



**DETEKSI SERANGAN *CROSS SITE SCRIPTING* (XSS) PADA
PLUGIN WORDPRES MENGGUNAKAN *MACHINE
LEARNING***

SKRIPSI

**Karya tulis sebagai salah satu syarat
untuk memperoleh gelar Sarjana dari
Universitas Indo Global Mandiri**

**Oleh
MIKKO
NPM: 2022310070
(Program Studi Sarjana Sistem Komputer)**

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS INDO GLOBAL MANDIRI
2026**

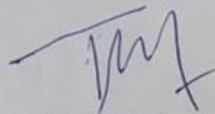
**Deteksi Serangan Cross Site Scripting (Xss) pada Plugin Wordpress
Menggunakan Machine Learning**

HALAMAN PENGESAHAN

Oleh
Mikko
NPM: 2022310070
(Program Studi Sarjana Sistem Komputer)
Universitas Indo Global Mandiri

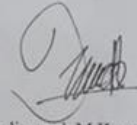
Menyetujui
Tim Pembimbing
Tanggal, 31 Juli 2026

Pembimbing 1



Dr. Tasmi, S.Si., M.Kom.
NIK. 2017.01.02.30

Pembimbing 2



Ferdiansyah M.Kom., Ph.D., CEH
NIK. 2024.01.03.75

Mengetahui,
Dekan Fakultas Ilmu Komputer & Sains
FAKULTAS ILMU KOMPUTER & SAINS

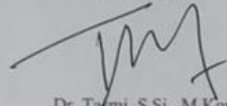


Dr. Herri Setiawan M.Kom.
NIK. 2003.00.01.67

LEMBAR PERSETUJUAN DEWAN PENGUJI

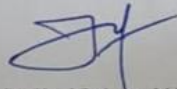
Pada hari ini Senin Tanggal 10 Agustus 2026 telah dilaksanakan Ujian Skripsi oleh Program Studi Sistem Komputer Fakultas Ilmu Komputer & Sains Universitas Indo Global Mandiri Palembang.

Menyetujui
Tim Penguji
Palembang, 14 Agustus 2026
Ketua Penguji



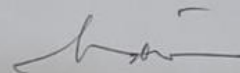
Dr. Tasmi, S.Si., M.Kom.
NIK. 2017.01.02.30

Penguji 1



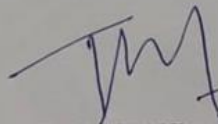
Dr. Herri Setiawan M.Kom.
NIK. 2003.00.01.67

Penguji 2



Ir. Hastha Sunardi, M.T.
NIK. 2005.01.00.72

Mengetahui
Ketua Program Studi Sistem Komputer



Tasmi, S.Si., M.Kom
NIK. 2017.01.02.30

SURAT KETERANGAN REVISI SKRIPSI

Kami yang bertanda tangan dibawah ini, menerangkan bahwa:

Nama : Mikko

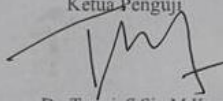
NPM : 2022310070

Judul Skripsi : Deteksi Serangan Cross Site Scripting (Xss) pada Plugin Wordpress
Menggunakan Machine Learning

Mahasiswa yang namanya tercantum diatas, telah selesai merevisi penulisan skripsi.

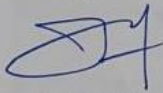
Menyetujui
Tim Penguji
Palembang, 14 Agustus 2026

Ketua Penguji



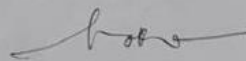
Dr. Tasmi, S.Si., M.Kom.
NIK. 2017.01.02.30

Penguji 1



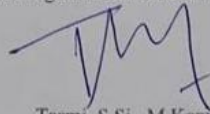
Dr. Herri Setiawan M.Kom.
NIK. 2003.00.01.67

