

**EVALUASI PERFORMA SUPPORT VECTOR MACHINE
DALAM DETEKSI SQL INJECTION DENGAN
CONFUSION MATRIX**

SKRIPSI

Diajukan untuk memenuhi salah satu syarat memperoleh gelar Sarjana Komputer
Pada Jenjang Program Sarjana – Program Studi Teknik Komputer



disusun oleh

FATA NAUFAL HANAN

20.83.0562

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2024

**EVALUASI PERFORMA SUPPORT VECTOR MACHINE
DALAM DETEKSI SQL INJECTION DENGAN
CONFUSION MATRIX**

SKRIPSI

Diajukan untuk memenuhi salah satu syarat memperoleh gelar Sarjana Komputer
Pada Jenjang Program Sarjana – Program Studi Teknik Komputer



disusun oleh

FATA NAUFAL HANAN

20.83.0562

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2024

HALAMAN PERSETUJUAN

SKRIPSI

**EVALUASI PERFORMA SUPPORT VECTOR MACHINE DALAM
DETEKSI SQL INJECTION DENGAN
CONFUSION MATRIX**

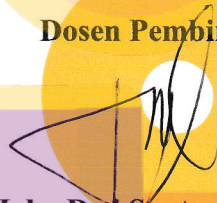
yang disusun dan diajukan oleh

Fata Naufal Hanan

20.83.0562

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 20 Agustus 2024

Dosen Pembimbing,



Joko Dwi Santoso, M.Kom

NIK. 190302181

HALAMAN PENGESAHAN
SKRIPSI
EVALUASI PERFORMA SUPPORT VECTOR MACHINE DALAM
DETEKSI SQL INJECTION DENGAN
CONFUSION MATRIX

yang disusun dan diajukan oleh

Fata Naufal Hanan

20.83.0562

Telah dipertahankan di depan Dewan Penguji
pada tanggal 20 Agustus 2024

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Muhammad Kopravi, S.Kom., M.Eng
NIK. 190302454

Ali Mustopa, M.Kom
NIK. 190302192

Joko Dwi Santoso, M.Kom
NIK. 190302181



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 20 Agustus 2024

DEKAN FAKULTAS ILMU KOMPUTER



Hanif Al Fatta, S.Kom., M.Kom., Ph.D.
NIK. 190302096

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Fata Naufal Hanan

NIM : 20.83.0562

Menyatakan bahwa Skripsi dengan judul berikut:

Evaluasi Performa Support Vector Machine dalam Deteksi SQL Injection dengan Confusion Matrix

Dosen Pembimbing : Joko Dwi Santoso, M.Kom.

