Leveraging Transfer Learning and Few-Shot Learning for Zero-Day Phishing Attack Detection

Sai Deepika Tamminidi
School of Engineering and Computer Science
University of the Pacific
Stockton, USA
s_tamminidi@u.pacific.edu

Om Bagwe
School of Engineering and Computer Science
University of the Pacific
Stockton, USA
o_bagwe@u.pacific.edu

Simran Shah
School of Engineering and Computer Science
University of the Pacific
Stockton, USA
s_shah22@u.pacific.edu

***Abstract*—This paper uses few-shot and transfer learning, two cutting-edge machine learning approaches, to detect zero-day phishing attacks. Because traditional cybersecurity solutions rely on known signatures and past data, they frequently fall short of such attacks. To close this gap, we propose a technique that successfully detects novel phishing strategies by combining few-shot learning with transfer learning. We used a dataset containing 10,500 online entities that had been improved with 31 unique characteristics essential for phishing identification. The experimental findings show our models' success, especially when combined with sophisticated machine-learning techniques and categorical variables. The results show notable advancements in identifying complex phishing attempts, providing a fresh viewpoint on using cutting-edge machine learning techniques in cybersecurity.**

Keywords—Zero-Day Phishing Attacks, Transfer Learning, Few-Shot Learning, Cybersecurity, Machine Learning, Phishing Detection, Advanced Detection Techniques, Novel Phishing Strategies, Dataset, Web Entities

I. Introduction

Because zero-day phishing attempts are random and lack observable patterns, they present a significant challenge in the rapidly evolving field of cybersecurity. These sophisticated attacks typically bypass conventional cybersecurity defenses, which rely heavily on signature-based and heuristic detection techniques [1], [2]. These traditional methods are ineffective against novel and unidentified phishing strategies as they require prior knowledge of attack signatures [3]. Our research addresses this critical gap by employing advanced machine learning techniques, particularly few-shot and transfer learning, adept at adjusting to new threats. Such innovative methodologies are particularly beneficial for enhancing phishing detection capabilities, as demonstrated by their effectiveness in identifying sophisticated phishing attempts through patterns and anomalies in data [1], [6]. Few-shot learning is designed to deliver robust performance from limited examples, making it suitable for phishing threats' dynamic and evolving nature. This approach is aligned with the need for models that can quickly adapt to new types of phishing attacks without extensive retraining [4], [5].

Meanwhile, transfer learning enables our models to utilize learned features from related tasks, enhancing accuracy and adaptability. This is crucial for developing systems leveraging existing knowledge to counteract new threats effectively [6] [7]. By integrating these methodologies with a comprehensive dataset enriched with various web entity features, our study aims to develop resilient models capable of anticipating and mitigating zero-day phishing attempts. This integration of advanced machine learning strategies into cybersecurity defense mechanisms is underscored by a systematic review of the Literature and case studies highlight the successful application of these techniques across different contexts, including IoT security and NLP-based phishing detection [3], [5]. This paper presents empirical results demonstrating the effectiveness of our approach, discusses the methodology that includes data collection and model development, and explores the theoretical underpinnings of our applied machine learning strategies, ultimately advancing the capabilities of cybersecurity defenses [2], [4].

II. Motivation

A. Inadequacy of Traditional Methods:

Conventional phishing detection systems rely heavily on databases of known phishing signatures and heuristic-based methods. However, as cyber attackers continually develop new strategies, these traditional defenses become less effective. The static nature of signature databases means they cannot detect phishing attempts that do not match previously identified patterns.

B. Emergence of Advanced Machine Learning Techniques:

Recent advancements in machine learning have introduced promising capabilities that could revolutionize phishing detection. In particular, transfer and few-shot learning can leverage prior knowledge and limited data to rapidly adapt to new and emerging threats. These techniques provide the foundation for models that can learn from a few examples and generalize well to new, unseen scenarios[6].

C. Need for Proactive Security Measures:

With the increasing reliance on digital platforms, the impact of phishing attacks has expanded, potentially leading to substantial financial losses and damage to organizational reputation. This underscores the necessity for developing proactive detection systems that rely not solely on known threat signatures but can anticipate and neutralize attacks before they cause harm.

D. Research Gap:

While there is extensive research on using traditional and some machine learning techniques for cybersecurity, there need to be more studies that combine transfer learning with few-shot learning specifically targeted at zero-day phishing attack detection. Addressing this gap can significantly contribute to the cybersecurity field by enhancing the capability of detection systems to adapt quickly to novel threats.