Penguji 2



Ir. Hastha Sunardi, M.T.
NIK. 2005.01.00.72

Mengetahui
Ketua Program Studi Sistem Komputer



Tasmi, S.Si., M.Kom
NIK. 2017.01.02.30

ABSTRAK

DETEKSI SERANGAN *CROSS SITE SCRIPTING* (XSS) PADA PLUGIN WORDPRES MENGGUNAKAN *MACHINE LEARNING*

WordPress merupakan salah satu *Content Management System* (CMS) yang paling banyak digunakan dalam pengembangan website, sehingga pemanfaatan plugin pihak ketiga menjadi hal yang umum untuk memperluas fungsionalitas sistem. Namun, plugin yang menerima dan memproses input pengguna sering kali menjadi celah keamanan, khususnya terhadap serangan *Cross-Site Scripting* (XSS). Serangan XSS memungkinkan penyerang menyisipkan skrip berbahaya yang dieksekusi pada sisi klien sehingga dapat menyebabkan pencurian data, pengambilalihan sesi pengguna, serta gangguan terhadap integritas sistem.

Penelitian ini bertujuan untuk mendeteksi serangan XSS pada plugin WordPress menggunakan pendekatan machine learning dengan algoritma *Random Forest* dan *Support Vector Machine* (SVM). Penelitian difokuskan pada tiga plugin WordPress yang merepresentasikan vektor input berbeda yaitu *Contact Form 7*, *Ultimate Member*, dan *wpDiscuz*. Dataset yang digunakan terdiri dari dataset primer yang diperoleh melalui perancangan testbed WordPress pada lingkungan localhost dengan merekam lalu lintas HTTP menggunakan Wireshark, serta dataset sekunder yang berasal dari *repositori* publik Kaggle. Data yang dikumpulkan kemudian melalui tahapan pra-pemrosesan meliputi pembersihan data dan normalisasi *case folding*, serta ekstraksi lima fitur *numerik* dari payload yaitu *payload_length*, *entropy*, *special_char_ratio*, *alpha_ratio*, dan *avg_word_length*. Optimasi *hyperparameter* dilakukan menggunakan *Grid Search* dengan *cross-validation* 5 lipatan untuk memperoleh parameter terbaik pada masing-masing algoritma.

Model *Random Forest* dan SVM dilatih dan diuji secara terpisah pada dataset primer yang berjumlah 891 data dan dataset sekunder yang berjumlah 10.913 data untuk mengklasifikasikan payload ke dalam kelas XSS dan Normal. Evaluasi kinerja model dilakukan menggunakan *confusion matrix* dengan metrik *accuracy*, *precision*, *recall*, *F1-score*, *balanced accuracy*, *macro average*, AUC-PR, dan ROC-AUC. Hasil penelitian menunjukkan bahwa algoritma *Random Forest* secara konsisten mengungguli algoritma SVM pada seluruh metrik evaluasi di kedua dataset. Pada dataset primer, Random Forest mencapai *accuracy* sebesar 95.53% dan ROC-AUC sebesar 0.9810, sedangkan pada dataset sekunder mencapai *accuracy* sebesar 98.26% dan ROC-AUC sebesar 0.9984. Hasil tersebut membuktikan bahwa pendekatan *machine learning* khususnya algoritma *Random Forest* efektif dalam mendeteksi serangan XSS pada plugin WordPress.

Kata kunci: *Cross-Site Scripting*, Plugin WordPress, *Machine Learning*, *Random Forest*, *Support Vector Machine*, Deteksi Serangan.

ABSTRACT

DETECTION OF CROSS-SITE SCRIPTING (XSS) ATTACKS ON WORDPRESS PLUGINS USING MACHINE LEARNING

WordPress is one of the most widely used Content Management Systems (CMS) in website development, making the use of third-party plugins a common practice to extend system functionality. However, plugins that receive and process user input often become security vulnerabilities, particularly against Cross-Site Scripting (XSS) attacks. XSS attacks allow attackers to inject malicious scripts that are executed on the client side, potentially leading to data theft, user session hijacking, and disruption of system integrity.

