Detecting Phishing in Text Messages

Aprajita Kumari
Tanu Gupta
Department of Computer Science & Engineering, Schhol of Engineering & Technology, K.R. Mangalam University, Gurugram , Haryana , India

Abstract:
Nowadays, there are many data security issues. Hackers are now very skilled at using their knowledge to hack into someone else's system and grab information. Phishing is one such methodology used to acquire this information. Phishing is a cybercrime in which emails, telephone calls, text messages, personally identifiable information, banking details, credit card details, and passwords are targeted. Phishing is mainly a form of online identity theft. Social engineering is used by the phisher to steal the victim's personal data and account details. This research paper provides a fair overview of phishing attacks, the types of phishing attacks through which these attacks are performed, and detection and prevention strategies for phishing.

Introduction:
Phishing is the act of attempting to acquire information, such as usernames, passwords, and credit card details, by impersonating a trustworthy entity in electronic communication. Communications purporting to be from popular social websites, auction sites, online payment processors, or IT administrators are commonly used to lure the unsuspecting public. Phishing emails may contain links to websites infected with malware.

Phishing is an example of social engineering. It is primarily used in email hacking. In email phishing, the hacker sends a link via email to a user, such as pretending to be from a bank or other service, asking for personal information. The user clicks the link, fills in their details, and the hacker gains access to their personal information. This is how phishing is typically carried out.

Related Work:
SMS phishing, also known as smishing, is a deceptive practice that tricks individuals into revealing sensitive information through fraudulent SMS messages. Attackers use various techniques, such as impersonation, fake promotions, malicious links, and urgent requests, to manipulate victims into clicking phishing links or sharing confidential data. Smishing is a growing cybersecurity threat, targeting financial institutions, businesses, and individuals worldwide. Several studies have explored machine learning approaches for detecting phishing SMS. Researchers have employed classification models such as Support Vector Machine (SVM), Naïve Bayes (NB), Random Forest (RF), and XGBoost to distinguish phishing SMS from legitimate messages. Feature extraction techniques, such as TF-IDF, Word2Vec, and GloVe embeddings, have been widely used to enhance model performance.

Gupta et al. (2020) demonstrated that Random Forest achieved higher accuracy than SVM when using TF-IDF features. Similarly, Sharma et al. (2021) compared Logistic Regression and XGBoost, showing that GloVe-based features improved classification accuracy. However, short text length and the lack of contextual information in SMS remain major challenges in phishing detection.

Additionally, researchers have experimented with dimensionality reduction techniques, such as PCA, to optimize feature representation and improve classification efficiency. Despite these advancements, challenges such as evasive phishing techniques, multilingual SMS phishing, and adversarial attacks require further research.

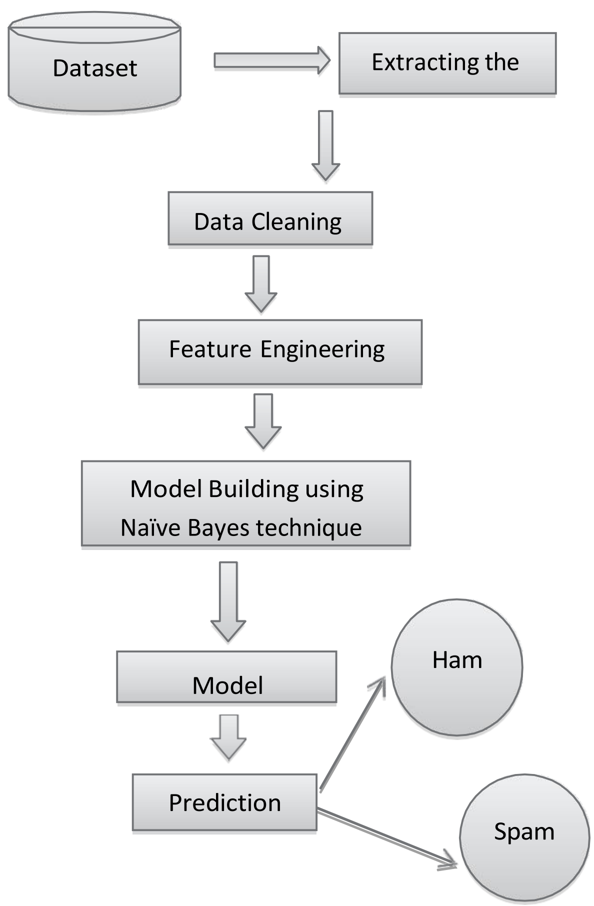
DOI: https://doi.org/10.5281/zenodo.15188546