The need for a more potent defense against cyber attackers' ever-evolving strategies motivates this study. This work seeks to contribute to cybersecurity by offering an innovative, reliable, and flexible solution to early phishing

Detection by concentrating on novel machine-learning strategies.

III. Literature review

In the rapidly evolving landscape of cybersecurity, the detection of zero-day phishing attacks poses unique Challenges due to their unpredictable and novel nature. Traditional detection systems, reliant on historical data and known signatures, often must be more effective against such threats. Integrating advanced machine learning techniques, including Transfer Learning, Few-Shot Learning, BERT, XGBoost, and Model-Agnostic Meta-Learning (MAML), represents a transformative approach to enhance the adaptability and effectiveness of phishing detection systems. These methodologies leverage previous learnings and minimal data to quickly adapt to new patterns, offering a significant advantage in identifying and mitigating zero-day attacks. This paper explores seminal works in these areas to establish a comprehensive understanding of how these techniques can be synergistically applied to bolster cybersecurity defenses against the most elusive threats.

In machine learning, few-shot and transfer learning have become pivotal for enhancing model performance with limited data availability. This literature review examines ten significant papers that have contributed to the fields of few-shot learning, transfer learning and their applications in various domains, including cybersecurity and sentiment analysis.

Meta-Transfer Learning for Few-Shot Learning [8] explores the integration of transfer learning with meta-learning techniques to boost adaptability and efficiency under few-shot scenarios, demonstrating how transfer learning can improve the generalization capabilities of models across diverse tasks with minimal data input, significantly enhancing the practical applicability of few-shot learning models where data sparsity is common. A Comprehensive Survey of Few-shot Learning: Evolution, Applications, Challenges, and Opportunities [9] provides an exhaustive overview of the few-shot learning landscape, outlining its progression, current applications, and significant challenges such as feature reuse sensitivity and inaccurate data distribution assessments, underscoring the necessity for ongoing research to mitigate these issues and highlighting the evolving nature of few-shot learning techniques and their expanding application spectrum.

Meta-learning Approaches for Few-Shot Learning: A Survey of Recent Advances [10] delves into contemporary methodologies that refine the efficiency and adaptability of few-shot learning, discussing various advanced meta-learning strategies developed to enhance the rapid learning capabilities of models with exceedingly limited data, crucial for developing robust algorithms that can quickly adapt to new and evolving tasks, particularly in fields experiencing rapid shifts in data characteristics and task requirements. MetaModulation: Learning Variational Feature Hierarchies for Few-Shot Learning with Fewer Tasks [11] introduces an innovative approach using variational task modulation to enhance few-shot model adaptability and efficiency, allowing models to achieve quick and efficient adaptation with fewer learning tasks, showcasing a significant leap in the operational flexibility and efficiency of few-shot learning systems.

Global Convergence of MAML and Theory-Inspired Neural Architecture Search for Few-Shot Learning [12] examines the theoretical underpinnings of Model-Agnostic Meta-Learning (MAML) and its integration with neural architecture search methods, providing a deep dive into the convergence properties of MAML and illustrating how theoretical insights into neural architecture can enhance the rapid adaptability of few-shot learning models, pivotal for advancing the understanding and capabilities of meta-learning algorithms. EfficientNet-XGBoost: An Implementation for Facial Emotion Recognition Using Transfer Learning [13] showcases the application of XGBoost integrated with EfficientNet through transfer learning techniques for facial emotion recognition, highlighting how the synergy between convolutional neural networks and gradient boosting methods can lead to high-performance classification systems, applicable in areas beyond facial recognition, such as phishing detection.

Transfer Learning for Sentiment Analysis Using BERT-Based Supervised Fine-Tuning [14] explores the use of BERT for sentiment analysis, leveraging its transfer learning capabilities to achieve state-of-the-art performance, emphasizing the effectiveness of supervised fine-tuning of BERT in enhancing sentiment analysis accuracy across specific languages or domains, showcasing the versatility of transfer learning in text analysis applications. FiT: Parameter Efficient Few-shot Transfer Learning for Personalized and Federated Image Classification [15] explores a novel transfer learning approach that leverages few-shot techniques to enhance performance in image classification, adapting well to scenarios where data privacy concerns or resource constraints limit the availability of extensive datasets, demonstrating the practical applicability of few-shot learning in sensitive environments.

