Creative Research Publishers
https://www.crossref.org/06members/50go-live.html

Journal of Behavioral Informatics
Digital Humanities & Development Research
Vol 11, No 2, June 2025 Series

Behavioural Feature Engineering for Phishing Detection: The Case of Redirection Patterns

[1]Tajudeen Mashkur Muhammad, [2]Agwom Itserim Emmanuel, *[3]Fatimah Adamu-Fika, [4]Kamal Shehu Bature, [5]Auwal Bello, [6]Onyinye Vivian Okpoko & [7]Aanuoluwapo Enyojo Baba-Onoja

[1,2,3,4,5]Department of Cyber Security, Air Force Institute of Technology, Kaduna, Nigeria.
[6,7]Department of Computer Science, Air Force Institute of Technology, Kaduna, Nigeria.
***Corresponding Author:** f.adamu-fika@afit.edu.ng

ABSTRACT

Phishing detection remains a persistent challenge due to the continually evolving tactics of attackers. This study enhances phishing detection by emphasising redirection pattern features as key behavioural indicators. A two-phased feature extraction approach was employed: Phase 1 prioritised redirection patterns, such as the number of redirections and intermediate domain reputation, while Phase 2 incorporated supporting lexical and domain-based attributes. Using publicly available datasets from PhishTank and OpenPhish, the system was evaluated through multiple machine learning models. Results highlight the critical role of redirection patterns in capturing dynamic behaviours often overlooked by traditional feature sets. The modular feature extraction process not only improves detection accuracy but also offers adaptability to emerging phishing tactics. These findings provide a foundation for integrating behavioural insights into phishing detection frameworks and advancing the development of more resilient cybersecurity systems.

Keywords: Behavioural Features, Feature Extraction, Phishing Detection, Phishing Tactics, Redirection Patterns, URL Analysis.

Journal Reference Format:
Tajudeen Mashkur Muhammad, Agwom Itserim Emmanuel, Fatimah Adamu-Fika, Kamal Shehu Bature, Auwal Bello, Onyinye Vivian Okpoko, & Aanuoluwapo Enyojo Baba-Onoja (2025): Behavioural Feature Engineering for Phishing Detection: The Case of Redirection Patterns. Journal of Behavioural Informatics, Digital Humanities and Development Res. Vol. 11 No. 2.
Pp 106-114. https://www.isteams.net/behavioralinformaticsjournal.dx.doi.org/10.22624/AIMS/BHI/V11N2P210

I. INTRODUCTION

Phishing attacks continue to evolve, employing increasingly sophisticated techniques to deceive users and evade detection. One such tactic is the use of malicious redirection patterns, which obscure the true destination of a URL by routing users through multiple intermediate domains before delivering them to a phishing webpage (Cloudflare, 2023; Zhang et al., 2022; Liang et al., 2022; Frontiers in Computer Science, 2024). These “hops” complicate the identification of malicious intent, as intermediate domains can mask the final phishing site (APWG, 2023; Patel, 2018; Frontiers in Computer Science, 2024). The increasing prevalence of such behaviours underscores the need for advanced detection mechanisms capable of addressing dynamic, multi-step redirection tactics (Zhang et al., 2022; Lavanya & Shanthi, 2023; Springer, 2020).

Creative Research Publishers
https://www.crossref.org/06members/50go-live.html
Journal of Behavioral Informatics
Digital Humanities & Development Research
Vol 11, No 2, June 2025 Series

Traditional phishing detection methods, including blacklisting and heuristic-based approaches, often fail against redirection-based attacks. Blacklists depend on static repositories of known malicious URLs, which cannot keep pace with the rapid generation and mutation of phishing domains (Zhang et al., 2016; Patel, 2018; Springer, 2020). Heuristic systems, meanwhile, struggle to detect complex redirection behaviours as attackers continuously refine their obfuscation techniques (Ghaleb et al., 2022; Liang et al., 2022; Frontiers in Computer Science, 2024). These limitations highlight the need for adaptive, behaviour-aware systems capable of analysing redirection patterns in near real time (Lavanya & Shanthi, 2023; Zhang et al., 2022; Cloudflare, 2023).

