***Korshunova Veronika Yuryevna***
Master's student
Reshetnev Siberian State University of Science and Technology
Krasnoyarsk, Krasnoyarsk region

# LOGICAL CLASSIFICATION ALGORITHM FOR PHISHING DOMAIN DETECTION IN CYBERSECURITY ANALYSIS

***Abstract:*** *this study investigates the application of decision trees for detecting phishing domains based on URL attributes. A logical classification algorithm searching for conjunctive patterns was applied to a dataset comprising 112 structural and server-related features. The decision tree model achieved an error rate of 3.26%. Key significant factors identified include the quantity of slashes in the URL (40.3%), domain activation time (16.8%), and directory length (14.06%). The results confirm the high efficiency and interpretability of rule-based machine learning for cybersecurity threat mitigation.*

***Keywords:*** *cybersecurity, phishing domains, machine learning, decision tree, threat detection.*

***Коршунова Вероника Юрьевна***
студентка
ФГБОУ ВО «Сибирский государственный университет
науки и технологий им. академика М.Ф. Решетнева»
г. Красноярск, Красноярский край

# ЛОГИЧЕСКИЙ АЛГОРИТМ КЛАССИФИКАЦИИ ДЛЯ ОБНАРУЖЕНИЯ ФИШИНГОВЫХ ДОМЕНОВ ПРИ АНАЛИЗЕ КИБЕРБЕЗОПАСНОСТИ

***Аннотация:*** *в статье рассматривается применение деревьев решений для обнаружения фишинговых доменов на основе атрибутов URL-адресов. Логический алгоритм классификации, осуществляющий поиск конъюнктивных закономерностей, был применен к набору данных, включающему 112 структурных и серверных признаков. Модель дерева решений достигла уровня ошибки 3,26%. Ключевыми значимыми факторами, выявленными в ходе исследования, стали:*



Content is licensed under the Creative Commons Attribution 4.0 license (CC-BY 4.0)

*количество косых черт (слэшей) в URL-адресе (40,3%), время активации домена (16,8%) и длина каталога (14,06%). Полученные результаты подтверждают высокую эффективность и интерпретируемость методов машинного обучения на основе правил для минимизации угроз кибербезопасности.*

***Ключевые слова:*** *кибербезопасность, фишинговые домены, машинное обучение, дерево решений, обнаружение угроз.*

*Introduction*: in the modern digital landscape, phishing remains one of the most prevalent cyberattack vectors, utilizing deceptive websites to harvest confidential information. Detecting malicious domains requires robust analytical tools capable of processing high-dimensional URL features. Machine learning (ML), particularly supervised classification, has emerged as a critical component in threat intelligence. Among various algorithms, decision trees offer a transparent, rule-based structure that facilitates the identification of hidden patterns characterizing phishing endeavors while maintaining high computational efficiency.

*Methodology*: the research utilizes a comprehensive dataset of URL attributes tailored for phishing domain identification. The dataset encompasses 112 columns, including character counts, domain registration parameters, SSL certificate statuses, and server response metrics. The target variable is a binary phishing indicator (0 or 1). Data preprocessing involved handling missing values, outlier elimination, and correlation analysis with a 0.05 threshold to remove insignificant features. The primary classification algorithm employed is a decision tree, which constructs a hierarchical model of decisions based on feature splits. This approach is highly suitable for cybersecurity tasks due to its ability to process both categorical and numerical data and automatically perform feature selection.

*Results*: the constructed decision tree model demonstrated high predictive accuracy, yielding an error rate of merely 3.26%. The algorithmic transparency allowed for the extraction of attribute significance, revealing the most critical indicators of phishing activity:


Содержимое доступно по лицензии Creative Commons Attribution 4.0 license (CC-BY 4.0)

- *Quantity of slashes in the URL (qty_slash_url):* Contributing 40.3% to the model, a high number of slashes often indicates deep nesting or complex site structures designed to mimic legitimate resources;

- *Domain activation time (time_domain_activation):* Accounting for 16.8%, recently registered domains are frequently utilized for transient phishing campaigns;

- *Directory length (directory_length):* With a 14.06% weight, elongated directory paths often signify obfuscation attempts to exploit user trust.

