone of smarter anti-phishing systems. Once full data access is granted, deeper models and better generalization can be achieved.

9. References

- S. Garera, N. Provos, M. Chew, and A. D. Rubin. "A framework for detection of phishing attacks." ACM workshop on Recurring Malicious Behavior, 2007.
- Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of deep bidirectional transformers for language understanding.