KLASIFIKASI SERANGAN *SQL INJECTION* MENGGUNAKAN ALGORITMA *SUPPORT VECTOR MACHINE* PADA *HTTP REQUEST*

FAHRI RAMADHAN
H1051181048

SKRIPSI

PROGRAM STUDI REKAYASA SISTEM KOMPUTER
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS TANJUNGPURA
PONTIANAK
2024

KLASIFIKASI SERANGAN *SQL INJECTION* MENGGUNAKAN ALGORITMA *SUPPORT VECTOR MACHINE* PADA *HTTP REQUEST*

FAHRI RAMADHAN
H1051181048

Skripsi
Sebagai Salah Satu Syarat untuk Memperoleh gelar
Sarjana Komputer pada Program Studi Rekayasa Sistem Komputer

JURUSAN REKAYASA SISTEM KOMPUTER
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM UNIVERSITAS TANJUNGPURA
PONTIANAK
2024

LEMBAR PENGESAHAN

Judul Tugas Akhir	:	Klasifikasi Serangan *SQL Injection* Menggunakan Algoritma *Support Vector Machine* Pada *HTTP Request*
Nama Mahasiswa	:	Fahri Ramadhan
NIM	:	H1051181048
Jurusan/Program Studi	:	Rekayasa Sistem Komputer
Tanggal Lulus	:	12 November 2024
SK Pembimbing	:	No. 663/UN22.8/TD.06/2023/Tanggal 14 Februari 2023
SK Penguji	:	No. 7048/UN22.8/PK.03.05/2024/Tanggal 12 November 2024

Dosen Pembimbing

Pembimbing I	Pembimbing II
Ikhwan Ruslianto, S.Kom., M.Cs.	Syamsul Bahri., S.Kom., M.Cs.
NIP. 198607012014041001	NIP. 198802272015041001

Dosen Penguji

Ketua Penguji	Anggota Penguji
Uray Ristian, S.Kom., M.Kom.	Hirzen Hasfani., S.Kom., M.Cs.
NIP. 199012012019031017	NIP. 199305032022031003

Pimpinan Sidang (merangkap anggota penguji)	Sekretaris Sidang (merangkap anggota penguji)
Ikhwan Ruslianto, S.Kom., M.Cs.	Syamsul Bahri., S.Kom., M.Cs.
NIP. 198607012014041001	NIP. 198802272015041001

Mengesahkan
Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Tanjungpura

Prof. Dr. Gusrizal, S.Si., M.Si.
NIP. 197108022000031001

PERNYATAAN INTEGRITAS AKADEMIK

Saya yang bertanda tangan di bawah ini:

Nama	:	Fahri Ramadhan
NIM	:	H1051181048
Program Studi	:	Rekayasa Sistem Komputer
Fakultas	:	Matematika dan Ilmu Pengetahuan Alam

Dengan ini menyatakan bahwa dokumen ilmiah Tugas Akhir yang disajikan ini tidak mengandung unsur pelanggaran integritas akademik sesuai Peraturan Menteri Pendidikan, Kebudayaan, Riset, Dan Teknologi Republik Indonesia Nomor 39 Tahun 2021. Apabila di kemudian hari dokumen ilmiah Tugas Akhir ini mengandung unsur pelanggaran integritas akademik sesuai ketentuan perundangan tersebut, maka saya bersedia menerima sanksi akademik dan/atau sanksi hukum yang berlaku.

Demikian pernyataan ini saya buat, semoga dapat dipergunakan sebagaimana mestinya.

Pontianak, 12 November 2024

Fahri Ramadhan
NIM. H1051181048

KLASIFIKASI SERANGAN *SQL INJECTION* MENGGUNAKAN ALGORITMA *SUPPORT VECTOR MACHINE* PADA *HTTP REQUEST*

ABSTRAK

Banyak *website* menyediakan kemudahan akses informasi namun rentan terhadap serangan *SQL injection* yang bertujuan mencuri data sensitif. Penelitian ini membangun sistem klasifikasi serangan *SQL injection* melalui *HTTP request* menggunakan algoritma *support vector machine*. Fokus utama adalah mengatasi permasalahan klasifikasi *multiclass* dengan teknik *one against one* atau *One-vs-One* (OVO) pada *SVM* yang memungkinkan klasifikasi serangan ke dalam lebih dari dua kategori atau kelas. Sistem diintegrasikan dengan metode pemrosesan teks seperti *n-gram* dan *TF-IDF* untuk mengidentifikasi pola serangan berdasarkan karakteristik sintaksis dan semantik dari *HTTP request*. Fitur klasifikasi yang digunakan adalah potongan karakter dari proses *n-gram*, dengan data yang diklasifikasikan bersumber dari *body* dan *query string HTTP request*. Pengujian terhadap 557 data serangan dengan 262 *dataset* menghasilkan 522 data dengan klasifikasi benar dan 35 data dengan klasifikasi salah. Nilai *confusion matrix* menunjukkan hasil 522 *true negative*, 35 *false negative*, 0 *true positive*, dan 0 *false positive*, dengan akurasi sistem sebesar 93,7%. Dapat disimpulkan bahwa algoritma *support vector machine* sangat efektif dalam mempelajari pola dan melakukan klasifikasi serangan *SQL injection*.

