Machine Learning & NLP: Modern Approaches to Phishing Detection

Course Name : Professional Foundation
Course Code : ECE 591

Instructors: Dr. Imen Bourguiba, David Woodward

Group 5
Chita (Yangxuedan) Xiang – V01047490
Eve Lu – V01073022
Mamta Rawat – V01066262
Prabhnoor Kaur – V01068454
Sarah Firouzabadi – V01043024

Acknowledgement

We acknowledge and respect the Lək̓ʷəŋən (Songhees and Xʷsepsəm/Esquimalt)Peoples on whose territory the university stands, and the Lək̓ʷəŋən and W̱SÁNEĆPeoples whose historical relationships with the land continue to this day.

What’s in Store!

- Acknowledgement
- Introduction
- Literature Review
- Email Dataset Analysis
- Phishing Email Word Cloud
- References

Introduction

Phishing is one of the most dangerous and persistent threats, with attackers constantly evolving techniques to evade traditional defenses.

Many successful attacks exploit human psychological weaknesses, such as:

- ➢ Reliance on memory
- ➢ Distraction by attention-grabbing elements
- ➢ Trust in familiar-looking content

There's a growing need to use advanced technologies like:

- ➢ Machine Learning (ML)
- ➢ Deep Learning (DL)
- ➢ Natural Language Processing (NLP)

Literature Review > Phishing Detection Techniques

Technique	Description	Limitations
Blacklist-Based Detection	Compares sender/domain to known phishing lists.	Cannot detect new or unknown phishing attacks (zero-day threats).
Heuristic-Based Detection	Analyzes structure and keywords using pre-set rules	High false positive rate; requires constant updates to remain effective.
Behavioural-Based Analysis	Profiles normal sender and recipient behaviors and identifies deviations from established email interaction patterns.	Continuous monitoring can lead to privacy issues and increased computational overhead.
Machine Learning Models	Utilizes supervised classification, NLP, and deep learning to analyze email patterns. Differentiates phishing emails from legitimate ones by learning and recognizing distinct features.	The system's effectiveness depends on diverse and high-quality training data. Consequently, new or emerging phishing tactics might not be fully captured.

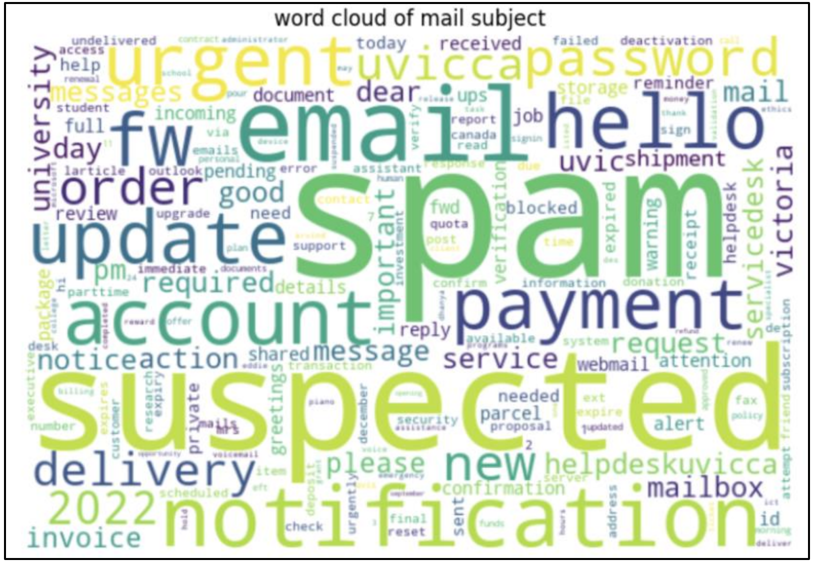
Literature Review > Deep Learning Approaches