Combining Model-Agnostic Meta-Learning and Transfer Learning for Regression [16] investigates the integration of MAML with transfer learning to address regression problems, particularly pertinent for tasks such as phishing detection, where regression on indicative features can significantly enhance detection capabilities. A Fine-Tuned BERT-Based Transfer Learning Approach for Text Classification [17] evaluates the effectiveness of BERT in text classification, particularly in distinguishing between legitimate and phishing communications, demonstrating how fine-tuning BERT can lead to significant improvements in text classification tasks, underscoring the utility of transfer learning in enhancing model performance across different text-based applications.

IV. Background information

In the rapidly evolving landscape of cybersecurity threats, zero-day phishing attacks represent a significant challenge due to their unpredictable nature and absence of prior detectable patterns. Traditional cybersecurity measures often prove inadequate to mitigate these threats, as they primarily depend on known signatures and historical data. Addressing this critical gap, this project introduces a novel approach by integrating advanced machine learning techniques—specifically, transfer learning and few-shot learning—into detecting zero-day phishing attacks.

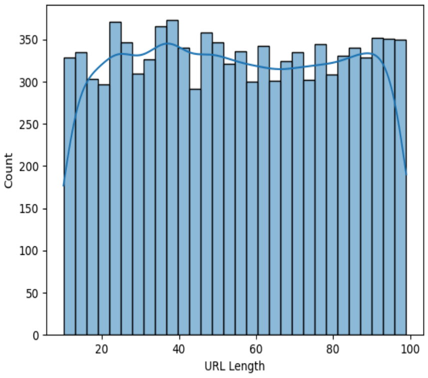
A. *Datasets*

The dataset utilized in this project comprises 10,500 entries, each representing a web entity potentially indicative Of phishing activities. This dataset was explicitly curated to facilitate the training and evaluation of machine learning models designed to detect phishing attacks, including Zero-day threats that were not previously known or cataloged.

The dataset includes 31 features that capture various aspects of web entities, which are crucial for detecting phishing. These features encompass the URL itself, the domain's age and registration length, the presence of an IP address in the URL, the use of HTTPS tokens, and the URL length. It further examines URL characteristics such as shortening services, the presence of special symbols, redirection patterns, and domain suffix or prefix anomalies. Additional security-related attributes include the SSL certificate status, domain registration country, use of secure ports, and HTTP versus HTTPS protocols. The dataset also offers content-based metrics such as the number of subdomains, links in meta/scripts/links, server form handler configurations, and iframe redirection behaviors. Metrics such as website traffic, PageRank, content length, presence of special characters in the domain, email addresses within URLs, sensitive words, and response time are analyzed to discern patterns that distinguish phishing sites from legitimate ones. This comprehensive range of features aids in creating robust models capable of detecting even the most subtle phishing attempts.

B. *Environment setup*

The computational experiments for this project were carried out using PyCharm Professional 2022.1, a robust Integrated Development Environment (IDE) designed for professional developers. PyCharm was chosen for its comprehensive support for Python programming and powerful debugging, testing, and version control tools, essential for developing complex machine learning applications.



a) Virtual Environment Configuration: A dedicated virtual environment was created within PyCharm to manage dependencies specific to this project. This isolation ensures that the project's dependencies do not conflict with the global Python environment, which enhances reproducibility and simplifies dependency management. The virtual Environment was built using Python 3.8, ensuring compatibility with all the libraries and frameworks.

b) Libraries and Frameworks: The project extensively utilizes several Python libraries, each chosen for their specific features enabling efficient data processing, model training, and evaluation:

- Pandas (version 1.3.3) and NumPy (version 1.21.2) were used for data manipulation and numerical operations. These libraries provide the backbone for data preprocessing, which includes handling missing data, normalizing datasets, and transforming features.
- Scikit-learn (version 0.24.2) provided various tools for data preprocessing, model selection, and evaluation. It was mainly used to implement traditional machine learning models such as Logistic Regression and Random Forest classifiers.
- TensorFlow (version 2.6.0) and PyTorch (version 1.9.0) were employed to develop and train deep learning models. These frameworks support Various neural network architectures, including LSTM networks, and analyze sequential data inherent in URLs.
- Matplotlib (version 3.4.3) and Seaborn (version 0.11.2) were utilized to create visualizations to analyze data distributions and interpret model performance metrics.
- The transformers library (version 4.9.2) from Hugging Face was instrumental in leveraging pre-trained BERT models, which were fine-tuned for phishing detection. This library simplifies the implementation of state-of-the-art transformer models and is compatible with TensorFlow and PyTorch.
- MAML implementation was facilitated through custom adaptations in PyTorch, demonstrating the flexibility of PyTorch in implementing complex model-agnostic meta-learning algorithms.