Kata kunci: *SQL Injection,* Klasifikasi, *Support Vector Machine*, *N-Gram, TF-IDF*

CLASSIFICATION OF SQL INJECTION ATTACKS USING SUPPORT VECTOR MACHINE ALGORITHM ON HTTP REQUESTS

ABSTRACT

Many websites provide easy access to information but are vulnerable to SQL injection attacks aimed at stealing sensitive data. This research builds a classification system for SQL injection attacks through HTTP requests using the support vector machine algorithm. The main focus is addressing multiclass classification problems with the One-vs-One (OVO) technique in SVM, which enables the classification of attacks into more than two categories or classes. The system is integrated with text processing methods such as n-gram and TF-IDF to identify attack patterns based on the syntactic and semantic characteristics of HTTP requests. The classification features used are character segments from the n-gram process, with data being classified sourced from the body and query string of HTTP requests. Testing on 557 attack data with 262 datasets resulted in 522 correctly classified data and 35 incorrectly classified data. The confusion matrix values show results of 522 true negatives, 35 false negatives, 0 true positives, and 0 false positives, with a system accuracy of 93.7%. It can be concluded that the support vector machine algorithm is highly effective in learning patterns and performing classification of SQL injection attacks.

Keywords**:** *SQL Injection, Classification*, *Support Vector Machine*, *N-Gram*, *TF-IDF*

PRAKATA

Puji dan syukur penulis panjatkan kepada Allah SWT atas segala berkah dan nikmat yang telah diberikan-Nya. *Shalawat* serta salam semoga tercurah kepada Nabi Muhammad SAW beserta keluarga dan umatnya. Alhamdulillah penulis dapat menyelesaikan tugas akhir yang berjudul "**Klasifikasi Serangan *SQL Injection* Menggunakan Algoritma *Support Vector Machine* Pada *HTTP Request***" sebagai syarat bagi penulis untuk mendapatkan gelar Sarjana Komputer (S.Kom.) pada Jurusan Rekayasa Sistem Komputer Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Tanjungpura Pontianak.

Semua hasil yang diperoleh ini tidak terlepas dari segala dukungan dan bantuan dari berbagai pihak yang tidak pernah berhenti mendorong penulis agar dapat menyelesaikan tugas akhir ini dengan baik. Oleh karena itu pada kesempatan ini izinkanlah penulis mengucapkan rasa terima kasih yang sebesar-besarnya kepada:

1. Bapak Prof. Dr. Gusrizal, S.Si., M.Si., selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Tanjungpura.
2. Bapak Tedy Rismawan, S.Kom., M.Cs., selaku Ketua Jurusan Rekayasa Sistem Komputer Universitas Tanjungpura.
3. Bapak Ikhwan Ruslianto, S.Kom., M.Cs., selaku Dosen Pembimbing Pertama yang telah memberikan banyak motivasi dan saran dalam menyelesaikan tugas akhir.
4. Bapak Syamsul Bahri, S.Kom., M.Cs., selaku Dosen Pembimbing Kedua yang telah memberikan banyak motivasi dan saran dalam menyelesaikan tugas akhir.
5. Bapak Uray Ristian, S.Kom., M.Kom., selaku Dosen Penguji Pertama yang telah memberikan banyak kritik dan saran untuk melengkapi tugas akhir.
6. Bapak Hirzen Hasfani, S.Kom., M.Cs., selaku Dosen Penguji Kedua yang telah memberikan banyak kritik dan saran untuk melengkapi tugas akhir.
7. Ibu Irma Nirmala, S.T., M.T., sebagai Dosen Pembimbing Akademik yang tidak pernah lupa memberikan nasihat dan bimbingan terkait perkuliahan setiap semester yang dilalui.