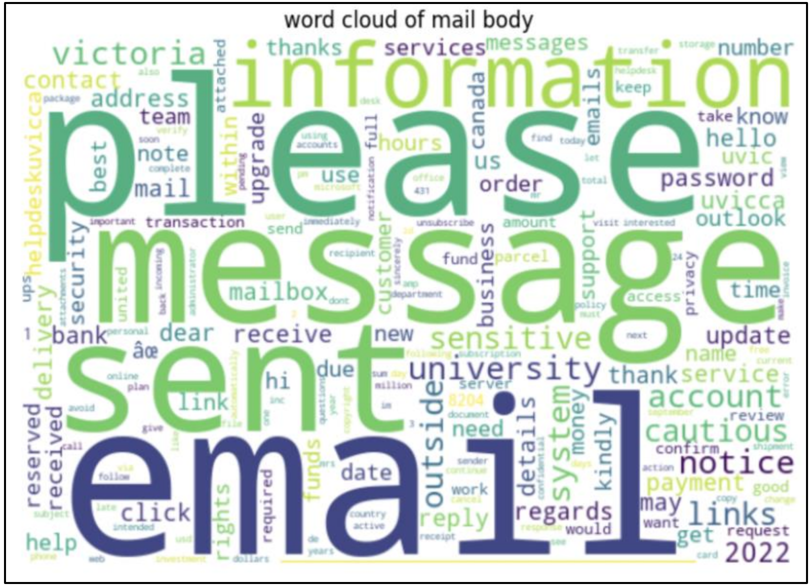
Technique	Description	Limitations
Improved RCNN Model With Multilevel Vectors and Attention Mechanism	Used a deep learning model named THEMIS to model the email header and the email body at both the character level and the word level.	Improving phishing email detection when only the email body is available, without access to the email header for THEMSIS model.
LSTM, Bi-LSTM, GRU, Bi-GRU	Compared CNN models augmented with recurrent layers best-performing CNN + Bi-GR	Require large datasets for training; performance might decline with noisy or adversarial data.
Transformer Models for Phishing Email Detection using BERT	Compared Transformer models like BERT, distilBERT, XLNet, ALBERT, and roBERTa, with traditional models. Model is a pretrained model that utilizes bidirectional transformer logic, allowing it to analyze provided text data in both directions.	There is a lack of sentiment analysis and transformer technology to capture both social engineering techniques and emotional intent in emails.

Literature Review > Natural Language Processing Techniques

Technique	Description	Limitations
TF-IDF, Word Embeddings (Salloum et al., 2022)	Provides foundational NLP features by quantifying word relevance and capturing semantics; widely used with ML models like SVM, Naive Bayes, RF.	Doesn't capture context; weak with short or multilingual emails .
TF-IDF + CountVectorizer (PhishGuard, Fahim et al., 2024)	Compared both vectorizers for phishing classification using Random Forest and Naive Bayes; TF-IDF showed higher accuracy (98.91%).	Both ignore word order and semantics; CountVectorizer less expressive than TF-IDF .
Composite NLP + Ensemble (DARTH) (Mittal et al., 2022)	Analyzes multiple email components (sender, subject, body) via individual NLP models, then combines them using an ensemble for near-perfect phishing classification.	High complexity; requires large labeled datasets and more resources .
Word2Vec & BERT Embeddings (SMU BERT paper, 2022)	Embeds words with context (BERT) or similarity (Word2Vec); used for deep phishing detection with high precision and F1 score (98.66%).	BERT is resource-intensive; Word2Vec lacks context awareness beyond local window

Phishing Email Word Cloud





Word Cloud of mail subject and mail body

Email Dataset Analysis

Total Rows: 2,576 emails

Unique Subjects: 1,940
Unique Bodies: 2,284

Most Common Subject: Empty subject > Appear 109 times
Most Common Body Content:

> Notice: *This message was sent from outside the University of Victoria email system. Please be cautious with links and sensitive information.*
>
> Maintenance required update your mailbox https://log.microsoftonline.com/adfs/ls/?login_hint?<https://suaccounservi.brizy.site/>
>
> Admin Team.

Email Dataset Analysis – Cont.

- Average word count: **~90 words**
- Maximum Words: **2,651 words**
- Typical range (IQR): **36 to 109 words**
- **109 emails** are missing subject and Only **5 emails** are missing body text
- **225** exact duplicate emails were found
- **1,406** emails contains URLs (~54.6%)

