МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ ХАРЧОВИХ ТЕХНОЛОГІЙ
НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ БІОРЕСУРСІВ І ПРИРОДОКОРИСТУВАННЯ УКРАЇНИ
SZKOŁA GŁÓWNA GOSPODARSTWA WIEJSKIEGO W WARSZAWIE
POZNAŃ UNIVERSITY OF LIFE SCIENCES
POLITECHNIKA WARSZAWSKA

Факультет автоматизації і комп'ютерних систем

XII Міжнародна науково-технічна
Internet-конференція

«Сучасні методи, інформаційне, програмне та технічне забезпечення систем керування організаційно-технічними та технологічними комплексами»

27 листопада 2025

КИЇВ НУХТ 2025

Матеріали XII Міжнародної науково-технічної Internet-конференції «Сучасні методи, інформаційне, програмне та технічне забезпечення систем керування організаційно-технічними та технологічними комплексами», 27 листопада 2025 [Електронний ресурс]. – К: НУХТ, 2025. – 390 с. – Режим доступу: https://nuft.edu.ua/naukova-diyalnist/naukovi-konferencii.

У матеріалах конференції наведено доповіді за напрямами: автоматизація процесів керування технологічними процесами та комплексами, інтелектуальні системи керування та аналізу даних, інтегроване автоматизоване керування організаційно-технічними системами, інформаційні системи керування у виробництві та освіті. Видання містить програму і матеріали Міжнародної науково-технічної конференції.

Матеріали конференції будуть корисні науковим та інженерно-технічним працівникам, виробничникам, потенційним інвесторам, студентам вищих закладів освіти та всім, хто пов'язаний з харчовою промисловістю та автоматизацією.

Подано в авторській редакції.

ISBN 978-966-612-391-9

UDC 004

Information System for Corporate Email Analysis Using Machine Learning Methods

Bereza M., Hrama M.
National University of Food Technologies, Kyiv, Ukraine

In today's environment of digital transformation in the corporate sector, email remains one of the key channels of business communication, but at the same time, it is the main tool for phishing attacks, account compromise, and the spread of malware. According to statistics from international analytical groups, up to 90% of cyberattacks begin with an email containing a link or attachment that prompts the recipient to visit a malicious website or disclose confidential information. The evolution of phishing techniques and the use of social engineering and generative artificial intelligence by attackers to create plausible messages has made it much more difficult to detect such threats using traditional signature-based mechanisms. To increase the security level of the corporate email environment, an information system has been proposed that provides intelligent analysis of emails using natural language processing (NLP) and machine learning methods. The developed solution is focused on automated detection of phishing signs, verification of sender domain reliability, link reputation analysis, and interactive blocking of potentially dangerous addresses. The system architecture includes components for receiving and routing mail (Postfix), a storage server (Dovecot), a message pre-checking mechanism (AI Milter), a REST API for rapid response, and a centralized monitoring subsystem using Loki and Grafana. Integration with SPF, DKIM, and DMARC DNS protection mechanisms provides additional confirmation of the authenticity of email sources. The DistilBERT machine learning model was reasonably chosen as the base classification model due to its ability to perform deep semantic analysis of texts, detect manipulative constructions, and adaptive phishing patterns formed using artificial intelligence. Previous studies have demonstrated the suitability of transformer models for detecting phishing even in cases of grammatically correct and stylistically authentic messages. The system is trained on a mixed set of corporate and open email datasets, which allows it to take into account the specifics of internal document flow. During experimental operation, the system processed over 17,000 messages, ensuring 94–97% classification accuracy, a noticeable reduction in false positives, and shorter response times thanks to the ability to instantly block suspicious domains via the Grafana interface. Message processing logs in JSON format are stored in Loki and used to build dashboards that reflect threat dynamics, key attack sources, domain geography, and typical phishing patterns. The system can be integrated with SIEM/SOAR solutions, generate analytical alerts, and serve as a basis for further expansion using threat intelligence methods. The proposed approach demonstrates the feasibility of combining machine learning, automated response, and central monitoring as an effective corporate email protection strategy. The system can be expanded by using ensemble classification models, adding multilingual training corpora, implementing sandbox attachment scanning, and adaptive real-time analysis.