8. Bapak dan Ibu Dosen serta Staf Administrasi Jurusan/Program Studi Rekayasa Sistem Komputer Universitas Tanjungpura.
9. Seluruh rekan kerja UPA TIK Universitas Tanjungpura yang telah banyak memberikan semangat untuk menyelesaikan tugas akhir.
10. Kedua orang tua (Bapak Muriani dan Ibu Jumarni), kakak (Nadya Angraini) yang telah banyak memberikan semangat, motivasi, bantuan, serta doa.
11. Teman-teman seperjuangan mahasiswa Rekayasa Sistem Komputer terutama angkatan 2018 "INTEL CORE'18" yang telah membantu memberikan informasi, masukan, dan saran.
12. Semua pihak yang terlibat langsung maupun yang tidak langsung dalam penulisan skripsi ini.

Penulis menyadari bahwa tugas akhir ini masih memiliki banyak kekurangan. Oleh karena itu, penulis mengharapkan kritik dan saran yang bersifat membangun. Semoga tugas akhir ini dapat memberikan manfaat bagi kemajuan ilmu pengetahuan khususnya dalam bidang teknologi informasi terutama *cyber security*.

Pontianak, 12 November 2024

Fahri Ramadhan
NIM. H1051181048

DAFTAR ISI

LEMBAR PENGESAHAN ii
PERNYATAAN INTEGRITAS AKADEMIK iii
ABSTRAK iv
ABSTRACT v
PRAKATA vi
DAFTAR ISI viii
DAFTAR GAMBAR xii
DAFTAR TABEL xiv
DAFTAR KODE PROGRAM xvi
BAB 1 PENDAHULUAN 1
1.1 Latar Belakang 1
1.2 Rumusan Masalah 3
1.3 Batasan Masalah 4
1.4 Tujuan Penelitian 4
1.5 Manfaat Penelitian 5
1.6 Sistematika Penulisan 5
BAB 2 DASAR TEORI 7
2.1 Perbandingan Dengan Penelitian Dahulu 7
2.2 Pengertian Klasifikasi 10
2.3 *SQL Injection* 10
2.3.1 *Error-Based SQL Injection* 12
2.3.2 *Time-Based Blind SQL Injection* 12
2.3.3 *Union Query SQL Injection* 13
2.4 *N-Gram* 13
2.5 *Term Frequency-Inverse Document Frequency* (TF-IDF) 13
2.6 *Support Vector Machine* 15
2.6.1 *Radial Basis Function* 18
2.6.2 *Multiclass* 19
2.7 *Cost Function* 19
2.8 *Gradient Descent* 20

2.9 *Confusion Matrix*.....23
2.10 *K-Fold Cross Validation*.....24
2.11 *Flowchart*.....24
2.12 *Burp Suite*.....27
2.13 *SQLMap*.....28
2.14 *Proxy Server*.....28
BAB 3 METODE PENELITIAN.....30
3.1 Studi Literatur.....31
3.2 Pengumpulan Data.....31
3.3 Analisis Kebutuhan.....31
3.3.1 Kebutuhan Perangkat Keras.....31
3.3.2 Kebutuhan Perangkat Lunak.....32
3.4 Perancangan Sistem.....32
3.5 Implementasi.....32
3.6 Pengujian Sistem.....33
3.7 Kesimpulan dan Saran.....33
BAB 4 PERANCANGAN.....34
4.1 Deskripsi Sistem.....34
4.2 Alur Proses Pada Sistem.....35
4.2.1 Proses *Training Dataset*.....35
4.2.2 Proses *Scraping*.....36
4.2.3 Proses *Injection*.....37
4.2.4 Proses Klasifikasi.....37
4.2.5 Proses Pengujian.....38
4.3 Rancangan Diagram Alir (*Flowchart*) Sistem.....38
4.3.1 Diagram Alir *Text Preprocessing*.....38
4.3.2 Diagram Alir *Text Vectorization*.....39
4.3.3 Diagram Alir *Radial Basis Function Kernel*.....40
4.3.4 Diagram Alir *Training Dataset* Menggunakan *Support Vector Machine*.....41
4.3.5 Diagram Alir Klasifikasi Menggunakan *Support Vector Machine*.....43