ISSNNO-2584-2706

Our study builds upon existing work by evaluating SVM and XGBoost classifiers using GloVe and GloVe+PCA feature representations.

Proposed methodology:
In this research, we use nltk, numpy , pandas , scipy, gensim, scikit-learn,spacy that is a library in Python for machine learning model development . It has a toolset for data preparation, such as word tokenization, and word embedding. The word tokenization technique is used for taking text inputs into sequential data as index values of the words. The word
.

Extraction to improve SMS phishing detection.
Embedding technique is used to make more dimension of sequence into vector. After data preparation process, wetrain the model based on SVM, LOGISTIC REGRESSION, RANDOMFOREST,
XGBoost algorithms. Then, we evaluate the performance of the models and compare their performance with the model based of machine learning algorithms. The working flow of the framework



- ***Dataset Collection***: Gather a dataset containing SMS messages labeled as phishing (spam) and legitimate (ham).
- ***Extracting the Data:*** Load and preprocess the dataset to make it suitable for further processing.
- ***Data Cleaning:*** Convert text to lowercase. Remove special characters, numbers, and unnecessary symbols. Remove stop words and apply tokenization.

DOI: https://doi.org/10.5281/zenodo.15188546

- ***Feature Engineering:*** Convert text data into numerical format using feature extraction techniques. Use TF-IDF, GloVe embeddings, or PCA for dimensionality reduction.

- ***Model Building:*** Train machine learning models for classification. Use algorithms such as Support Vector Machine (SVM) and XG Boost.

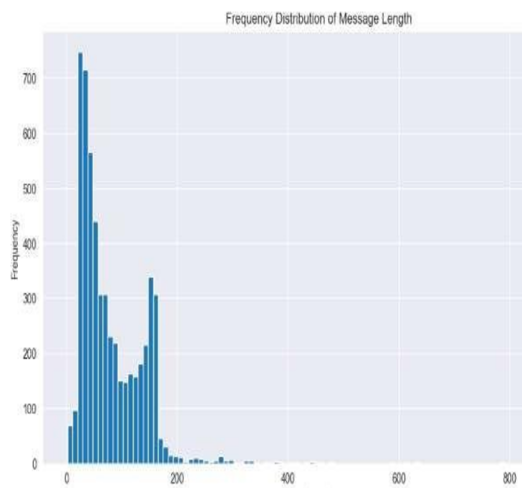
- ***Model Evaluation:*** Assess the model performance using evaluation metrics such as accuracy, precision, recall, and F1- score.

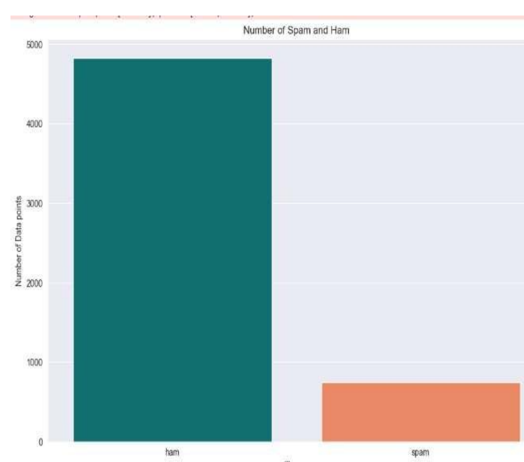
- ***Prediction:*** Deploy the trained model to classify incoming SMS messages as ham (legitimate) or spam (phishing).