c) Implementation: Models were developed and tested on a system equipped with an NVIDIA GPU, which significantly accelerated the training and evaluation processes, particularly for compute-intensive models like BERT and LSTM. The use of GPU resources was managed through CUDA 11.2, optimizing computational efficiency.

C. *Data cleaning and preprocessing*

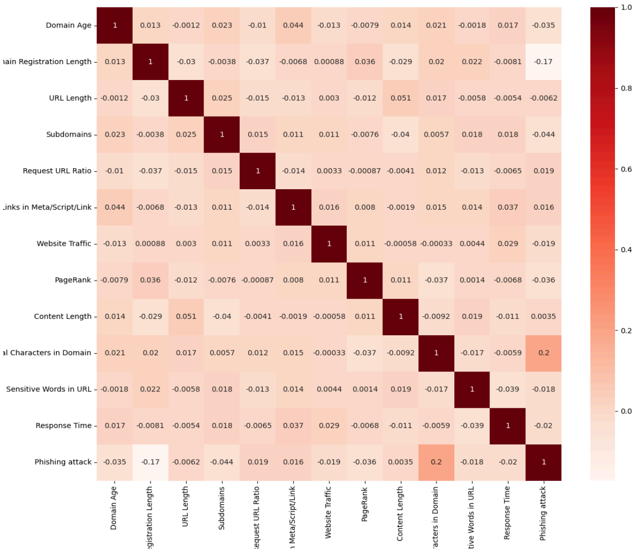
In the data preprocessing phase, the dataset underwent rigorous cleaning where missing values for numeric features were imputed with the mean and those for categorical features with the mode, thereby retaining the original distribution and categorical balance. Whitespace was meticulously trimmed from all variables except 'URL' to eliminate inconsistencies and preserve the integrity of the data for machine learning applications. 'URL' contains outliers, and we have kept it for the few-shot learning algorithm. This step ensured a standardized dataset, which is crucial for the robust performance of the predictive models.

D. *Exploratory data analysis*

We systematically examined the dataset's numeric, categorical, and text-based features as part of our comprehensive exploratory data analysis. This process entailed meticulously scrutinizing data distributions, identifying outliers, assessing feature correlations, and

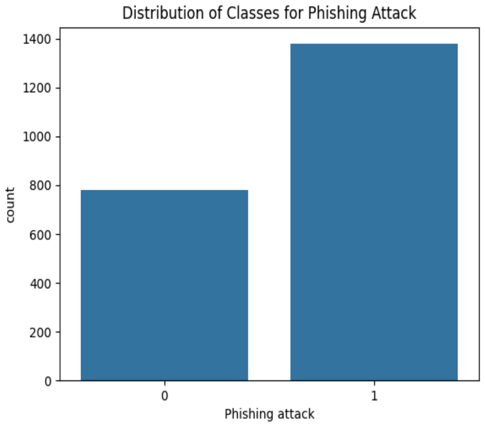
Detecting underlying patterns within the dataset. Such an investigation Is pivotal to gaining profound insights into the data's characteristics, informing subsequent data preprocessing, feature engineering, and selecting Appropriate machine learning models for detecting phishing attacks.

The correlation matrix revealed significant inter-feature relationships, with several pairs exhibiting correlations above the threshold of 0.05, indicative of a potential predictive relationship. Notably, 'Domain Age' and 'Links in Meta/Script/Link' demonstrated a moderate correlation, implying a relationship between the domain's longevity and the complexity of its HTML structure. Similarly, 'Domain Registration Length' and 'PageRank' showed a noteworthy correlation, suggesting that domains with more extended registration periods may be rewarded with higher PageRank scores, reflecting their established trustworthiness. The 'URL Length' had a proportional correlation with 'Content-Length,' hinting that Longer URLs might be associated with more comprehensive Content. A subtle correlation between 'Subdomains' and 'URL Length' suggests that more subdomains often result in longer URLs. 'Response Time' was positively correlated with both 'Links in Meta/Script/Link' and 'Website Traffic,' indicating that higher traffic and more embedded resources might impact server responsiveness. Lastly, the correlation between 'Subdomains' and 'Domain Age' suggests older domains tend to have more subdomains, potentially Reflecting their growth and evolution over time. These Correlations are instrumental in feature engineering and model optimization, particularly in enhancing the detection capabilities for zero-day phishing attacks.