This study aims to detect XSS attacks on WordPress plugins using a machine learning approach with Random Forest and Support Vector Machine (SVM) algorithms. The research focuses on three WordPress plugins representing different input vectors, namely Contact Form 7, Ultimate Member, and wpDiscuz. The dataset used consists of a primary dataset obtained through a WordPress testbed designed in a localhost environment by recording HTTP traffic using Wireshark, and a secondary dataset obtained from the Kaggle public repository. The collected data underwent preprocessing stages including data cleaning and case folding normalization, as well as extraction of five numerical features from payloads namely `payload_length`, `entropy`, `special_char_ratio`, `alpha_ratio`, and `avg_word_length`. Hyperparameter optimization was performed using Grid Search with 5-fold cross-validation to obtain the best parameters for each algorithm.

The Random Forest and SVM models were trained and tested separately on the primary dataset of 891 records and the secondary dataset of 10,913 records to classify payloads into XSS and Normal classes. Model performance evaluation was conducted using a confusion matrix with metrics including accuracy, precision, recall, F1-score, balanced accuracy, macro average, AUC-PR, and ROC-AUC. The results showed that the Random Forest algorithm consistently outperformed the SVM algorithm across all evaluation metrics on both datasets. On the primary dataset, Random Forest achieved an accuracy of 95.53% and ROC-AUC of 0.9810, while on the secondary dataset it achieved an accuracy of 98.26% and ROC-AUC of 0.9984. These results demonstrate that the machine learning approach, particularly the Random Forest algorithm, is effective in detecting XSS attacks on WordPress plugins.

Keywords: Cross-Site Scripting, WordPress Plugin, Machine Learning, Random Forest, Support Vector Machine, Attack Detection, Balanced Accuracy

KATA PENGANTAR

Segala puji dan syukur senantiasa penulis panjatkan kepada Allah SWT dan tidak hentinya sholawat serta salam penulis limpah curahkan kepada baginda besar Nabi Muhammad SAW. Sehingga penulis dapat menyelesaikan Skripsi ini dengan judul “DETEKSI SERANGAN *CROSS SITE SCRIPTING* (XSS) PADA PLUGIN WORDPRES MENGGUNAKAN *MACHINE LEARNING*” sebagai salah satu syarat dalam menyelesaikan pendidikan Sarjana Program Studi Sistem Komputer di Universitas Indo Global Mandiri.

Pada kesempatan ini, penulis ingin menyampaikan ucapan terima kasih yang tak terhingga kepada pihak-pihak yang telah memberikan dukungan, bimbingan, semangat dan kemauan kepada penulis untuk menyelesaikan skripsi ini, yaitu:

1. Bapak Dr. H. Marzuki Alie, S.E., M.M. Selaku Rektor Universitas Indo Global Mandiri Palembang
2. Bapak Dr. Herri Setiawan, M.Kom. Selaku Dekan Fakultas Ilmu Komputer dan Sains atas kebijakan dan dorongan yang di berikan.
3. Bapak Tasmi, S.Si., M.Kom. selaku Dosen Pembimbing I yang dengan sabar telah memberikan motivasi dan bimbingan kepada penulis untuk menyelesaikan skripsi ini.
4. Bapak Ferdiansyah., S.Kom., M.Kom., Ph.D. selaku Dosen Pembimbing II yang dengan sabar telah memberikan motivasi dan bimbingan kepada penulis untuk menyelesaikan skripsi ini.
5. Bapak dan Ibu Dosen yang selama ini telah memberikan ilmunya kepada penulis selama proses belajar mengajar di Fakultas Ilmu Komputer Universitas Indo Global Mandiri.
6. Kepada Kedua orang tua yang selalu memberikan ketulusan doanya dan dukungan baik secara moril maupun material, juga motivasi sehingga penulis dapat menyelesaikan skripsi ini.

7. Kepada seluruh keluarga besar penulis yang selalu memberikan semangat, doa, dan dukungan dalam setiap langkah perjuangan penulis selama menyelesaikan pendidikan hingga tahap akhir ini.
8. Teman-teman seperjuangan yang ada di fakultas ilmu komputer khususnya program studi sistem komputer.

Penulis berharap skripsi ini dapat membawa dampak positif bagi para pembaca, karena di dalam skripsi ini memuat pembelajaran yang penulis dapatkan selama penelitian berlangsung. Dalam hal ini penulis tidak menutup diri untuk menerima kritik dan saran yang sekiranya bisa menjadi pembelajaran bagi penulis untuk berkembang menjadi lebih baik lagi.