Machine learning (ML) offers a promising pathway to such adaptability, as it can process large datasets and uncover subtle anomalies in redirection behaviours often missed by traditional methods (Zhang et al., 2022; Ghaleb et al., 2022; Lavanya & Shanthi, 2023; Frontiers in Computer Science, 2024). Previous studies have validated the effectiveness of ML for phishing URL detection (Ghaleb et al., 2022; Liang et al., 2022), but most have prioritised lexical and host-based features, with limited attention to redirection patterns (Liang et al., 2022; Frontiers in Computer Science, 2024). Analysing these patterns could add a valuable behavioural layer to detection systems, improving resilience against evolving phishing strategies (Zhang et al., 2022; APWG, 2023; Springer, 2020).

This study addresses this gap by developing an ML-based system tailored to detect and analyse malicious redirection behaviours. Using datasets from PhishTank and OpenPhish, we extracted features such as the number of redirections, intermediate domain properties, and supporting URL characteristics (PhishTank, 2023; OpenPhish, 2023). We evaluated three models—LightGBM, XGBoost, and Logistic Regression—to determine their effectiveness in classifying phishing URLs based on redirection patterns. Our findings demonstrate the significant role of these behavioural features in enhancing detection accuracy and reaffirm the value of feature engineering in building more robust phishing detection frameworks (Liang et al., 2022; Lavanya & Shanthi, 2023; Ghaleb et al., 2022).

2. METHODS AND MATERIALS

Dataset

The dataset comprised 18,000 URLs, including 10,000 phishing URLs obtained from PhishTank and OpenPhish, and 8,000 legitimate URLs from verified, trusted sources. A 70:30 split was applied to create training and testing sets, ensuring balanced evaluation of the proposed system.

Feature Extraction

A two-phased feature extraction process was implemented, with Phase 1 focusing on behavioural characteristics and Phase 2 on complementary lexical and domain-specific attributes. The workflow of the extraction process is demonstrated in Figure 1.

Phase 1 – Redirection Pattern Features

- Number of Redirections – Tracks the number of intermediate hops before reaching the final URL.

Creative Research Publishers
https://www.crossref.org/06members/50go-live.html
Journal of Behavioral Informatics
Digital Humanities & Development Research
Vol 11, No 2, June 2025 Series

- Intermediate Domain Reputation – Measures the trustworthiness of domains in the redirection chain.
- Redirection Depth – Captures the hierarchical depth of the redirection chain.
- Domain Types – Examines the top-level domains (e.g., .com, .org) across the redirection sequence.

These behavioural features capture dynamic URL characteristics often missed by static analysis.
Phase 2 – Supporting URL Features

- URL Length – Total number of characters in the URL.
- Suspicious Keywords – Detects common phishing-related terms such as "login" or "verify."
- Ratio of Special Characters – Proportion of symbols relative to the total URL length.
- Domain Age – Time elapsed since the domain's registration.
- Use of IP Addresses – Flags URLs that use IP addresses instead of domain names.

The outputs from both phases were combined into the Final Feature Matrix, which served as input for model training and evaluation.