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, 20 Agustus 2024

Yang Menyatakan,

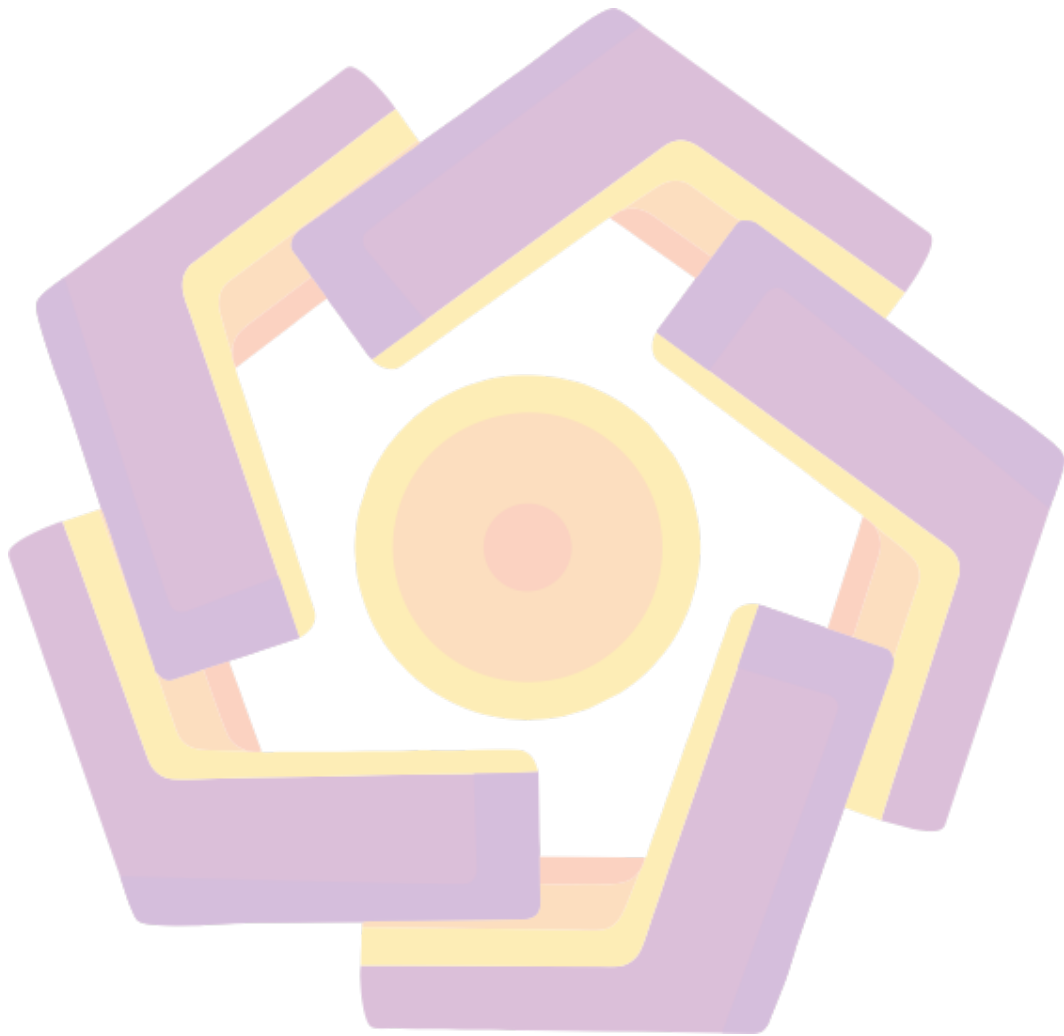


Fata Naufal Hanan

HALAMAN MOTTO

“Man Jadda Wajada (Siapa yang bersungguh-sungguh, dia akan dapat)”