*Conclusion*: the logical classification algorithm based on decision trees proves to be a powerful tool for phishing domain detection. The low error rate (3.26%) combined with high interpretability enables security analysts to understand attacker strategies and develop adaptive blocking rules. Future research will focus on integrating ensemble methods to further reduce false positives in dynamic threat landscapes.

***References***

1. Razrabotka algoritma avtomaticheskoi klassifikatsii narushennykh gornymi rabotami zemel' po dannym distantsionnogo zondirovaniya [Development of an algorithm for automatic classification of lands disturbed by mining operations based on remote sensing data] / O. S. Ermolaeva, A. A. Stupina, S. I. Gorbachev, G. V. Korableva // Izvestiya Tul'skogo gosudarstvennogo universiteta. Nauki o Zemle [News of the Tula State University. Earth Sciences]. – 2026. – No. 3. – P. 167–180.

2. Kleshko I. I. Primenenie algoritmov mashinnogo obucheniya dlya vyyavleniya anomal'nykh pokazanii datchikov kar'ernogo vodootliva [Application of machine learning algorithms for detecting anomalous readings of quarry drainage sensors] / I. I. Kleshko, I. A. Vakhrusheva, G. V. Korableva // Izvestiya Tul'skogo gosudarstvennogo universiteta. Nauki o Zemle [News of the Tula State University. Earth Sciences]. – 2026. – No. 3. – P. 235–248.

3. Interpretiruemye modeli mashinnogo obucheniya v zadachakh otsenki ekologicheskoi i promyshlennoi bezopasnosti gornodobyvayushchikh predpriyatii [Interpretable machine learning models in tasks of assessing environmental and industrial safety of mining enterprises] / L. A. Kazakovtsev, A. S. Prudkii, I. A. Vakhrusheva, A. N. Losev // Izvestiya Tul'skogo gosudarstvennogo universiteta. Nauki o Zemle



Content is licensed under the Creative Commons Attribution 4.0 license (CC-BY 4.0)

[News of the Tula State University. Earth Sciences]. – 2026. – No. 1. – P. 169–174. – EDN GNIXFD.

4. Integratsiya geoinformatsionnykh sistem i algoritmov iskusstvennogo intellekta dlya monitoringa geotekhnicheskoi ustoichivosti ob»ektov gornoi infrastruktury [Integration of geographic information systems and artificial intelligence algorithms for monitoring geotechnical stability of mining infrastructure objects] / L. A. Kazakovtsev, A. S. Prudkii, I. A. Vakhrusheva, M. S. Nikanorov // Izvestiya Tul'skogo gosudarstvennogo universiteta. Nauki o Zemle [News of the Tula State University. Earth Sciences]. – 2026. – No. 1. – P. 354–359. – EDN EGFSPA.

5. Intellektual'naya model' prognozirovaniya ob»emov obrazovaniya otkhodov na osnove metodov mashinnogo obucheniya [Intelligent model for forecasting waste generation volumes based on machine learning methods] / T. A. Panfilova, V. S. Tynchenko, B. V. Potapov, M. S. Nikanorov // Ustoichivoe razvitie gornykh territorii [Sustainable Development of Mountain Territories]. – 2026. – Vol. 18, No. 2 (68). – P. 1107–1120. – DOI 10.21177/1998-4502-2026-18-2-1107-1120. – EDN GTTEPG.

6. Panfilov I. A. Neirosetevoi metod operezhayushchego vyyavleniya predavariinykh peregruzok v rudnichnykh raspredelitel'nykh setyakh [Neural network method for early detection of pre-emergency overloads in mine distribution networks] / I. A. Panfilov, M. V. Zhuravlev, E. F. Malykha // Bezopasnost' truda v promyshlennosti [Occupational Safety in Industry]. – 2026. – No. 7. – P. 7–13. – DOI 10.24000/0409-2961-2026-7-7-13.


Содержимое доступно по лицензии Creative Commons Attribution 4.0 license (CC-BY 4.0)