Palembang, 9 Juli 2026

Penulis

Mikko

NPM.2022310070

DAFTAR ISI

SURAT HALAMAN PENGESAHAN	i
SURAT PERSETUJUAN DEWAN PENGUJ.....	ii
SURAT KETERANGAN REVISI SKRIPSI.....	iii
ABSTRAK	iv
ABSTRACT	v
KATA PENGANTAR.....	vi
DAFTAR ISI.....	viii
DAFTAR GAMBAR.....	xi
DAFTAR TABEL	xiii
DAFTAR SINGKATAN DAN LAMBANG	xiv
DAFTAR LAMPIRAN	xv
BAB I.....	1
PENDAHULUAN.....	1
I.1 Latar Belakang	1
I.2 Rumusan Masalah	3
I.3 Batasan Masalah.....	3
I.4 Tujuan Penelitian	3
I.5 Manfaat Penelitian	4
I.6 Metodologi Penelitian	4
I.7 Sistematika Penulisan	5
BAB II	6
TINJAUAN PUSTAKA	6
II.1 <i>Wordpress</i>	6
II.2 Plugin	6
II.2.1 <i>Contact Form 7</i>	7
II.2.2 <i>Ultimate Member</i>	8
II.2.3 <i>WpDiscuz</i>	8
II.3 <i>Cross-Site Scripting (XSS)</i>	9

II.3.1	Tiga jenis utama serangan XSS:.....	10
II.4	<i>Machine learning</i> (ML).....	10
II.4.1	<i>Random Forest</i>	11
II.4.2	<i>Support Vector Machine</i> (SVM)	12
II.5	<i>Confusion matrix</i>	13
II.5.1	<i>Accuracy</i>	14
II.5.2	<i>Precision</i>	14
II.5.3	<i>Recall</i>	14
II.5.4	<i>F1-score</i>	14
II.5.5	<i>Area Under Precision –Recall Curve</i> (AUC-PR)	15
II.5.6	<i>Receiver Operating Characteristic - Area Under Precision</i> (ROC)	15
II.5.7	<i>Balanced accuracy</i> (akurasi Seimbang).....	16
II.5.8	<i>Macro-Average</i> (Rata-rata Makro).....	17
II.6	Penelitian Terdahulu	17
BAB III		20
METODE PENELITIAN		20
III.1	Kerangka Penelitian	20
III.2	Studi Literatur.....	21
III.3	Analisis Kebutuhan Sistem.....	21
III.3.1	Analisis Kebutuhan Perangkat Lunak (<i>Software</i>).....	21
III.3.2	Analisis Kebutuhan Perangkat Keras (<i>Hardware</i>).....	21
III.4	Identifikasi Masalah	22
III.5	Pengumpulan Data	23
III.6	Perancangan Testbed.....	23
III.7	Proses dan Pengujian Model <i>Machine Learning</i>	24
III.7.1	Pra-pemrosesan Data.....	25
III.7.2	Ekstraksi Fitur	27
III.7.3	Pemisahan Data	27
III.7.4	Optimasi Model.....	27
III.7.5	Evaluasi Model.....	28

BAB IV	29
HASIL DAN PEMBAHASAN	29
IV.1 Hasil Pengumpulan Dataset	29
IV.2 Hasil Pra-Pemrosesan Dataset Primer.....	30
IV.2.1 Hasil Ekstraksi Fitur	30
IV.2.2 Proses Parser	32
IV.2.3 Proses Menentukan Label	33
IV.2.4 Pembersihan Dataset	34
IV.2.5 Hasil Pemisahan Dataset	37
IV.3 Hasil Pra-Pemrosesan Dataset Sekunder	39
IV.3.1 Pembersihan Dataset	39
IV.3.2 Hasil Pemisahan Dataset	43
IV.4 Pencarian Nilai Hiperparameter Dengan <i>Grid Search</i>	44
IV.5 Hasil Pengujian Performa Algoritma <i>Machine learning</i>	47
IV.5.1 <i>Confusion Matrix</i> dan Hasil Klasifikasi Algoritma <i>Random Forest</i>	48
IV.5.2 <i>Confusion Matrix</i> dan Hasil Klasifikasi Algoritma <i>Support Vector Machine</i>	50
IV.5.3 <i>Area Under Precision – Recall Curve (AUC-PR)</i>	52
IV.5.4 <i>Receiver Operating Characteristic - Area Under the Curve (ROC-AUC)</i>	53
IV.5.5 <i>Balanced accuracy</i>	53
IV.6 Analisis Hasil Pengujian	56
BAB V.....	59
KESIMPULAN DAN SARAN	59
V.1 Kesimpulan	59
V.2 Saran.....	60
DAFTAR PUSTAKA	62

