CODEBERT DENGAN TRANSFER LEARNING UNTUK MENDETEKSI KERENTANAN CROSS-SITE SCRIPTING PADA KODE SUMBER

M. RAFI SYAFRINALDI

21/472772/PA/20335

ABSTRAK

Cross-site scripting (XSS) masih menjadi masalah kritis dalam keamanan web, karena penyerang memanfaatkan kerentanan dalam aplikasi web untuk menjalankan skrip berbahaya, yang mengakibatkan pelanggaran keamanan yang serius. Metode keamanan tradisional sering kali kurang efektif dalam mendeteksi serangan XSS yang canggih, menghasilkan tingkat deteksi palsu yang tinggi dan kerentanan yang terlewat. Untuk mengatasi kekurangan ini, tesis ini meningkatkan CodeBERT, model bahasa yang telah dilatih sebelumnya, dengan menyetel ulang menggunakan dataset khusus yang dirancang untuk meningkatkan identifikasi kerentanan XSS. Pendekatan ini bertujuan untuk secara signifikan mengurangi deteksi palsu dan memperkuat keamanan aplikasi web.

Untuk mengatasi masalah kritis ini, tesis ini mengusulkan pendekatan inovatif menggunakan CodeBERT, model bahasa yang telah dilatih sebelumnya, yang ditingkatkan dengan pembelajaran transfer. Metodologi ini melibatkan pengumpulan dan prapemrosesan dataset kode aplikasi web yang beragam, diikuti dengan penyetelan ulang lapisan atas CodeBERT untuk mengkhususkan dalam pengenalan pola XSS. Pendekatan ini memanfaatkan kemampuan pemahaman bahasa lanjutan CodeBERT untuk meningkatkan akurasi deteksi serangan XSS.

Evaluasi model CodeBERT yang telah disetel ulang menunjukkan kinerja yang luar biasa, terutama pada dataset PHP, di mana ia secara signifikan mengungguli pendekatan terkini sebelumnya. Model ini mencapai akurasi, presisi, recall, dan skor F1 yang tinggi, menunjukkan kemampuannya untuk meminimalkan negatif palsu yang mana sebuah atribut penting dalam konteks tugas-tugas yang berhubungan dengan keamanan. Kontribusi yang diharapkan dari penelitian ini adalah peningkatan substansial dalam deteksi XSS, yang berpotensi mempengaruhi strategi keamanan siber di masa depan.

Kata kunci: Kerentanan XSS, CodeBERT, Transfer learning, Keamanan aplikasi web, Keamanan siber

CODEBERT WITH TRANSFER LEARNING FOR CROSS-SITE SCRIPTING VULNERABILITY DETECTION IN SOURCE CODE

M. RAFI SYAFRINALDI

21/472772/PA/20335

ABSTRACT

Cross-site scripting (XSS) remains a critical issue in web security, as attackers exploit vulnerabilities in web applications to execute malicious scripts, leading to severe security breaches. Traditional security methods often fall short in detecting sophisticated XSS attacks, resulting in high rates of false detections and overlooked vulnerabilities. To address these shortcomings, this thesis enhances CodeBERT, a pre-trained language model, by fine-tuning it with a custom dataset specifically designed to improve XSS vulnerability identification. This approach aims to reduce false detections significantly and strengthen web application security.

To address this critical issue, this thesis proposes an innovative approach employing CodeBERT, a pre-trained language model, augmented with transfer learning. The methodology involves collecting and preprocessing a diverse dataset of web application code, followed by fine-tuning CodeBERT's upper layers to specialize in XSS pattern recognition. This approach capitalizes on CodeBERT's advanced language understanding capabilities to enhance the detection accuracy of XSS attacks.

The evaluation of the fine-tuned CodeBERT model demonstrates exceptional performance, particularly on PHP datasets, where it significantly outperforms previous state-of-the-art approaches. The model achieves high accuracy, precision, recall, and F1-scores, showcasing its ability to minimize false negatives which is a crucial attribute in the context of security-related tasks. The expected contribution of this research is a substantial improvement in XSS detection, potentially influencing future cybersecurity strategies.

Key words: XSS vulnerabilities, CodeBERT, Transfer learning, Web application security, Cybersecurity