Datasets

In this experiment, we use a SMS spam dataset proposed by mohitgupta-1O1/Kaggle-SMS-Spam-Collection-Dataset.

This dataset consists of approximately 5,574 records. It contains SMS text messaging conversationsinEnglishlanguage, which include text and number in different length of sentences. All records in this dataset already labeled. The spam messages are labelled as 1 (747 records) and the normal messages are labeled as 0 (4,825 records). The example of the dataset illustrated.





DOI: https://doi.org/10.5281/zenodo.15188546

Models and Algorithms Used:

1. Support Vector Machine (SVM)

SVM is a powerful supervised learning algorithm that works by finding the optimal hyperplane to separate different classes. It is effective for high-dimensional text data.It uses the kernel trick to transform non-linearly separable data in to a higher-dimensional space.

- **Advantage:** Works well with small to medium-sized datasets and handles text classification efficiently.
- **Limitation:** Computationally expensive for large datasets.

2. Naïve Bayes (NB):

Naïve Bayes is a probabilistic classifier based on Bayes' theorem with an assumption of independence between features. Commonly used for spam detection due to its simplicity and efficiency. Uses term frequency and conditional probabilities to classify SMS messages.

- **Advantage:** Fast and performs well even with small datasets.
- **Limitation:** Assumption of feature independence may not always hold in real-world text data.

3. Random Forest (RF):

Random Forest is an ensemble learning method that combines multiple decision trees to improve classification accuracy. Works by aggregating Predictions from multiple trees to reduce over fitting. Suitable for handling non-linear relationships in data.

- **Advantage:** Provides high accuracy and robustness to noisy data.
- **Limitation:** Can be computationally expensive and slow for very large datasets.

4. Logistic Regression (LR):

Logistic Regression is a linear model used for binary classification tasks. Computes the probability of an SMS being phishing or legitimate using the sigmoid function. Works well when features are linearly separable.

- **Advantage:** Simple, interpretable, and effective for text classification.
- **Limitation:** May not perform well on complex, non-linear relationships.

5. XG Boost (Extreme Gradient Boosting) XG Boost is a boosting algorithm that proves classification performance by training weak models iteratively. Uses gradient boosting to minimize errors and enhance model accuracy. Handles missing data and large-scale datasets efficiently.

- **Advantage:** Highly efficient, scalable, and outperforms traditional models in many text classification tasks.
- **Limitation:** Requires careful hyperparameter tuning to avoid over fitting.

DOI: https://doi.org/10.5281/zenodo.15188546

Experiment and Results

This section presents the findings of the proposed framework in this study. The experiments evaluate the performance of different machine learning models, including SVM, Naïve Bayes, Random Forest, Logistic Regression, and XGBoost. The models are analyzed and compared based on accuracy, precision, recall, F1-score, and AUC-ROC.

Experiment1:
Performance Comparison of Machine Learning Models using GloVe and GloVe + PCA

In this experiment, we compare the performance of different machine learning models, including SVM and XGBoost, using GloVe and GloVe + PCA embeddings for phishing SMS detection. The models are evaluated based on accuracy, precision, recall, F1-score, and AUC-ROC. The results indicate that XGBoost with GloVe + PCA achieves the highest accuracy, demonstrating its effectiveness in feature extraction and classification. The table below presents the detailed comparison of these models.

Model	Accuracy	Precision	Recall	F1-Score	AUC-ROC
SVM(Glove)	0.949776	0.861538	0.746667	0.800000	0.966370
SVM(Glove+ PCA)	0.937220	0.803030	0.706667	0.751773	0.965406
XGBoost (Glove)	0.964126	0.923077	0.800000	0.857143	0.980197
XGBoost (Glove+ PCA)	0.969507	0.946154	0.820000	0.878571	0.981440

Experiment2:
Impact of Feature Extraction on Model Performance

In this experiment, we analyze the impact of different feature extraction techniques on model performance. We compare the results of models using GloVe and GloVe + PCA to assess how dimensionality reduction affects classification. The results demonstrate that while GloVe provides Strong performance, incorporating PCA enhances generalization, particularly for SVM and XGBoost. The table below summarizes the performance variations.