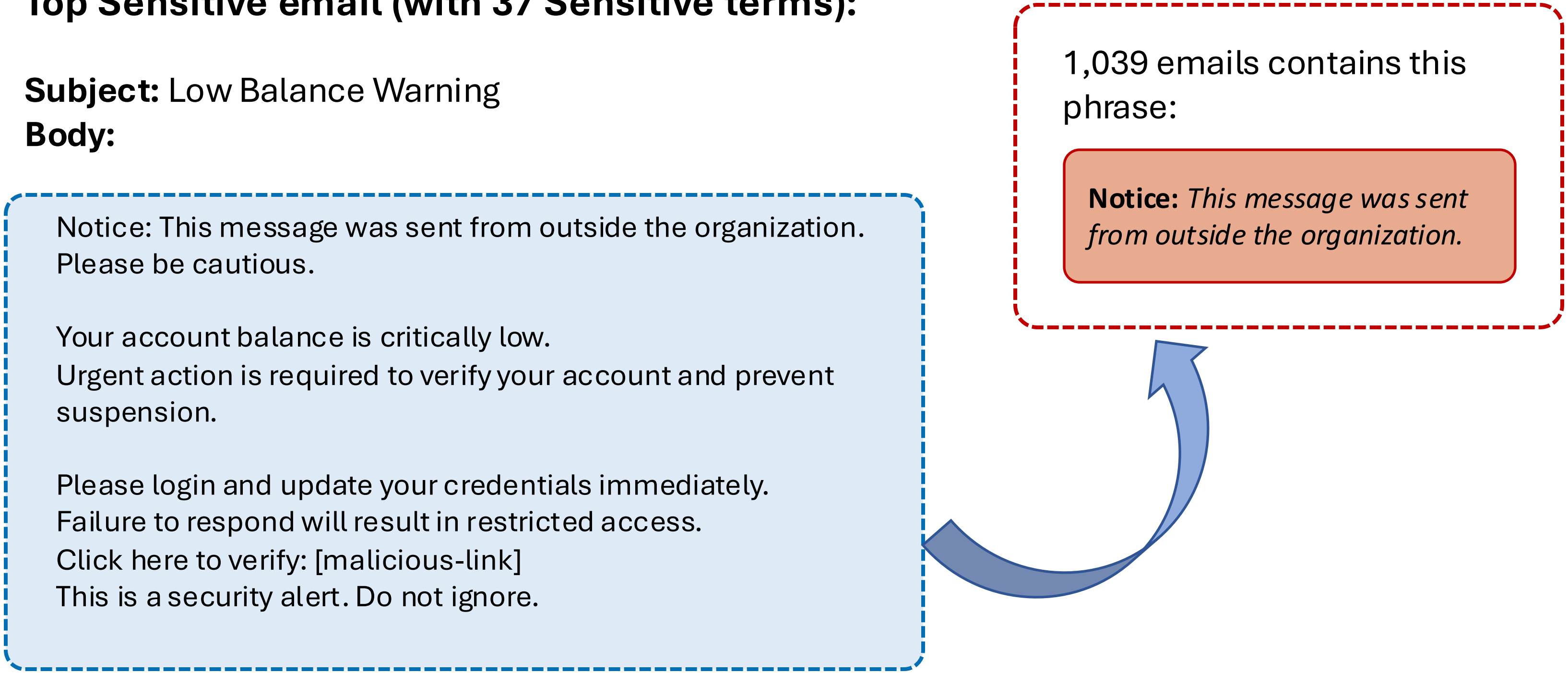
Top 10 Most Common Subjects

Subject Line	Frequency
Fw:	65
Re: IT Servicedesk	39
University of Victoria	26
Hello	24
University of Victoria Webmail	20
Good day	20
UPDATE	20
SPAM Suspected	16
Fwd:	12
Greetings	7

Email Dataset Analysis – Cont.

Top Sensitive email (with 37 Sensitive terms):

Subject: Low Balance Warning
Body:



Notice: This message was sent from outside the organization.
Please be cautious.

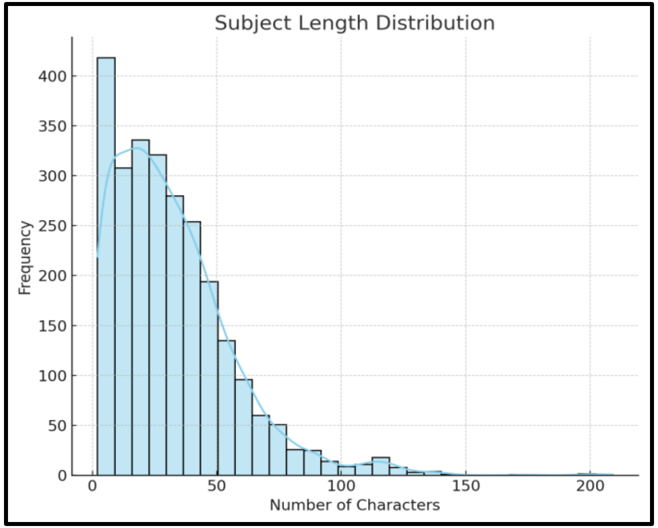
Your account balance is critically low.
Urgent action is required to verify your account and prevent suspension.