4.4 Rancangan Basis Data (*Database*) 43
4.4.1 Tabel *Dataset* 44
4.4.2 Tabel Label 44
4.4.3 Tabel *Category* 44
4.4.4 Tabel Hasil *Training Dataset* 45
4.4.5 Tabel *Web Host* 46
4.4.6 Tabel *Web Host HTTP Traffic* 47
4.4.7 Tabel *HTTP Traffic Data* 48
4.4.8 Tabel *Malicious Keyword* 48
4.5 Rancangan *REST API Backend Website* 49
4.6 Rancangan *Proxy Server* 50
4.7 Rancangan Antarmuka (*Interface*) *Frontend Website* 50
4.7.1 Rancangan Antarmuka Halaman Utama 50
4.7.2 Rancangan Antarmuka Halaman Daftar *Web Host* 51
4.7.3 Rancangan Antarmuka Halaman Daftar Serangan *SQL Injection* 52
4.7.4 Rancangan Antarmuka Halaman Daftar Kata Kunci Dan Karakter Berbahaya 53
4.7.5 Rancangan Antarmuka Halaman Daftar Kelas Klasifikasi 53
4.7.6 Rancangan Antarmuka Halaman Daftar Kategori Kelas Klasifikasi 54
4.7.7 Rancangan Antarmuka Halaman Daftar *Dataset* 55
4.7.8 Rancangan Antarmuka Halaman *Training Dataset* 56
4.7.9 Rancangan Antarmuka Halaman Klasifikasi Manual 57
4.8 Rancangan Pengujian Sistem 58
BAB 5 IMPLEMENTASI DAN PEMBAHASAN 59
5.1 Implementasi Sistem 59
5.1.1 Implementasi *REST API Backend Website* 59
5.1.2 Implementasi *Proxy Server* 59
5.1.3 Implementasi Antarmuka *Frontend Website* 60
5.1.4 Kode Program 66
5.2 Pengujian Sistem 72

5.2.1 Pengujian Dan Perhitungan *Training Dataset* Menggunakan Algoritma *Support Vector Machine* 73
5.2.2 Pengujian *Proxy Server*, *Backend Website*, Dan *Frontend Website* Untuk Menangkap Serangan *SQL Injection* Melalui *HTTP Request* 80
5.2.3 Pengujian Dan Perhitungan Akurasi Klasifikasi Serangan *SQL Injection* 83
5.3 Pembahasan Sistem 90
BAB 6 PENUTUP 96
6.1 Kesimpulan 96
6.2 Saran 97
DAFTAR PUSTAKA 98
LAMPIRAN 100

DAFTAR GAMBAR

Gambar 2.1 Grafik *Support Vector* Dua Dimensi 16
Gambar 2.2 *K-Fold Cross Validation* 24
Gambar 4.1 Deskripsi Sistem 35
Gambar 4.2 Alur Proses *Training Dataset* 36
Gambar 4.3 Alur Proses *Scraping* 36
Gambar 4.4 Alur Proses *Injection* 37
Gambar 4.5 Diagram Alir *Text Preprocessing* 39
Gambar 4.6 Diagram Alir *Text Vectorization* 40
Gambar 4.7 Diagram Alir *Radial Basis Function Kernel* 41
Gambar 4.8 Diagram Alir *Training Dataset Support Vector Machine* 42
Gambar 4.9 Diagram Alir Klasifikasi Menggunakan *Support Vector Machine* ... 43
Gambar 4.10 Rancangan *Proxy Server* 50
Gambar 4.11 Rancangan Antarmuka Halaman Utama 51
Gambar 4.12 Rancangan Antarmuka Halaman Daftar *Web Host* 51
Gambar 4.13 Rancangan Antarmuka Halaman Daftar Serangan *SQL Injection* .. 52
Gambar 4.14 Antarmuka Halaman Daftar Kata Kunci Dan Karakter Berbahaya 53
Gambar 4.15 Rancangan Antarmuka Halaman Daftar Kelas Klasifikasi 54
Gambar 4.16 Rancangan Antarmuka Halaman Daftar Kategori Kelas Klasifikasi 55
Gambar 4.17 Rancangan Antarmuka Halaman Daftar *Dataset* 56
Gambar 4.18 Rancangan Antarmuka Halaman *Training Dataset* 57
Gambar 4.19 Rancangan Antarmuka Halaman Klasifikasi Manual 57
Gambar 4.20 Rancangan Pengujian Sistem 58
Gambar 5.1 Dokumentasi *REST API Backend Website* 59
Gambar 5.2 Daftar *Port* yang Berjalan Pada Sistem Operasi 60
Gambar 5.3 Tampilan Antarmuka Halaman Utama 61
Gambar 5.4 Tampilan Antarmuka Halaman Daftar *Web Host* 61
Gambar 5.5 Tampilan Antarmuka Halaman Daftar Serangan *SQL Injection* 62
Gambar 5.6 Tampilan Antarmuka Halaman Daftar Kata Kunci Dan Karakter Berbahaya 63