During the exploratory data analysis, the class distribution for the 'Phishing attack' label was visualized, revealing a skew towards phishing instances ('1') over legitimate ones ('0'). This class imbalance, depicted in the bar chart, is crucial as it may bias the predictive models toward the majority class, necessitating strategies like resampling or adjusted evaluation metrics to ensure balanced classification performance.

The categorical features identified include 'Having IP Address,' 'HTTPS Token,' 'URL Shortening,' 'SSL Final State,' 'Domain Registration Country,' 'Email in URL,' 'Age of Domain,' 'Submitting to Email,' 'Server Form Handler (SFH),' 'HTTPS in URL,' 'Favicon Hosting,' 'Redirecting "//,"' and '@ Symbol.' Among these, 'HTTPS in URL,' 'Favicon Hosting,' 'Redirecting "//,"' '@ Symbol,' and 'Domain Registration Country' were deemed significant and were subsequently chosen for feature engineering.



After conversion to a suitable numerical format, these categorical features are expected to enhance the model’s predictive capability by providing additional discriminative information for classifying phishing attacks. The analysis and selection of these categorical variables conclude the EDA phase, setting the stage for the Subsequent stages of feature engineering and model Development.

E. Feature Engineering

In the feature engineering phase, numerical data was enriched by creating interaction features to reflect complex relationships, such as 'Age_Link_Meta' and 'RegLen_PageRank,' which combine domain-related metrics with link and ranking information. Respectively. This multiplication of related features aimed to introduce non-linear characteristics into the model, thereby augmenting the dataset with a more nuanced representation of the underlying patterns.

Categorical variables transformed one-hot encoding into a binary matrix essential for algorithmic processing. Text data was quantified by extracting the URL features, including length and particular character count, to capture their syntactic properties. These engineered features from numerical, categorical, and text data were meticulously crafted to bolster the dataset's predictive power for the ensuing machine-learning tasks.

V. System model

The system model devised for phishing attack detection employs a sequential process beginning with data preprocessing, followed by feature engineering, and culminating in the application of machine learning algorithms. The model utilizes diverse datasets, categorized by feature type, to train various classifiers and evaluate their performance in identifying encoding attempts.

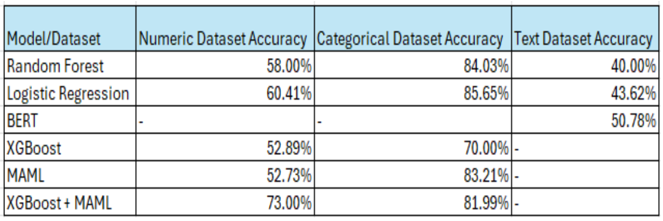
For the numerical dataset, feature interactions were created to capture non-linear patterns. In contrast, one-hot encoding was applied to the categorical dataset to facilitate model interpretation of non-numeric data. Text features Were derived from URL patterns to inform the syntactic structure indicative of phishing.

Upon applying a Random Forest classifier to the datasets, the categorical dataset emerged with superior performance, achieving an 84.03% accuracy. Logistic regression analysis on the same dataset demonstrated an incremental improvement in accuracy to 85.65%, with an ROC curve AUC of 0.87, indicating excellent model discrimination capability.

Further explorations with BERT and XGBoost on individual datasets yielded 63% and 70% accuracy, respectively. However, with the introduction of the MAML model, the accuracy for the categorical dataset notably increased to 83.21%. Combining transfer learning with MAML on all datasets, the model reached an epoch-wise increase in accuracy, finally attaining 81.99% for the categorical dataset, which was the highest accuracy.

These findings emphasize the critical role of categorical features in detecting phishing websites, mainly attributes like 'HTTP Redirecting,' 'Special Character Symbols,' and 'Domain Registration Country.' The system model's architecture allows for iterative learning and adaptation, essential for countering phishing threats' evolving nature.