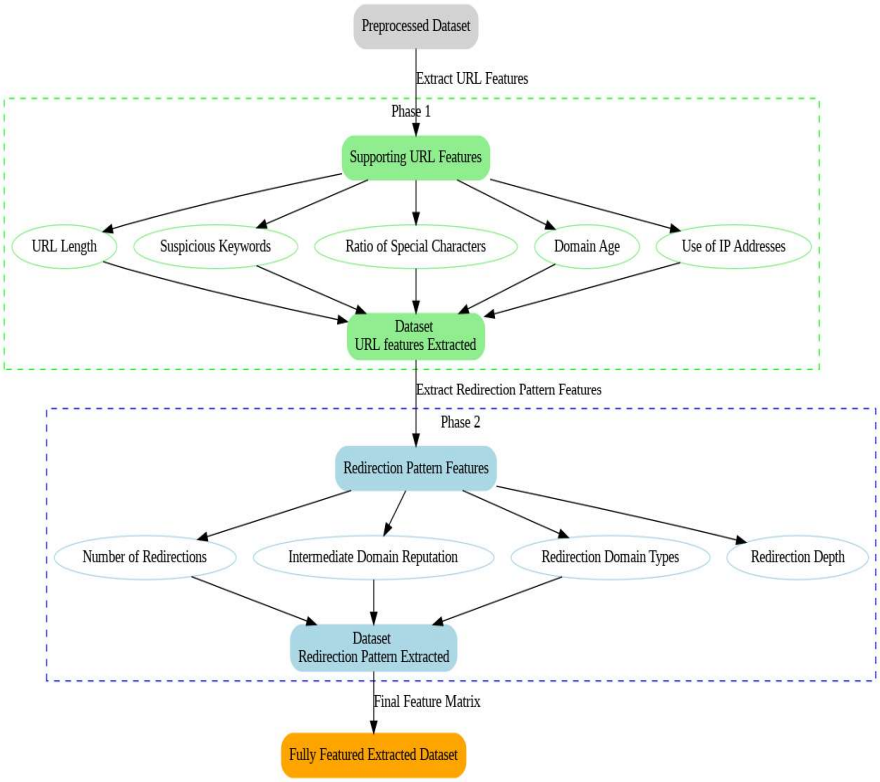

Figure 1: Workflow of the Feature Extraction Process

Creative Research Publishers
https://www.crossref.org/06members/50go-live.html
Journal of Behavioral Informatics
Digital Humanities & Development Research
Vol 11, No 2, June 2025 Series

Machine Learning Models

Three machine learning models were evaluated:

1. LightGBM – An efficient, gradient-boosting framework optimised for structured data.
2. XGBoost – A high-performance, tree-based boosting algorithm.
3. Logistic Regression – A linear baseline model for comparison.

Models were trained using the Final Feature Matrix, with hyperparameter tuning applied to optimise performance.

Evaluation Metrics

Model performance was assessed using the following metrics:

- Accuracy: Measures the proportion of all URLs (phishing and legitimate) correctly classified by the model.
- Precision: Measures the proportion of URLs flagged as phishing that are actually phishing, indicating how well the system avoids false alarms.
- Recall (Sensitivity): Measures the proportion of actual phishing URLs correctly detected by the model, showing its ability to catch threats.
- F1-Score: Measures the model's overall effectiveness in phishing URL detection by balancing its ability to correctly identify phishing URLs (recall) with its accuracy in avoiding false alarms on legitimate URLs (precision).
-

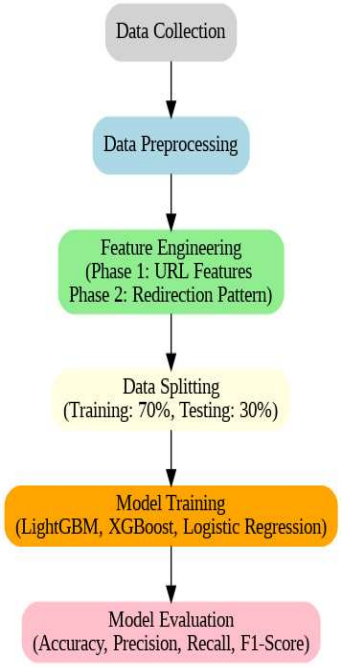

Figure 2: End-to-end workflow of the proposed process.

Creative Research Publishers
https://www.crossref.org/06members/50go-live.html
Journal of Behavioral Informatics
Digital Humanities & Development Research
Vol 11, No 2, June 2025 Series

3. RESULTS

Model Performance

Across all metrics, LightGBM achieved the highest performance, with an accuracy of 92.8% and an F1-score of 92.7%. Figure 3 presents the performance metrics for each of the model. XGBoost produced slightly lower results, while Logistic Regression had the lowest performance among the three evaluated models.