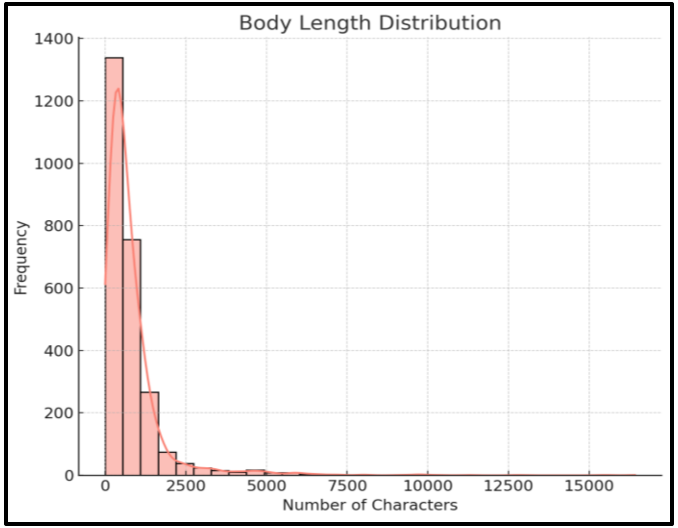
Please login and update your credentials immediately.
Failure to respond will result in restricted access.
Click here to verify: [malicious-link]
This is a security alert. Do not ignore.

1,039 emails contains this phrase:

Notice: *This message was sent from outside the organization.*

Email Dataset Analysis – Cont.

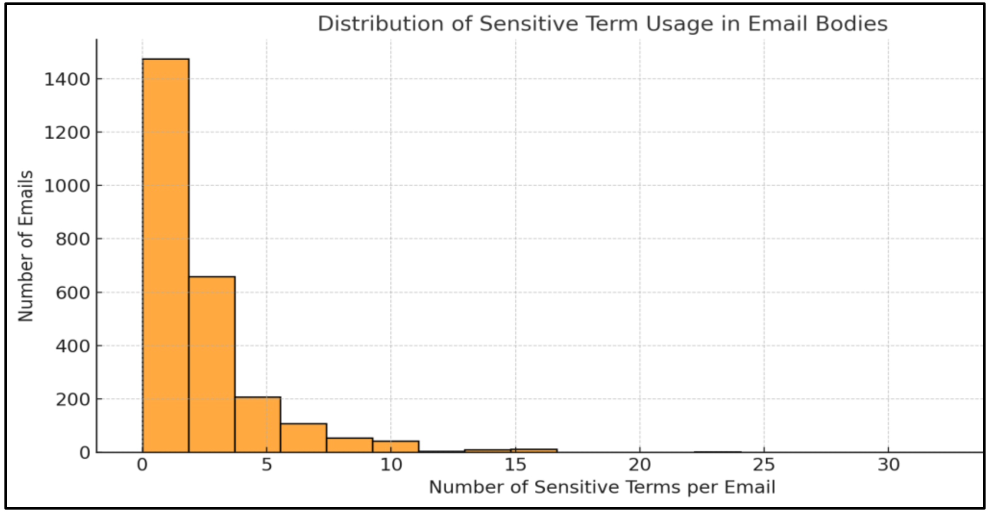




- Shows a strong right-skew, with most subjects clustered below **50 characters**.
- Indicates that short and concise subject lines dominate the dataset, though outliers up to **~200 characters** appear.
- Suggests subjects are typically brief, possibly reflecting common email norms or phishing strategies to grab attention quickly.

- Shows a strong right-skew, with most emails under **2,500 characters.**
- Some outliers reach up to **15,000 characters**, signaling that certain emails can be extremely lengthy.
- Highlights significant variability in message content size, requiring flexible handling for text analysis or filtering approaches.

Email Dataset Analysis – Cont.



The chart displays how many "sensitive terms" (like "password," "account," or "bank") appear in each email body.