Pepatah Arab



HALAMAN PERSEMBAHAN

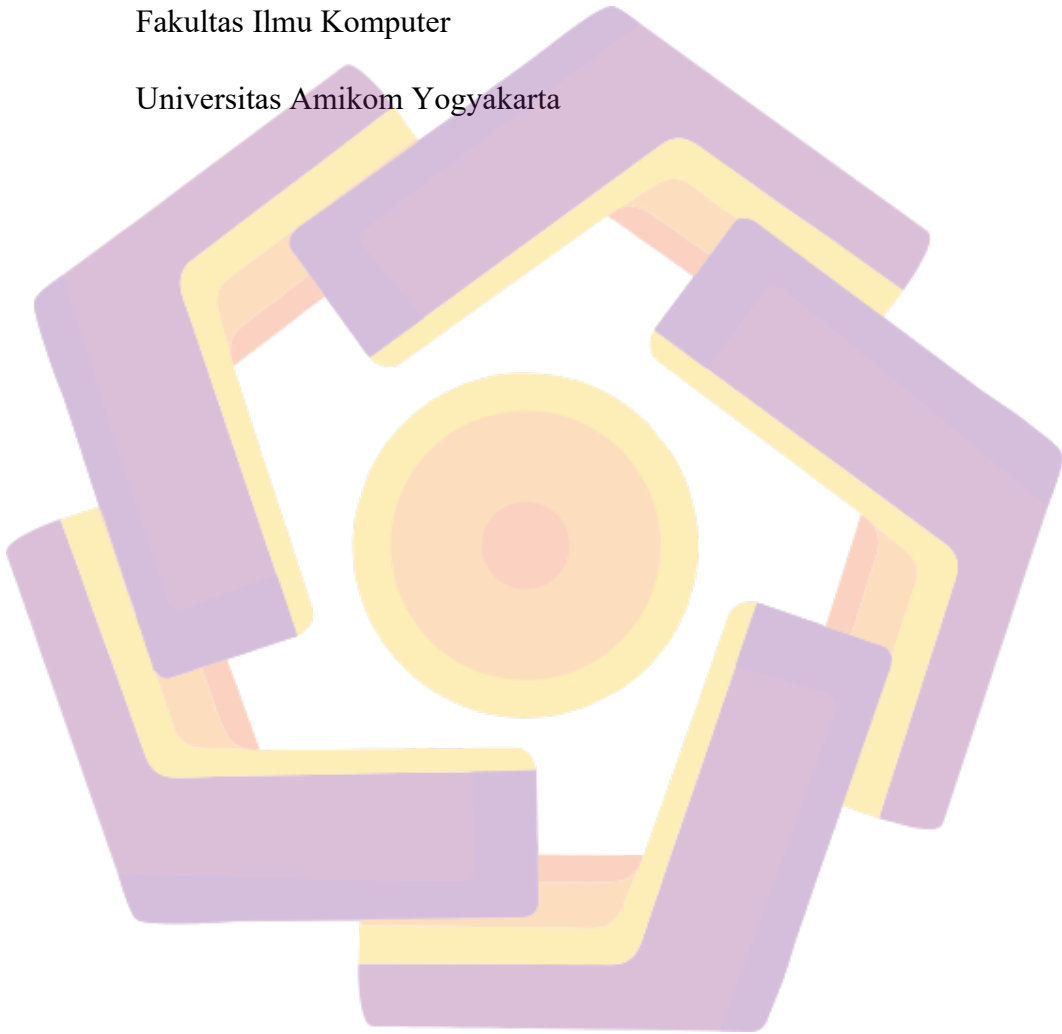
Skripsi ini peneliti persembahkan kepada:

Almameter tercinta

Program Studi Teknik Komputer

Fakultas Ilmu Komputer

Universitas Amikom Yogyakarta



KATA PENGANTAR

Alhamdulillah Rabbil ‘Alamiin, puji syukur ke hadirat Allah Swt. atas limpahan rahmat, taufik dan hidayah-Nya. Peneliti menyadari bahwa dalam proses penyusunan skripsi ini tidak terlepas dari kemudahan dan pertolongan dari Allah Swt. Selawat serta salam senantiasa tercurahkan kepada Nabi Muhammad SAW. Beserta keluarga, sahabat dan sampai kepada ummatnya.