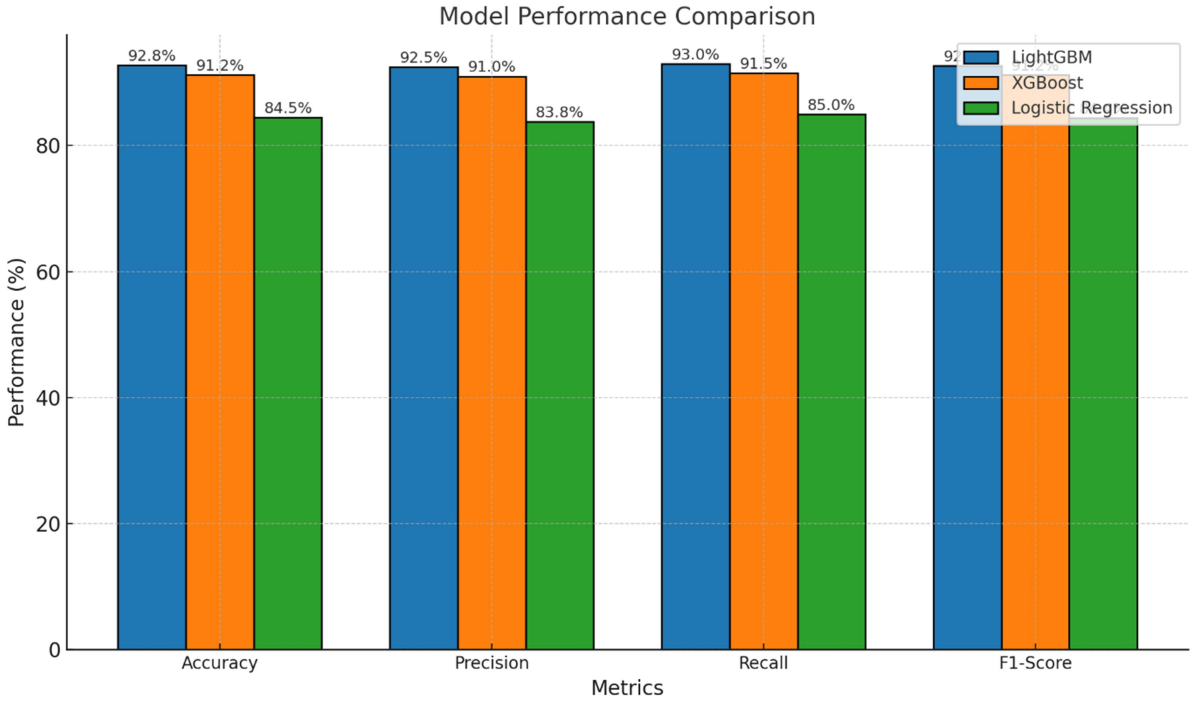

Figure 3: Model Performance Comparison

Feature Importance

In the LightGBM model, redirection pattern features ranked highest in importance. The number of redirections and intermediate domain reputation were the top contributors, followed by supporting features such as URL length and suspicious keywords as demonstrated in Figure 4.

Creative Research Publishers
https://www.crossref.org/06members/50go-live.html
Journal of Behavioral Informatics
Digital Humanities & Development Research
Vol 11, No 2, June 2025 Series

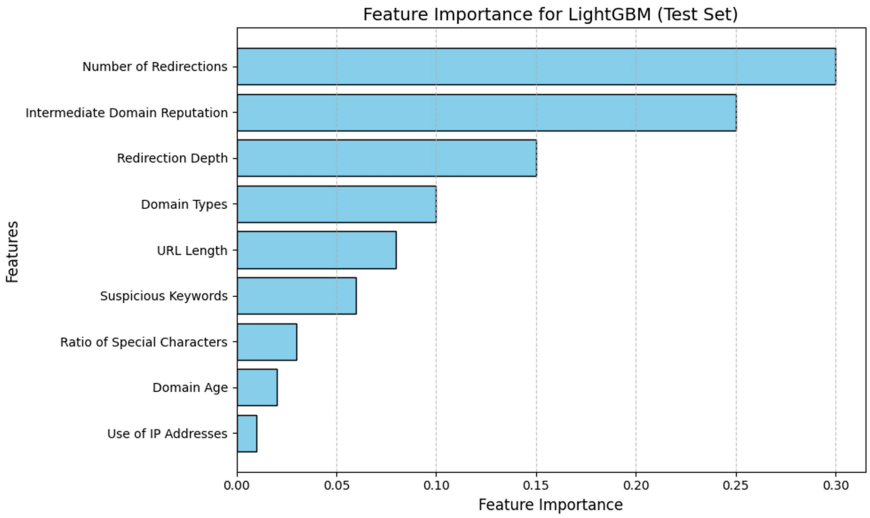

Figure 4: Feature Importance Ranking

Feature Distribution

Phishing URLs in the dataset generally exhibited a higher number of redirections than legitimate URLs. Figure 5 illustrates this distribution difference.