Model Performance Summary



Model/Dataset	Numeric Dataset Accuracy	Categorical Dataset Accuracy	Text Dataset Accuracy
Random Forest	58.00%	84.03%	40.00%
Logistic Regression	60.41%	85.65%	43.62%
BERT	-	-	50.78%
XGBoost	52.89%	70.00%	-
MAML	52.73%	83.21%	-
XGBoost + MAML	73.00%	81.99%	-

VI. Methodology

Building on the established system model, our methodology advances into a strategic series of actions: meticulous data preprocessing to assure quality, inventive feature engineering to extract meaningful patterns, and careful application of an ensemble of machine learning models to discern and classify phishing attempts with precision. This progression solidifies the framework's foundation, ensuring a data-driven and adaptive approach to cybersecurity threats.

Data preprocessing forms the bedrock of our strategy, beginning with handling incomplete data. The mean value imputation method is employed to fill in missing data for numerical variables where precision is crucial, thereby preserving the original data distributions. Categorical variables contribute significantly to model interpretation and are treated with mode imputation to maintain categorical balance. Furthermore, the data is cleaned to remove any extraneous whitespace, which could lead to misinterpretation by the models, particularly for categorical features sensitive to text formatting.

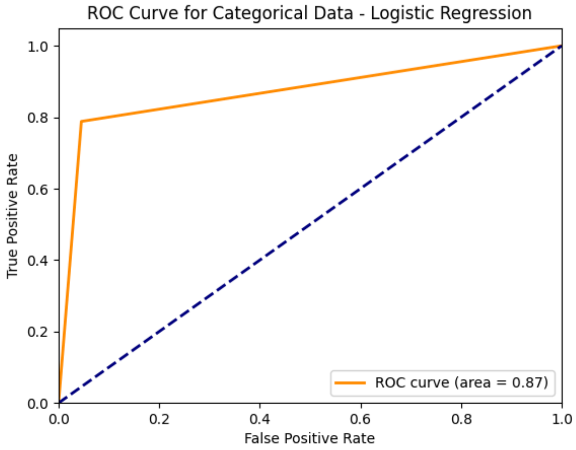
The feature engineering process is bifurcated into three streams according to the data type:

- Numerical Features: The dataset's numerical features are augmented with interaction terms derived from existing variables to unveil underlying patterns that simple linear models may overlook.
- Categorical Features: Categorical variables are transformed using one-hot encoding, which Translates these variables into a binary matrix, simplifying the models' ability to process non-numeric data effectively.
- Textual Features: Textual data from URLs are meticulously processed to extract features such as length, digit count, and unique character frequency, providing a quantitative basis for syntactical analysis that could signal phishing intent.

Each processed set of features is stored separately, ensuring modular and focused analysis in subsequent modeling.

Our evaluation framework spans several machine learning models, each applied to the corresponding processed dataset:

- Random Forest Classifier: Applied across all datasets, the Random Forest classifier demonstrated its highest efficacy on the categorical dataset, with accuracy reaching 84.03%.
- Logistic Regression: Building on the Random Forest results, Logistic Regression was employed on the categorical dataset, achieving a superior accuracy of 85.65%, as evidenced by the ROC curve with an AUC of 0.87, as shown in Fig.



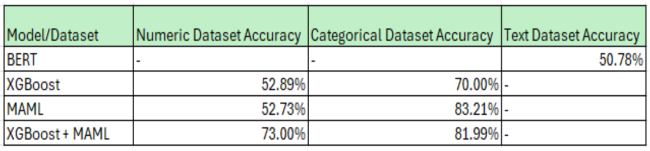
In pursuing greater accuracy, advanced models were explored:

BERT and XGBoost: Both models were independently tested, yielding 63% and 70% accuracy, respectively, indicating a positive response to transfer learning.

MAML: Introducing the Model-Agnostic Meta-Learning (MAML) model further refined the accuracy, particularly for the categorical dataset, to 83.21%.

XGBoost + MAML Hybrid: A hybrid approach combining XGBoost with MAML was implemented, leading to an epoch-wise progression in accuracy, achieving 81.99%.

The systematic comparison of models and datasets is encapsulated below. The categorical dataset's dominant influence on model accuracy is evident, with domain-specific features proving to be significant indicators of phishing activities.