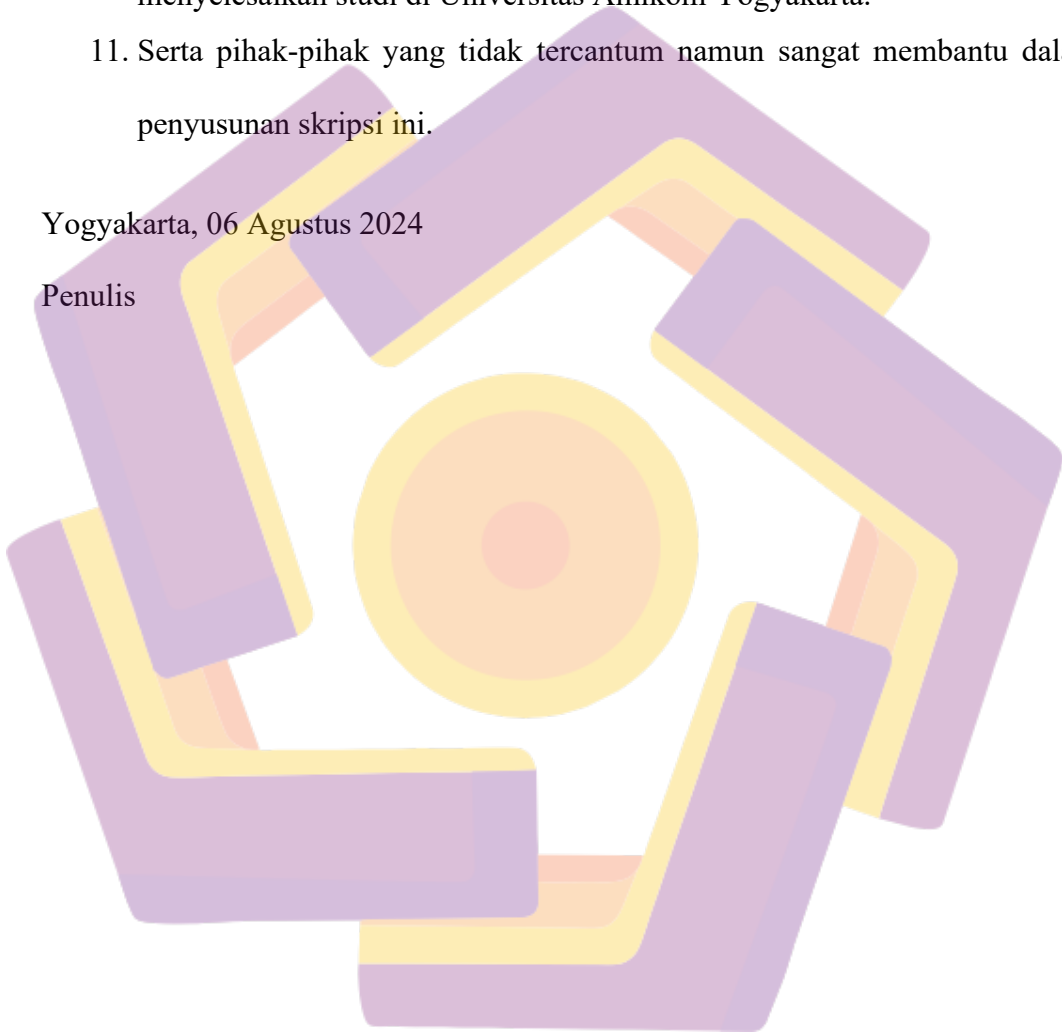
Penyusunan skripsi ini sebagai salah satu persyaratan dalam menyelesaikan program Sarjana Komputer (S.Kom.) Strata S-1 Program Studi Teknik Komputer Fakultas Ilmu Komputer Universitas Amikom Yogyakarta, tak lepas dari dukungan dan doa dari berbagai pihak yang telah membantu dalam penyelesaiannya. Maka dari itu dengan kerendahan hati, penulis mengucapkan terima kasih kepada:

1. Prof, Dr. M. Suyanto, MM., selaku Rektor Universitas Amikom Yogyakarta.
2. Hanif Al Fatta, S.Kom., M.Kom., Ph.D., selaku Dekan Fakultas Ilmu Komputer Universitas Amikom Yogyakarta.
3. Dony Ariyus, M.Kom., selaku Ketua Program Studi Teknik Komputer Universitas Amikom Yogyakarta.
4. Banu Santoso, S.T., M.Eng., selaku Sekretaris Program Studi Teknik Komputer Universitas Amikom Yogyakarta.
5. Joko Dwi Santoso, M.Kom., selaku Dosen Pembimbing Skripsi yang telah meluangkan waktunya untuk berbagi ilmu, memberikan arahan, doa serta dukungan, sehingga skripsi ini dapat terselesaikan.
6. Sumarni Adi, S.Kom., M.Cs., selaku Dosen Pembimbing Akademik yang telah memberikan arahan selama masa perkuliahan.
7. Para Dosen Program Studi Teknik Komputer Universitas Amikom Yogyakarta yang telah memberikan ilmu yang bermanfaat, pengalaman dan nasehat-nasehat yang berguna.
8. Orang tua tercinta Bapak Muslihudin dan Ibu Nur Ngazizah yang selama ini telah mendidik, memberikan dukungan baik moral maupun materil secara ikhlas, serta doa yang senantiasa dipanjatkan.

9. Kepada teman-teman satu angkatan Program Studi Teknik Komputer yang yang senantiasa memberikan dukungan sehingga selalu termotivasi untuk menyelesaikan skripsi ini.
10. Kepada sahabat-sahabat seperjuangan, dengan NIM 20.83.0561 dan 20.83.0574, yang telah bersusah payah berjuang bersama untuk menyelesaikan studi di Universitas Amikom Yogyakarta.
11. Serta pihak-pihak yang tidak tercantum namun sangat membantu dalam penyusunan skripsi ini.