DAFTAR GAMBAR

Gambar 2.1 Logo Wordpress	6
Gambar 2.2 Skema Penyerangan <i>Cross-Site Scripting</i> (XSS)	9
Gambar 2.3 Konsep <i>Random Forest</i>	12
Gambar 2.4 <i>Confusion Matrix</i> Serta Metrik Terkait.	13
Gambar 3.1 Diagram Alir Kerangka Penelitian.	20
Gambar 3.2 Perancangan Testbed.	24
Gambar 3.3 Proses dan Pengujian Model <i>Machine Learning</i>	25
Gambar 3.4 Diagram Alir Pra-Pemrosesan Dataset.	26
Gambar 4.1 Tampilan Data Awal Dataset Primer.....	30
Gambar 4.2 Tampilan Data Awal Dataset Sekunder.....	30
Gambar 4.3 Hasil <i>Capture Traffic</i> Wireshark.	31
Gambar 4.4 Hasil Ekstraksi Fitur.	32
Gambar 4.5 Proses Parser dari File .Pcap dan Ekstraksi Fitur.	32
Gambar 4.6 Kode Program Proses Menentukan XSS = 1 dan Normal = 0.....	33
Gambar 4.7 Penghapusan <i>Missing Value</i>	34
Gambar 4.8 Penghapusan Data Salah Kategori.....	35
Gambar 4.9 Penghapusan Data Duplikat.....	35
Gambar 4.10 Normalisasi <i>Case Folding</i>	36
Gambar 4.11 Penghapusan Duplikat setelah <i>Case Folding</i>	36
Gambar 4.12 Hasil Akhir Pembersihan Dataset Primer.	37
Gambar 4.13 Grafik Hasil Proses Pemisahan Daset Primer.....	39
Gambar 4.14 Penghapusan Kolom Tidak Relevan.....	40
Gambar 4.15 Penyeragaman Nama Kolom.	40
Gambar 4.16 Penghapusan <i>Missing Value</i>	40
Gambar 4.17 Penghapusan Payload Kosong.....	41
Gambar 4.18 Penghapusan Data Duplikat.....	41
Gambar 4.19 Normalisasi <i>Case Folding</i>	42
Gambar 4.20 Penghapusan Duplikat Pasca <i>Case Folding</i>	42
Gambar 4.21 Ekstraksi Fitur Dataset Sekunder.....	43

Gambar 4.22 Hasil Akhir Pembersihan Dataset Sekunder.....	43
Gambar 4.23 Hasil Pencarian Nilai Hiperparameter Dengan Pendekatan <i>Grid Search</i> Dataset Primer (a) <i>Random Forest</i> , (b) <i>Support Vector Machine</i>	46
Gambar 4.24 Hasil Pencarian Nilai Hiperparameter Dengan Pendekatan <i>Grid Search</i> Dataset Sekunder (a) <i>Random Forest</i> , (b) <i>Support Vector Machine</i>	47
Gambar 4.25 Lokasi Nilai TP, FP, TN, dan FN Pada <i>Confusion Matrix</i> (a) Label Normal, (b) Label XSS.	48
Gambar 4.26 <i>Confusion Matrix</i> Pada Algoritma <i>Random Forest</i> (a) Dataset Primer, (b) Dataset Sekunder.	49
Gambar 4.27 <i>Confusion Matrix</i> Pada Algoritma <i>Support Vector Machine</i> (a) Dataset Primer, (b) Dataset Sekunder.	51
Gambar 4.28 Grafik <i>Area Under Precision – Recall Curve</i> (AUC-PR).	52
Gambar 4.29 Grafik <i>Receiver Operating Characteristic - Area Under the Curve</i> (ROC-AUC).	53
Gambar 4.30 Grafik Perbandingan Teknik Resampling Dataset Primer.....	55
Gambar 4.31 Grafik Perbandingan Teknik Resampling Dataset Sekunder.	56