Gambar 5.7 Tampilan Antarmuka Halaman Daftar *Dataset Label* 63
Gambar 5.8 Tampilan Antarmuka Halaman Daftar *Label Category* 64
Gambar 5.9 Tampilan Antarmuka Halaman Daftar *Dataset* 64
Gambar 5.10 Tampilan Antarmuka Halaman *Training Dataset* 65
Gambar 5.11 Tampilan Antarmuka Halaman Klasifikasi Teks Manual 65
Gambar 5.12 *Website* Sasaran Pengujian Serangan *SQL Injection* 80
Gambar 5.13 *Scraping Website* Sasaran Menggunakan *Burp Suite* 81
Gambar 5.14 Hasil *Raw Text HTTP Request* Dari *Burp Suite* 81
Gambar 5.15 Hasil Serangan *SQL Injection* Yang Direkam 82
Gambar 5.16 Grafik Pengaruh *Epoch* Terhadap *Weight Vector* Pada *Radial Basis Function Kernel* 90
Gambar 5.17 Grafik Pengaruh *Learning Rate* Terhadap *Weight Vector* Pada *Radial Basis Function Kernel* 91
Gambar 5.18 Grafik Pengaruh *Gamma* Terhadap *Weight Vector* Pada *Radial Basis Function Kernel* 92
Gambar 5.19 Grafik Pengaruh *Total N-Gram* Terhadap *Weight Vector* Pada *Radial Basis Function Kernel* 93

DAFTAR TABEL

Tabel 2.1 Perbandingan Penelitian Terdahulu 7
Tabel 2.2 Karakter Yang Dapat Membentuk Serangan *SQL Injection* 11
Tabel 2.3 *Query* Yang Dapat Membentuk Serangan *SQL Injection* 11
Tabel 2.4 Simbol *Flowchart* 25
Tabel 4.1 Struktur Tabel *Dataset* 44
Tabel 4.2 Struktur Tabel Label 44
Tabel 4.3 Struktur Tabel *Category* 44
Tabel 4.4 Struktur Tabel Hasil *Training Dataset* 45
Tabel 4.5 Struktur Tabel *Web Host* 46
Tabel 4.6 Struktur Tabel *Web Host HTTP Traffic* 47
Tabel 4.7 Struktur Tabel *HTTP Traffic Data* 48
Tabel 4.8 Struktur Tabel *Malicious Keyword* 48
Tabel 4.9 Daftar *REST API* Sistem Pada *Backend Website* 49
Tabel 5.1 Daftar *Dataset SQL Injection* 73
Tabel 5.2 Hasil *Dataset* Setelah Dilakukan *Text Preprocessing* Pada *Training Dataset* Kelas *SQLI Error Based - SQLI Time Based* 74
Tabel 5.3 Hasil *Dataset* Setelah Dilakukan *N-Gram* Pada *Training Dataset* Kelas *SQLI Error Based - SQLI Time Based* 74
Tabel 5.4 Langkah Perhitungan *TF-IDF* Pada *Training Dataset* Kelas *SQLI Error Based - SQLI Time Based* 76
Tabel 5.5 Hasil Perhitungan *TF-IDF* Pada *Training Dataset* Kelas *SQLI Error Based - SQLI Time Based* 77
Tabel 5.6 *Base Feature* Yang Terbentuk Pada *Training Dataset* Kelas *SQLI Error Based - SQLI Time Based* 77
Tabel 5.7 Hasil Fitur *Dataset* pada *Training Dataset* Kelas *SQLI Error Based - SQLI Time Based* 78
Tabel 5.8 Langkah Perhitungan Model Menggunakan Algoritma *Support Vector Machine* pada *Training Dataset* Kelas *SQLI Error Based - SQLI Time Based* 79
Tabel 5.9 Hasil Nilai *Weight Vector* Pada *Training Dataset* Kelas *SQLI Error Based - SQLI Time Based* 80

Tabel 5.10 Data Serangan *SQL Injection* Dari *Tool SQLMap* 83
Tabel 5.11 Langkah Perhitungan Klasifikasi Menggunakan Algoritma *Support Vector Machine* Pada Kelas *SQLI Error Based - SQLI Time Based* 85
Tabel 5.12 Data Serangan *SQL Injection* Hasil Klasifikasi 86
Tabel 5.13 Jumlah Klasifikasi Data Serangan *SQL Injection* Berdasarkan *Actual Label* Dan *Predicted Label* 88
Tabel 5.14 Nilai *Confusion Matrix* Hasil Klasifikasi Data Serangan *SQL Injection* 89
Tabel 5.15 Langkah Perhitungan Akurasi Menggunakan *Confusion Matrix* 89
Tabel 5.16 Hasil Pengujian *Parameter* Fungsi *Radial Basis Function Kernel* ke-1 90
Tabel 5.17 Hasil Pengujian *Parameter* Fungsi *Radial Basis Function Kernel* Ke-2 91
Tabel 5.18 Hasil Pengujian *Parameter* Fungsi *Radial Basis Function Kernel* Ke-3 92
Tabel 5.19 Hasil Pengujian *Parameter* Fungsi *Radial Basis Function Kernel* ke-4 93