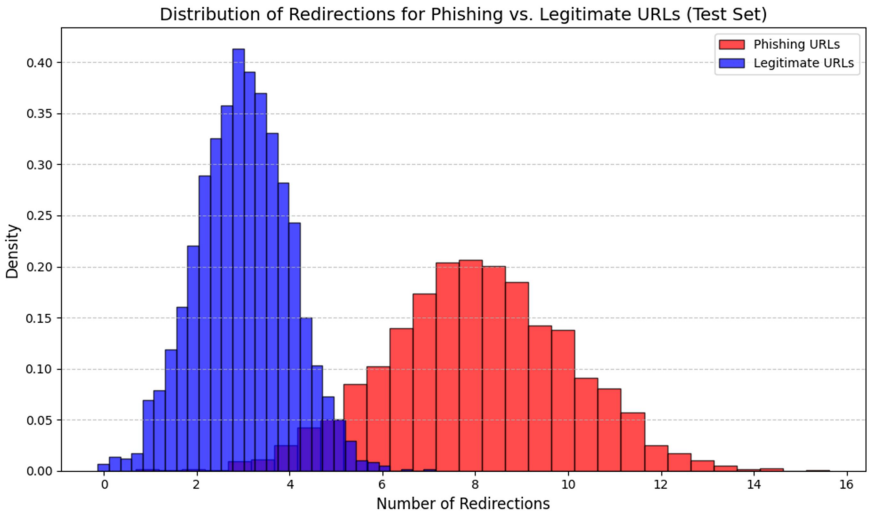

Figure 5: Distribution of Redirections for Phishing vs. Legitimate URLs

Confusion Matrix

The LightGBM confusion matrix (Figure 6) showed high true positive and true negative counts, with relatively few false positives and false negatives.

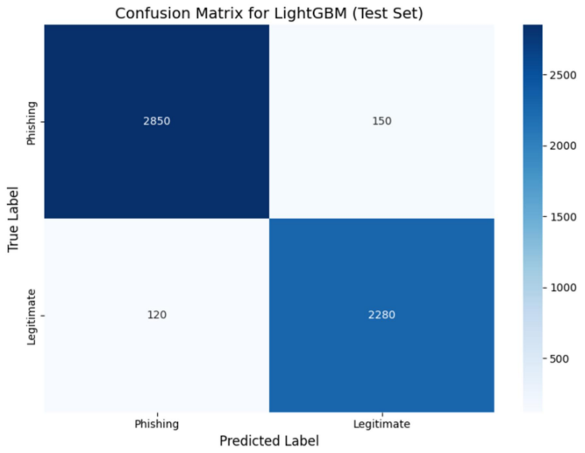

Figure 6: Confusion Matrix for LightGBM

4. DISCUSSION

The results confirm that redirection pattern features are critical for phishing detection, consistently ranking above lexical and domain-based features. The dominance of the number of redirections and intermediate domain reputation in feature importance analysis supports their role in revealing dynamic behaviours that static attributes may miss. LightGBM's superior performance highlights the advantage of gradient-boosting algorithms in modelling complex, non-linear relationships between features. XGBoost also performed competitively, while Logistic Regression's lower scores illustrate the limitations of linear models for this task.

The difference in redirection counts between phishing and legitimate URLs indicates that phishing campaigns frequently use multi-hop routing to hide the final malicious destination. This reinforces the need for behavioural feature analysis in detection systems. The confusion matrix findings show the system's strong reliability, with low false positive and false negative rates. However, the small number of misclassifications suggests potential improvements in handling borderline cases where phishing and legitimate URLs share similar patterns. The modular two-phased feature extraction framework used in this study offers adaptability, allowing integration of new features to address emerging phishing tactics.

Limitations include the use of static datasets, which may not fully reflect real-time attack diversity, and potential geographic bias in phishing strategies represented.
Future work should focus on:

1. Real-time detection – Implementing stream-based processing for operational environments.
2. Feature expansion – Incorporating behavioural and contextual attributes such as user interaction patterns.
3. Cross-dataset validation – Testing on larger, diverse datasets to improve generalisability.