DAFTAR TABEL

Tabel 2.1 Keunggulan <i>Random Forest</i> dalam Proses Deteksi XSS.....	11
Tabel 2.2 Keunggulan <i>Support Vector Machine</i> dalam Proses Deteksi XSS.	12
Tabel 2.3 Penelitian Terdahulu.....	17
Tabel 3.1 Daftar Perangkat Lunak (<i>Software</i>).....	21
Tabel 3.2 Daftar Perangkat Keras (<i>Hardware</i>).	22
Tabel 4.1 Ringkasan Dataset Primer dan Sekunder.	29
Tabel 4.2 Fitur yang di gunakan dalam penelitian.	31
Tabel 4.3 Perbandingan Rasio Pembagian Dataset Primer.	38
Tabel 4.4 Perbandingan Rasio Pembagian Dataset Sekunder.	44
Tabel 4.5 Hasil Nilai <i>Convusion Matrix</i> Pada <i>Random Forest</i>	49
Tabel 4.6 Laporan Pengklasifikasian Pada Algoritma <i>Random Forest</i>	49
Tabel 4.7 Hasil <i>Macro Averange</i> Pada Algoritma <i>Random Forest</i>	50
Tabel 4. 8 Hasil Nilai <i>Convusion Matrix</i> Pada <i>Support Vector Machine</i>	51
Tabel 4.9 Laporan Pengklasifikasian Pada Algoritma <i>Support Vector Machine</i>	51
Tabel 4.10 Hasil <i>Macro Averange</i> Pada Algoritma <i>Support Vector Machine</i>	52
Tabel 4.11 Hasil <i>Balanced accuracy</i> Algoritma <i>Random Forest</i> dan <i>Support Vector Machine</i> pada Dataset Primer dan Dataset Sekunder.....	54

DAFTAR SINGKATAN DAN LAMBANG

Singkatan	Nama	Pemakaian Pertama kali pada Halaman
AJAX	<i>Asynchronous JavaScript and XML</i>	8
AUC	<i>Area Under the Curve</i>	3
AUC-PR	<i>Area Under Precision-Recall Curve</i>	3
CMS	<i>Content Management System</i>	1
CSV	<i>Comma-Separated Values</i>	23
CVE	<i>Common Vulnerabilities and Exposures</i>	7
DOM	<i>Document Object Model</i>	10
FN	<i>False Negative</i>	13
FP	<i>False Positive</i>	13
FPR	<i>False Positive Rate</i>	2
HTTP	<i>Hypertext Transfer Protocol</i>	23
ML	<i>Machine Learning</i>	2
ROC	<i>Receiver Operating Characteristic</i>	15
ROC-AUC	<i>Receiver Operating Characteristic - Area Under the Curve</i>	3
ROS	<i>Random Over Sampling</i>	16
RUS	<i>Random Under Sampling</i>	16
SMOTE	<i>Synthetic Minority Oversampling Technique</i>	16
SVM	<i>Support Vector Machine</i>	2
TN	<i>True Negative</i>	13
TP	<i>True Positive</i>	13
XSS	<i>Cross-Site Scripting</i>	1

DAFTAR LAMPIRAN

Lampiran 1 Daftar Riwayat Hidup	66
Lampiran 2 Kartu Bimbingan.....	67
Lampiran 3 Surat Pernyataan Tidak Plagiat.....	68
Lampiran 4 Persetujuan Ujian Skripsi	69
Lampiran 5 Rekomendasi Sidang Sidang	70
Lampiran 6 Surat Keterangan Siap Sidang	71
Lampiran 7 Lembar Revisi Ujian Sidang Skripsi	72
Lampiran 8 Hasil Cek Plagiasi	73