These findings highlight the effectiveness of XGBoost in phishing detection and demonstrate that combining GloVe with PCA enhances model performance.

Model	Accuracy	Precision	Recall	F1-Score	AUC-ROC
SVM	0.949776	0.861538	0.746667	0.800000	0.966366
XGBoost	0.964126	0.923077	0.800000	0.857143	0.980197
SVM+ XGBoost	0.968610	0.945736	0.813333	0.874552	0.975320

DOI: https://doi.org/10.5281/zenodo.15188546

Conclusion:

"In this study, we explored machine learning approaches for phishing SMS detection. Our analysis demonstrated that SVM, Random Forest, XGBoost, Naïve Byes achieved the highest accuracy using GloVe-based feature representation. The results indicate that word embeddings combined with dimensionality reduction techniques can improve classification performance. However, the study was limited to English-language SMS and a relatively small dataset. In the future, we aim to extend this research to multilingual datasets, deep learning-based approaches, and real-time phishing detection systemsto enhance security against evolving cyber threats."

Acknowledgement:

"I would like to express my sincere gratitude to my supervisor Tanu Gupta, for their invaluable guidance, constructive feedback, and continuous support throughout this research. I am also grateful to K.R. Mangalam University for providing the necessary resources and academic environment to conduct this study. Additionally, I acknowledge the contributions of researchers in the field whose work has inspired and informed my research."

References:

[1] K. Haynes, H. Shirazi, and I. Ray (2021). Lightweight URL-based phishing detection using natural language processing transformers for mobile devices. *Procedia Computer Science, 191*, 127–134. https://doi.org/10.1016/j.procs.2021.07.040

[2] A. Abbasi, D. Dobolyi, A. Vance, and F. M. Zahedi (2021). The Phishing Funnel Model: A Design Artifact to Predict User Susceptibility to Phishing Websites. *Information Systems Research, 32(2),* 410–436. https://doi.org/10.1287/isre.2020.0973

[3] A. Aljofey, Q. Jiang, Q. Qu, M. Huang, and J. P. Niyigena (2020). An Effective Phishing Detection Model Based on Character Level Convolutional Neural Network from URL. *Electronics (Basel), 9(9),* 1514. https://doi.org/10.3390/electronics9091514

[4] C. Jones (2022, January 18). 50 phishing stats you should know in 2022. *50 Phishing Stats You Should Know in 2022*. Retrieved January 27, 2022, from https://expertinsights.com/insights/50-phishing-stats-you-should-know/

[5] A. Hannousse and S. Yahiouche (2021). Web page phishing detection. *Mendeley Data, V3*. doi: 10.17632/c2gw7fy2j4.3

[6] G. K. Soon, C. O. Kim, N. M. Rusli, T. S. Fun, R. Alfred, and T. T. Guan (2020). Comparison of simple feedforward neural network, recurrent neural network, and ensemble neural networks in phishing detection. *Journal of Physics: Conference Series, 1502(1),* 12033. https://doi.org/10.1088/1742-6596/1502/1/012033

[7] Anti-Phishing Working Group (APWG) (2014). Phishing Activity Trends Report, 3rd Quarter 2021. [Online] Retrieved Feb 9, 2022, from https://docs.apwg.org/reports/apwg_trends_report_q3_2021.pdf

[8] M. Kearns, and A. Roth (2022, March 9). Ethical Algorithm Design Should Guide Technology Regulation. *Brookings*. Retrieved June 9, 2022, from https://www.brookings.edu/research/ethical-algorithm-design-should-guide-technology-regulation/#footnote-3