- ➢ The distribution is heavily **skewed toward zero**, meaning most emails contain few or no sensitive terms. However, a noticeable minority includes multiple such terms.
- ➢ These higher counts often point to potentially malicious or phishing-related content. Tracking the frequency of sensitive terms can help **in prioritizing suspicious emails for deeper inspection**.

References

[1] M. A. Uddin and I. H. Sarker, “An Explainable Transformer-based Model for Phishing Email Detection: A Large Language Model Approach,” *Preprint*, arXiv:2402.13871, Feb. 2024. [Online]. Available: https://arxiv.org/abs/2402.13871

[2] P. H. Kyaw, J. Gutierrez, and A. Ghobakhlou, “A Systematic Review of Deep Learning Techniques for Phishing Email Detection,” *Electronics*, vol. 13, no. 19, p. 3823, Sep. 2024. [Online]. Available: https://doi.org/10.3390/electronics13193823

[3] S. Salloum, T. Gaber, S. Vadera, and K. Shaalan, “A Systematic Literature Review on Phishing Email Detection Using Natural Language Processing Techniques,” *IEEE Access*, vol. 10, pp. 65703–65723, 2022. [Online]. Available: https://doi.org/10.1109/ACCESS.2022.3183083

[4] M. Khonji, Y. Iraqi, and A. Jones, “Phishing Detection: A Literature Survey,” *IEEE Communications Surveys & Tutorials*, vol. 15, no. 4, pp. 2091–2121, 2013. [Online]. Available: https://doi.org/10.1109/SURV.2013.032213.00009

[5] K. Singh, P. Aggarwal, P. Rajivan, and C. Gonzalez, “What Makes Phishing Emails Hard for Humans to Detect?” in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 67, no. 1, pp. 1097–1101, 2023. [Online]. Available: https://doi.org/10.1177/1071181320641097

[6] A. Mittal, T. Chowdhury, R. Sivaraman, D. W. Engels, and H. Kommanapalli, “Phishing Detection Using Natural Language Processing and Machine Learning,” *SMU Data Science Review*, vol. 6, no. 2, Article 14, 2022. [Online]. Available: https://scholar.smu.edu/datasciencereview/vol6/iss2/14

[7] R. A. Fahim, M. A. Al Hafiz, and S. Azam, “PhishGuard: Leveraging NLP and Machine Learning for Email Phishing Detection,” *Electronics*, vol. 13, no. 19, p. 3823, 2024. [Online]. Available: https://doi.org/10.3390/electronics13193823

[8] M. AlJamal and H. Al-Mubaid, “Harnessing ML and NLP for Enhanced Cybersecurity: A Comprehensive Approach for Phishing Email Detection,” *IEEE Access*, vol. 12, pp. 18657–18668, 2024. [Online]. Available: https://doi.org/10.1109/ACCESS.2024.3358374

[9] Y. Fang, C. Zhang, C. Huang, L. Liu and Y. Yang, "Phishing Email Detection Using Improved RCNN Model With Multilevel Vectors and Attention Mechanism," in IEEE Access, vol. 7, pp. 56329-56340, 2019, doi: 10.1109/ACCESS.2019.2913705.

[10] A. K. Jain and B. B. Gupta, ‘‘*Phishing detection: Analysis of visual similarity based approaches*,’’ Secure Communication Network, vol. 2017, pp. 1–20, Jan. 2017.

[11] K. L. Chiew, K. S. C. Yong, and C. L. Tan, ‘‘*A survey of phishing attacks: Their types, vectors and technical approaches,*’’ Exp. Syst. Appl., vol. 106, pp. 1–20, Sep. 2018.

[12] [9] N. Altwaijry, I. Al-Turaiki, R. Alotaibi, and F. Alakeel, “Advancing Phishing Email Detection: A Comparative Study of Deep Learning Models,” Sensors, vol. 24, no. 7, p. 2077, Mar. 2024. [Online]. Available: https://doi.org/10.3390/s24072077

[13] K. Thakur, M. L. Ali, M. A. Obaidat, and A. Kamruzzaman, “A Systematic Review on Deep-Learning-Based Phishing Email Detection,” Electronics, vol. 12, no. 21, p. 4545, Oct. 2023. [Online]. Available: https://doi.org/10.3390/electronics12214545

[14] M. Alanezi, “Phishing Detection Methods: A Review,” Technium Science, vol. 3, no. 4, pp. 57–69, 2021. [Online]. Available: https://techniumscience.com/index.php/technium/article/view/4973/1723

Thanks!