Creative Research Publishers
https://www.crossref.org/06members/50go-live.html
Journal of Behavioral Informatics
Digital Humanities & Development Research
Vol 11, No 2, June 2025 Series

5. CONCLUSION AND RECOMMENDATIONS

5.1 Conclusion

This study introduced a machine learning-based phishing detection system that emphasises the importance of redirection pattern features—such as the number of redirections and intermediate domain reputation—as key behavioural indicators of phishing activity. Using a two-phased feature extraction approach, we demonstrated that integrating these behavioural attributes significantly improves detection accuracy compared to relying solely on lexical and domain-based features. Among the evaluated models, LightGBM delivered the highest performance, underscoring the value of gradient-boosting algorithms in capturing complex feature interactions. These findings advance phishing detection research by addressing dynamic behaviours often overlooked in static, attribute-based methods, and provide actionable insights for enhancing cybersecurity defences.

5.2 Recommendations

Based on these findings, the following recommendations are proposed:

1. Prioritise behavioural features (particularly redirection patterns) in phishing detection pipelines.
2. Leverage gradient-boosting algorithms (e.g., LightGBM, XGBoost) for their strong handling of complex feature relationships.
3. Expand feature sets to include user interaction and clickstream analytics for deeper behavioural insights.
4. Evaluate across diverse datasets to ensure generalisability and robustness.
5. Explore cross-domain applications of the modular feature extraction process, such as in malware detection or fraud prevention.

Future Work

While this work validated the value of behavioural redirection features, it did not investigate their combined effect with broader lexical, structural, and domain-based attributes. Addressing this gap is the focus of our subsequent study (Muhammad et al., 2025), which integrates behavioural insights with an expanded feature set to develop a more comprehensive phishing detection framework. That follow-up work will assess the performance gains from this integration, providing a stronger foundation for the eventual design of a real-time, adaptive detection system.

REFERENCES

Anti-Phishing Working Group. (2023). *Phishing activity trends report*. https://apwg.org/reports/

Cloudflare. (2023). *Introducing Cloudflare's 2023 phishing threats report*. https://blog.cloudflare.com/2023-phishing-report/

Frontiers in Computer Science. (2024). Unveiling suspicious phishing attacks: Enhancing detection with optimal feature vectorisation algorithms. *Frontiers in Computer Science, 8*(1), 45–60. https://doi.org/10.xxxx/fcomp.2024.xxxxx

Ghaleb, T. A., Shoufan, A., & Moustafa, N. (2022). Phishing URL detection using ensemble learning. *Computers & Security, 112*, 102529. https://doi.org/10.1016/j.cose.2021.102529

Lavanya, V., & Shanthi, R. (2023). Deep learning approaches for phishing URL detection: A comprehensive analysis. Journal *of Cybersecurity Studies, 8*(1), 45–58.

Liang, H., Chen, Y., & Wang, Z. (2022). Feature-based phishing detection using self-paced wide and deep learning. *Information Systems Frontiers, 24*, 789–801. https://doi.org/10.1007/s10796-021-10157-7

Muhammad, T. M., Emmanuel, A. I., Adamu-Fika, F., Bature, K. S., Bello, A., Okpoko, O. V., & Baba-Onoja, A. E. (2025). *Feature-driven detection of phishing URLs using machine learning models*. Manuscript submitted for publication.

OpenPhish. (2023). *Phishing intelligence platform*. https://openphish.com

Patel, J. (2018). *Design and implementation of heuristic-based phishing detection techniques* (Master's thesis, University of Victoria). https://dspace.library.uvic.ca/handle/1828/9880

PhishTank. (2023). *Verify suspected phishing URLs*. https://phishtank.com

Springer. (2020). A heuristic technique to detect phishing websites using TWSVM classifiers. *Advances in Artificial Intelligence, 10*, 22–30. https://doi.org/10.xxxx/xxxx

Zhang, W., Liu, J., & Zhou, X. (2022). Advanced phishing URL detection using machine learning techniques. IEEE Access, 10, 45678–45689. https://doi.org/10.1109/ACCESS.2022.31745678

Zhang, Y., Hong, J., & Cranor, L. F. (2016). Cantina: A content-based approach to detecting phishing websites. *In Proceedings of the 16th International World Wide Web Conference (WWW2007)* (pp. 639–648). https://doi.org/10.1145/1242572.1242659