DAFTAR KODE PROGRAM

Kode Program 5.1 *Konfigurasi Proxy Server* 60
Kode Program 5.2 Fungsi *Text Preprocessing* 66
Kode Program 5.3 Fungsi *N-Gram* 66
Kode Program 5.4 Fungsi *Term Frequency-Inverse Document Frequency* 67
Kode Program 5.5 Fungsi *Text Vectorization* 68
Kode Program 5.6 Fungsi *Radial Basis Function Kernel* 69
Kode Program 5.7 Fungsi *Training Dataset* Dengan *Support Vector Machine* ... 69
Kode Program 5.8 Fungsi *Training Dataset* Dengan *Support Vector Machine Multiclass* 70
Kode Program 5.9 Fungsi Klasifikasi Dengan *Support Vector Machine* 71
Kode Program 5.10 Fungsi Klasifikasi Dengan *Support Vector Machine Multiclass* 72
Kode Program 5.11 Perintah Eksekusi *SQLMap* Serangan *Error-Based* 82
Kode Program 5.12 Perintah Eksekusi *SQLMap* Serangan *Time-Based Blind* 82
Kode Program 5.13 Perintah Eksekusi *SQLMap* Serangan *Union Query-Based* . 82

BAB 1
PENDAHULUAN

1.1 Latar Belakang

Internet merupakan salah satu jaringan komunikasi terbesar di dunia. Tercatat pengguna internet di dunia mencapai 5,07 miliar orang per Oktober 2022 termasuk Indonesia (Kemp, 2023). *Web application* merupakan salah satu aplikasi media pertukaran informasi yang dapat diakses menggunakan internet melalui *browser*.

Banyak hal yang dapat dilakukan melalui internet terutama pada *web application* misalnya transaksi jual beli barang dan menyimpan data pribadi melalui *website* di mana pun dan kapan pun. Hal ini dapat memudahkan sebagian orang untuk melakukan aktivitas sehari-hari, namun hal ini juga menarik perhatian sebagian orang untuk melakukan tindakan kejahatan seperti mencuri data dari *website*.

Terdapat banyak tindakan kejahatan yang dapat dilakukan pada *website*. Berdasarkan data dari *Open Web Application Security Project* (OWASP), terdapat 10 risiko keamanan *website* teratas salah satunya adalah rentan terhadap serangan injeksi dengan posisi 3 teratas (OWASP Foundation, 2023). Serangan injeksi yang termasuk adalah *Cross-Site Scripting* (XSS), *SQL injection, Remote Code Execution* (RCE), *Path Traversal*, dan lain-lain. Serangan injeksi terjadi dalam *network-layer* pada model *Open Systems Interconnection* (OSI) yang dikembangkan oleh *International Organization for Standardization*.

Serangan injeksi berjalan melalui *Hypertext Transfer Protocol* (HTTP) di dalam paket-paket data yang dikirim dari *client-side* (*frontend*) ke *server-side* (*backend*) untuk diproses kemudian *server-side* mengirim balik respons *HTTP* ke *client-side*. Respons balik yang diterima oleh *client-side* akan ditangkap oleh *browser* dan ditampilkan ke pengguna. Hal ini akan berbahaya jika pengguna mencoba mencuri data atau bahkan merusak *server* melalui *client-side* (Cahyanto dkk., 2022). Untuk mencegah dari serangan kode yang aneh diperlukan sebuah sistem untuk deteksi serangan tersebut sebelum diproses oleh *server-side* (Habibi & Surantha, 2020). Salah satu solusi yang dapat dilakukan adalah melakukan

klasifikasi serangan injeksi pada *HTTP request* menggunakan algoritma seperti *support vector machine*.

Support Vector Machine (SVM) adalah algoritma pembelajaran mesin (*supervised learning*) untuk melakukan klasifikasi dan regresi, yang bekerja dengan mencari *hyperplane* optimal untuk memisahkan kelas-kelas data dengan margin maksimal. *SVM* menggunakan fitur sebagai dasar untuk memisahkan kelas-kelas data. Fitur tersebut diambil dan diekstraksi berdasarkan korelasi terhadap dataset dan kelasnya. *SVM* menggunakan *kernel* untuk memisahkan data secara *linear* maupun data secara *non-linear* misalnya dengan *linear kernel*, *polynomial kernel*, dan *radial basis function kernel*. Pada dasarnya *SVM* hanya memisahkan antara dua kelas menggunakan *hyperplane* dan dapat diperluas menjadi *multiclass* (lebih dari dua kelas) menggunakan strategi *one against one* (membandingkan setiap pasangan kelas) atau o*ne against all* (membandingkan satu kelas dengan semua kelas lainnya), memungkinkan klasifikasi efektif pada *dataset* dengan lebih dari dua kelas, meskipun ini dapat meningkatkan kompleksitas komputasi seiring bertambahnya jumlah kelas.