Yogyakarta, 06 Agustus 2024

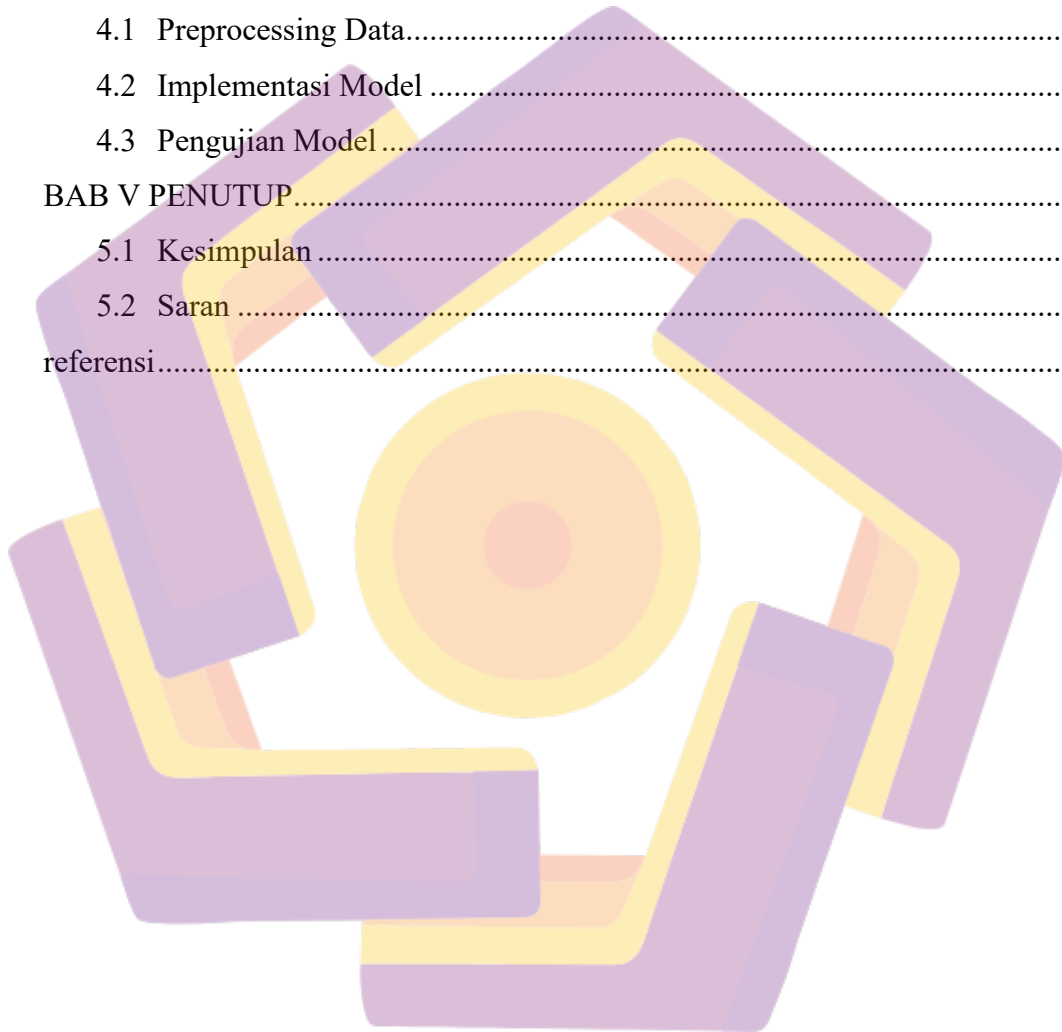
Penulis



DAFTAR ISI

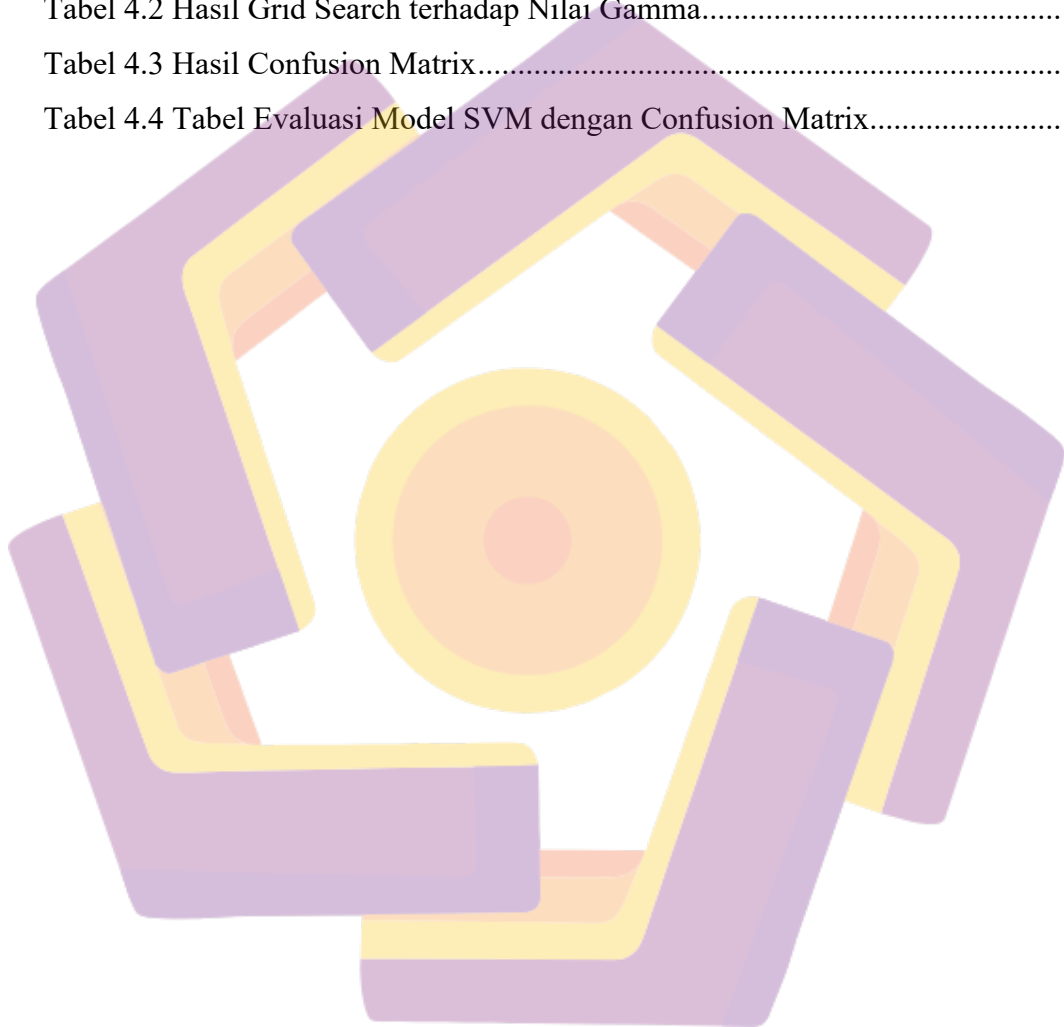
HALAMAN JUDUL.....	2
HALAMAN PERSETUJUAN.....	iii
HALAMAN PENGESAHAN.....	iv
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	v
HALAMAN MOTTO	vi
HALAMAN PERSEMBAHAN	vii
KATA PENGANTAR	viii
DAFTAR ISI.....	x
DAFTAR TABEL.....	xii
DAFTAR GAMBAR.....	xiii
DAFTAR LAMBANG DAN SINGKATAN	xiv
DAFTAR ISTILAH	xv
INTISARI.....	xvi
ABSTRACT.....	xvii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah.....	2
1.3 Batasan Masalah	3
1.4 Tujuan Penelitian	3
1.5 Sistematika Penulisan	4
BAB II TINJAUAN PUSTAKA.....	5
2.1 Studi Literatur	5
2.2 Basis data Relasional	8
2.3 SQL (<i>Structured Query Language</i>)	9
2.4 Serangan SQLI (<i>Structured Query Language Injection</i>)	10
2.5 Metode Deteksi SQL Injection	12
2.7 SVM (<i>Support Vector Machine</i>).....	14
2.8 Evaluasi Sistem Klasifikasi.....	17
BAB III METODE PENELITIAN	19
3.1 Alur Penelitian	19