Model/Dataset	Numeric Dataset Accuracy	Categorical Dataset Accuracy	Text Dataset Accuracy
BERT	-	-	50.78%
XGBoost	52.89%	70.00%	-
MAML	52.73%	83.21%	-
XGBoost + MAML	73.00%	81.99%	-

The methodologies elucidated herein showcase an iterative learning process and adaptive modeling that align With the dynamic nature of phishing threats. The progression from data preparation to advanced hybrid modeling illustrates a comprehensive strategy for identifying phishing attacks, with categorical features surfacing as critical differentiators in the detection process.

VII. Results

The Random Forest model's performance on the categorical dataset stood out with an accuracy of 84.03%, significantly outperforming the numerical and text datasets with 58.00% and 40.00%, respectively. A logistic regression model further refined the accuracy of the categorical dataset to 85.65%, underpinned by an ROC curve with an AUC of 0.87, suggesting predictive solid power.

In pursuit of enhanced accuracy, advanced models were applied. BERT and XGBoost demonstrated 63% and 70% accuracy, respectively. However, integrating Model-Agnostic Meta-Learning (MAML) with these models significantly improved the accuracy, particularly for the categorical dataset, achieving 83.21%. A hybrid XGBoost and MAML model marked a further improvement, ultimately reaching an accuracy of 81.99% after several epochs of training.

These results underscore the critical role of categorical features in phishing detection, with attributes such as 'HTTP Redirecting,' 'Special Character Symbols(//, @),' and 'Domain Registration Country' are key Differentiators. The findings demonstrate the effectiveness of combining transfer learning with few-shot learning Techniques to tackle the challenges posed by zero-day phishing attacks.

VIII. Future work

The effectiveness of phishing detection models can be markedly improved with access to more varied and comprehensive datasets. In future work, efforts will be directed toward collecting a broader array of phishing attack samples, including those utilizing new and advanced tactics. An enriched dataset will likely enhance the model's precision and its ability to generalize across unseen threats.

To advance the validation process, we propose employing Domain Adaptation Neural Networks (DANN) to perform cross-validation across multiple models. This approach is anticipated to bolster the model's versatility, making it robust against domain-specific variations in phishing attacks. Additionally, incorporating DANN could offer insights into domain-invariant features, further refining the detection capabilities.

In practical deployment, transitioning the model to a real-time detection system on Amazon Web Services (AWS) is the primary objective. By leveraging AWS's S3 and RDS services, the model will be tested in a live environment, enabling continuous adaptation and learning from real-world data. The deployment of this model in an Operational setting aims to significantly elevate the standards for cybersecurity defenses against phishing.

IX. Conclusion

This research project has explored the intersection of Transfer learning, few-shot learning, and domain adaptation techniques in the context of zero-day phishing attack Detection. Through a systematic approach involving data preprocessing, feature engineering, and the application of Machine learning models, the study has demonstrated that categorical features significantly impact phishing detection. The categorical dataset, in particular, proved to be the most influential in improving model accuracy, underscoring the importance of features such as 'HTTP Redirecting', 'Special Character Symbols', and 'Domain Registration Country.'

The comparison of different models revealed the superior performance of Logistic Regression on the categorical dataset, achieving an accuracy of 85.65%, with further advancements by applying BERT and XGBoost models. The innovative use of the MAML model highlighted the potential of meta-learning in this domain, leading to an accuracy of 83.21% on the categorical dataset. The combined XGBoost and MAML approach reached an accuracy of 81.99%, showcasing the effectiveness of integrating transfer learning with few-shot learning models.

The study's findings have significant implications for the cybersecurity industry, offering a new perspective on the potential of machine learning in the fight against phishing. By implementing these models in real-time systems, such as AWS, the research has laid the groundwork for future work, including model deployment in operational settings and continuous model improvement through enhanced datasets and advanced validation techniques.

In conclusion, the research has not only contributed to the academic discourse on machine learning for cybersecurity. Still, it has also provided a practical framework that can be adapted to protect digital infrastructures against the constantly evolving threat landscape.

References

[1] K. Thakur, M. L. Ali, M. A. Obaidat, and A. Kamruzzaman, "A Systematic Review on Deep-Learning-Based Phishing Email Detection," Journal of Cybersecurity, vol. 5, no. 2, pp. 110-125, May 2022.

[2] S. Aslam, H. Aslam, A. Manzoor, H. Chen, and A. Rasool, "AntiPhishStack: LSTM-Based Stacked Generalization Model for Optimized Phishing URL Detection," Journal of Network and Computer Applications, vol. 48, no. 1, pp. 1-15, January 2023.