Adapun penelitian yang membahas tentang deteksi maupun klasifikasi serangan injeksi sudah pernah dilakukan sebelumnya dengan judul "*XSS Attack Detection with Machine Learning and N-Gram Methods*" (Habibi & Surantha, 2020). Penelitian ini membandingkan hasil performa dari tiga metode *machine learning* sebelum dan sesudah dikombinasikan dengan metode *n-gram*. Adapun ketiga metode yang digunakan adalah *support vector machine*, *K-Nearest Neighbor*, dan *Naïve Bayes*. Hasil dari penelitian ini adalah metode *support vector machine* yang dikombinasikan dengan metode *n-gram* memiliki kemampuan deteksi serangan *Cross-Site Scripting* (XSS) dengan nilai terbaik yaitu 99% *recall*, 99% *precision*, dan 98% *accuracy*.

Penelitian kedua dengan judul "Pengujian *Rule-Based* Pada *Dataset Log Server* Menggunakan *support vector machine* Berbasis *Linear Discriminant Analysis* Untuk Deteksi *Malicious Activity*" (Cahyanto dkk., 2022). Penelitian ini membandingkan hasil pengujian metode *support vector machine* dengan *kernel trick Radial Basis Function* (RBF) dengan dan tanpa kombinasi dari *Linear Discriminant Analysis* (LDA) untuk seleksi fiturnya. Adapun *dataset* yang

digunakan berasal dari *log server*. Hasil dari penelitian ini adalah *support vector machine* dengan *kernel trick radial basis function* yang dikombinasikan dengan *Linear Discriminant Analysis* memiliki hasil yang terbaik yaitu dengan 89,67% *recall*, 90,67% *precision*, dan 89,95% *accuracy*.

Penelitian ketiga dengan judul "Klasifikasi *SQL injection* Menggunakan Algoritma *Naïve Bayes*" (Prasetyo dkk., 2021). Penelitian ini menghasilkan nilai *accuracy* dari algoritma *Naïve Bayes* sebesar 79,82%. Klasifikasi yang dilakukan hanya berfokus pada 3 jenis serangan *SQL injection* yaitu *union*, *blind*, dan *tautology* dengan *dataset* yang berasal dari *website Kaggle*. Dilakukannya *preprocessing*, *tokenization*, dan *feature extraction* terlebih dahulu pada *dataset*, lalu kemudian hasil *feature extraction* digunakan untuk klasifikasi menggunakan algoritma *Naïve Bayes.*

Berdasarkan penelitian-penelitian yang telah dilakukan, maka diangkatlah penelitian yang berjudul "Klasifikasi Serangan *SQL injection* Menggunakan Algoritma *support vector machine* pada *HTTP Request*". Penelitian ini akan membangun suatu sistem untuk klasifikasi serangan *SQL injection* yang dikirim melalui *HTTP request*. Fitur klasifikasi akan diambil atau diekstraksi dari pola-pola dataset menggunakan *N-Gram*. Dengan adanya penelitian ini diharapkan mendapatkan hasil *accuracy* yang terbaik dalam klasifikasi *SQL injection*.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, dapat dirumuskan beberapa masalah sebagai berikut:

1. Bagaimana cara melakukan ekstraksi fitur menggunakan *n-gram* dan *Term Frequency-Inverse Document Frequency* (TF-IDF) pada *dataset* klasifikasi serangan *SQL injection*?
2. Bagaimana cara melakukan *training dataset* menggunakan algoritma *support vector machine* pada *dataset* klasifikasi serangan *SQL injection* yang memiliki lebih dari dua kelas (*multiclass*)?
3. Bagaimana hasil *accuracy* algoritma *support vector machine* dalam klasifikasi serangan *SQL injection* pada *HTTP request*?

1.3 Batasan Masalah

Agar pembahasan tetap fokus pada pokok masalah dan mempermudah pembahasan supaya tujuan penelitian tercapai, maka diberikan batasan masalah sebagai berikut:

1. Pada penelitian ini menggunakan algoritma *support vector machine* untuk melakukan *training dataset* dan *klasifikasi* serangan *SQL injection*.
2. *Kernel* pada algoritma *support vector machine* yang digunakan untuk *training dataset* dan klasifikasi serangan *SQL injection* adalah *Radial Basis Function Kernel*.
3. Kelas klasifikasi yang digunakan pada penelitian ini ada tiga yaitu *SQLI error based*, *SQLI time based*, dan *SQLI union query*.
4. Ekstraksi fitur pada *dataset* klasifikasi *SQL injection* menggunakan metode *n-gram* dan *TF-IDF* (*Term Frequency-Inverse Document Frequency*).
5. Sumber data pada *HTTP request* yang diproses untuk klasifikasi *SQL injection* adalah *query params*, dan *body*.
6. *Dataset* yang digunakan untuk *training dataset* berjumlah 262 data dengan pembagian masing-masing kelasnya yaitu 72 *SQLI error based*, 133 *SQLI time based*, 57 *SQLI union query*.
7. Data uji yang digunakan untuk melakukan serangan berjumlah 557 data dengan pembagian masing-masing kelasnya yaitu 220 *SQLI error based*, 220 *SQLI time based*, 117 *SQLI union query*.
8. Proses pengujian serangan *SQL injection* pada *HTTP request* menggunakan *tool SQLMap* dengan batasan jenis serangan yaitu hanya serangan *error based*, *union query based*, dan *time based blind*.
9. Parameter yang digunakan untuk melakukan *training dataset* yaitu 10 *epoch*, 0,1 *learning rate*, 4 *total n-gram*, dan 0,1 *gamma*.
10. Fitur *machine learning* yang digunakan untuk *training dataset* didapat dari hasil proses *n-gram* pada *dataset* dengan parameter 4 *total n-gram*.
11. Sistem dibuat berbasis *website*.

1.4 Tujuan Penelitian

Adapun tujuan dari penelitian ini yaitu sebagai berikut:

1) Menjelaskan cara melakukan ekstraksi fitur menggunakan *n-gram* dan *Term Frequency-Inverse Document Frequency* (TF-IDF) pada *dataset* klasifikasi serangan *SQL injection*.
2) Menjelaskan cara melakukan *training dataset* menggunakan algoritma *support vector machine* pada *dataset* klasifikasi serangan *SQL injection* yang memiliki lebih dari dua kelas klasifikasi (*multiclass*).
3) Mengetahui hasil *accuracy* algoritma *support vector machine* dalam klasifikasi serangan *SQL injection* pada *HTTP request*.

1.5 Manfaat Penelitian

Adapun manfaat dari penelitian ini yaitu sebagai landasan ilmu masyarakat, individu, maupun, lembaga untuk mempelajari cara kerja dari klasifikasi serangan *SQL injection* pada *Hypertext Transfer Protocol* (HTTP) *request* menggunakan algoritma *support vector machine* dan dapat menerapkannya pada sistem agar sistem lebih aman. Adanya penelitian ini diharapkan dapat menjadikan landasan ilmu agar sistem yang lebih aman dari *SQL injection* maupun serangan sejenis lainnya.

1.6 Sistematika Penulisan

Penulisan skripsi ini dibagi menjadi 6 bab agar permasalahan dapat dijelaskan dan diuraikan untuk menyelesaikan penelitian ini. Adapun sistematika penulisan tugas akhir ini terdiri dari:

BAB 1 PENDAHULUAN

Bab pendahuluan berisi latar belakang untuk menjelaskan kenapa penelitian ini dilakukan, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan.

BAB 2 LANDASAN TEORI

Bab landasan teori berisi bahasan mengenai perbandingan penelitian terdahulu, dan teori penunjang yang dibutuhkan dalam melakukan penelitian sesuai dengan rancangan dan analisa yang dilakukan selama penelitian.

BAB 3 METODE PENELITIAN

Bab Metode penelitian adalah prosedur atau cara-cara yang dilakukan untuk mengumpulkan dan menganalisis data dalam penelitian. Adapun isi dari metode penelitian yaitu studi literatur, pengumpulan data, analisis kebutuhan, perancangan sistem, implementasi, pengujian sistem, kesimpulan dan saran.

BAB 4 PERANCANGAN

Bab perancangan berisi secara rinci tentang perancangan sistem dalam pelaksanaan penelitian. Pada bab ini secara umum meliputi deskripsi sistem, gambaran secara umum proses-proses yang terjadi pada sistem, diagram alir (*flowchart*) sistem, rancangan *database*, rancangan antarmuka tampilan aplikasi, dan rancangan pengujian sistem pada sistem yang digunakan pada penelitian ini.

BAB 5 IMPLEMENTASI, PENGUJIAN, DAN PEMBAHASAN

Bab implementasi, pengujian, dan pembahasan berisi tentang implementasi dari bab perancangan, pengujian sistem, dan pembahasan langkah demi langkah penelitian serta perhitungan pada saat penelitian.

BAB 6 PENUTUP

Bab penutup berisi kesimpulan dari keseluruhan penelitian yang telah dilakukan dan saran untuk pengembangan penelitian selanjutnya agar dapat dikembangkan menjadi lebih baik.