3.1.1. Pengumpulan Data.....	20
3.1.2. Exploratory Data Analysis (EDA).....	20
3.1.3. Preprocessing Data	21
3.1.4. Perancangan Model <i>Machine Learning</i>	21
3.2 Alat dan Bahan.....	24
BAB IV Hasil dan PEMBAHASAN.....	25
4.1 Preprocessing Data.....	25
4.2 Implementasi Model	31
4.3 Pengujian Model	34
BAB V PENUTUP.....	39
5.1 Kesimpulan	39
5.2 Saran	39
referensi.....	40



DAFTAR TABEL

Tabel 2.1 Keaslian Penelitian.....	7
Tabel 2.1 Tabel <i>Confusion Metrix</i>	17
Tabel 4.1 Tabel Ekstraksi Fitur.....	29
Tabel 4.1 Lanjutan	30
Tabel 4.2 Hasil Grid Search terhadap Nilai Gamma.....	34
Tabel 4.3 Hasil Confusion Matrix.....	36
Tabel 4.4 Tabel Evaluasi Model SVM dengan Confusion Matrix.....	37

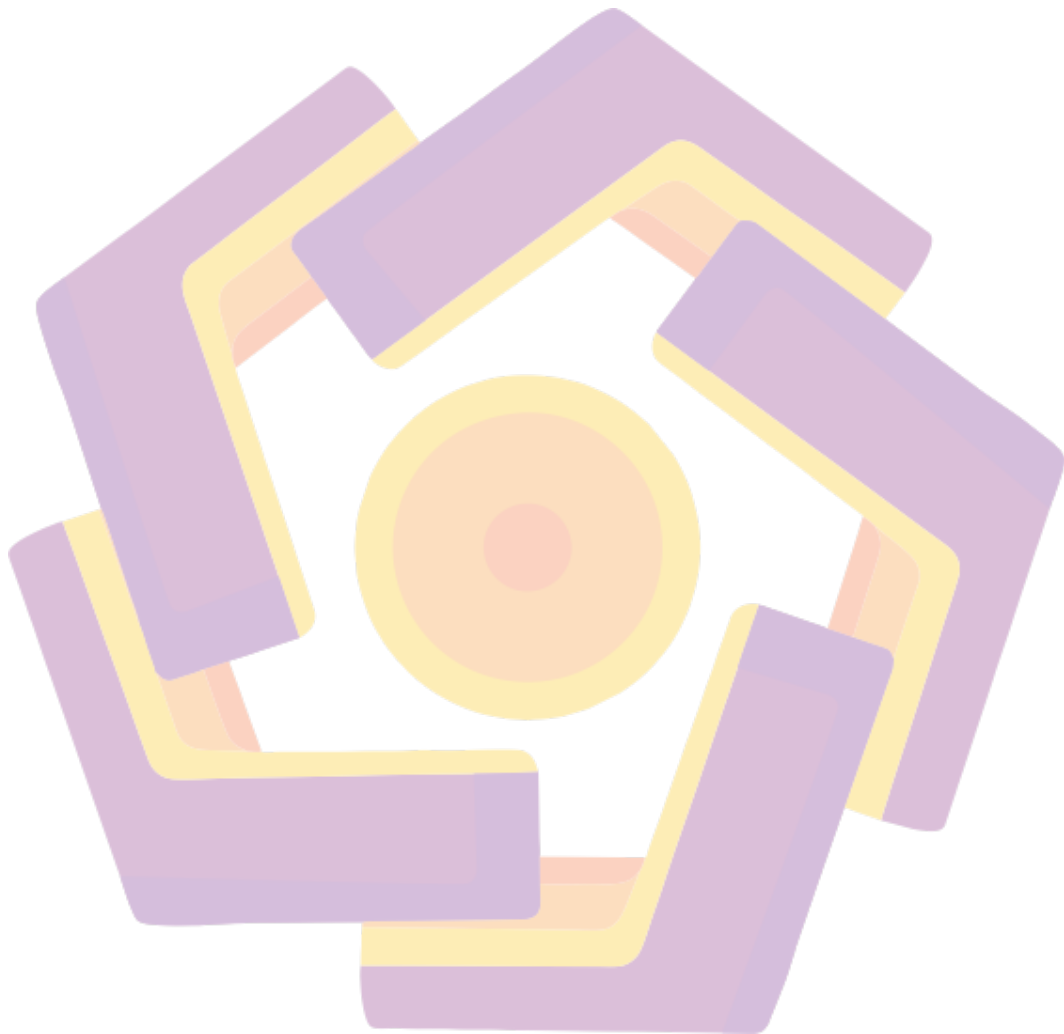


DAFTAR GAMBAR

Gambar 2.1 Visualisasi Klasifikasi Biner SVM [29].....	15
Gambar 2.2 <i>Feature Space</i> pada SVM [30].....	16
Gambar 3.1 Flowchart Penelitian.....	19
Gambar 3.2 Diagram Proses Preprocessing.....	21
Gambar 3.3 Flow Chart Perancangan Model.....	24
Gambar 4.1 Struktur dan Tipe Data dari Setiap Kolom Dataset.....	25
Gambar 4.2 Kode untuk Menghapus Data Bernilai Null.....	26
Gambar 4.3 Kode untuk Menemukan Data Duplikat.....	26
Gambar 4.4 Kode untuk Menghapus Data Duplikat.....	26
Gambar 4.5 Kode untuk Menemukan Data Duplikat dengan Label Berbeda.....	26
Gambar 4.6 Kode untuk Menghapus Data Duplikat dengan Label Berbeda.....	27
Gambar 4.7 Distribusi Data terhadap Label Sebelum Pembersihan Data	27