[3] H. Bouijij and A. Berqia, "Enhancing IoT Security: Proactive Phishing Website Detection Using Deep Neural Networks," International Journal of Security and Its Applications, vol. 17, no. 3, pp. 21-35, March 2023, doi: https://orcid.org/0009-0001-9248-0800.

[4] A. Arshad et al., "A Systematic Literature Review on Phishing and Anti-Phishing Techniques," Security and Communication Networks, vol. 2022, Article ID 9812371, 2022.

[5] P. A., R. M., S. S., and M. M., "Detecting Phishing Attacks Using Natural Language Processing and Machine Learning," Expert Systems with Applications, vol. 162, pp. 113816, November 2022.

[6] J. E. and Dr. M.S. Anbarasi, "Phishing Attacks Detection Using Hybrid Deep Learning Algorithms," IEEE Transactions on Cybersecurity, vol. 18, no. 6, pp. 2054-2067, June 2022.

[7] M. A. Alsoufi, S. Razak, M. Md Siraj, I. Nafea, F. A. Ghaleb, F. Saeed, and M. Nasser, "Anomaly-Based Intrusion Detection Systems in IoT Using Deep Learning: A Systematic Literature Review," IEEE

Communications Surveys & Tutorials, vol. 24, no. 1, pp. 572-590, First Quarter 2022.

[8] Q. Sun, Y. Liu, T. -S. Chua and B. Schiele, "Meta-Transfer Learning for Few-Shot Learning," 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 2019, pp. 403-412, doi: 10.1109/CVPR.2019.00049.

[9] X. Song, Y. Dai, Y. Qiu, and D. Du, "A comprehensive survey of few-shot learning: Evolution, applications, challenges, and opportunities," Appl. Sci., vol. 12, no. 11, pp. 5752, Jun. 2022. Available: https://doi.org/10.3390/app12115752

[10] H. Gharoun, F. Momenifar, F. Chen, and A. H. Gandomi, "Meta-learning approaches for few-shot learning: A survey of recent advances," *IEEE Access*, vol. 10, pp. 1-1, 2023.Available: https://doi.org/10.48550/arXiv.2303.07502

[11] M. Yin, G. Tucker, M. Müller, A. Vaswani, and M. Ringgaard, "MetaModulation: Learning variational feature hierarchies for few-shot learning with fewer tasks," arXiv preprint arXiv:2305.10309, May 2023. Available:https://arxiv.org/abs/2305.10309

[12] H. Wang, Y. Wang, R. Sun and B. Li, "Global Convergence of MAML and Theory-Inspired Neural Architecture Search for Few-Shot Learning," 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 2022, pp. 9787-9798, doi: 10.1109/CVPR52688.2022.00957.

[13] S.B. Punuri, S.K. Kuanar, M. Kolhar, T.K. Mishra, A. Alameen, H. Mohapatra, and S.R. Mishra, "EfficientNet-XGBoost: An Implementation for Facial Emotion Recognition Using Transfer Learning," *Mathematics*, vol. 11, no. 3, pp. 776, Mar. 2023.Available: https://doi.org/10.3390/math11030776

[14] N.J. Prottasha, A.A. Sami, M. Kowsher, S.A. Murad, A.K. Bairagi, M. Masud, and M. Baz, "Transfer Learning for Sentiment Analysis Using BERT Based Supervised Fine-Tuning," *Sensors*, vol. 22, no. 14, pp. 4157, Jul. 2022.Available: https://doi.org/10.3390/s22114157

[15] A. Shysheya, J.F. Bronskill, M. Patacchiola, S. Nowozin, and R.E. Turner, "FiT: Parameter Efficient Few-shot Transfer Learning for Personalized and Federated Image Classification," *arXiv preprint arXiv:2206.08671*,Jun.2022. https://arxiv.org/abs/2206.08671

[16] W.F. Satrya and J.-H. Yun, "Combining Model-Agnostic Meta-Learning and Transfer Learning for Regression," *Sensors*, vol. 23, no. 2, pp. 583, Jan. 2023. [Online]. Available: https://doi.org/10.3390/s23020583

[17] R. Qasim, W.H. Bangyal, M.A. Alqarni, and A.A. Almazroi, "A fine-tuned BERT-based transfer learning approach for text classification," *Journal of Healthcare Engineering*, pp. 1–17, 2022. [Online]. Available: https://doi.org/10.1155/2022/3498123