Gambar 4.8 Distribusi Data Terhadap Label Setelah Pembersihan Data	28
Gambar 4.9 Kata dengan Frekuensi Terbanyak pada Kueri SQL Injection	28
Gambar 4.10 Kata dengan Frekuensi Terbanyak pada Kueri non-SQL Injection	29
Gambar 4.11 Kode Case Folding	29
Gambar 4.12 Persebaran Fitur Yang Telah Diekstraksi.....	31
Gambar 4.13 Kode Pembagian Dataset	32
Gambar 4.14 Teknik <i>Bag of Words</i>	32
Gambar 4.15 Kode Encoding Fitur <code>no_single_qts</code>	33
Gambar 4.16 Kode Implementasi Model SVM	33
Gambar 4.17 Kode Teknik Grid Search.....	33
Gambar 4.18 Confusion Matrix SVM dengan $\text{Gamma}=0.034$	35
Gambar 4.19 Confusion Matrix SVM dengan $\text{Gamma}=0.23$	35
Gambar 4.20 Learning Curve SVM setelah Hyperparameter Tuning	38

DAFTAR LAMBANG DAN SINGKATAN

SVM	Support Vector Machine
SQL	Structure Query Language
σ	Standar deviasi



DAFTAR ISTILAH

<i>Machine Learning</i>	Pembelajaran Mesin
<i>Accuracy</i>	Tingkat akurasi
<i>Precision</i>	Tingkat presisi
<i>Recall</i>	Tingkat Sensitivitas
Alternatif	Pilihan lain
Implikasi	Akibat langsung
File	Kumpulan data tersimpan
Server	Sebuah sistem komputer atau perangkat lunak
Kueri	intruksi untuk melakukan permintaan terhadap basis data



INTISARI

Aplikasi web sudah umum digunakan untuk membantu berbagai aktivitas masyarakat di era modern. Basis data menjadi bagian penting dari aplikasi web untuk menyimpan berbagai data, termasuk data yang sensitif. Menjaga keamanan basis data menjadi tantangan serius terhadap berbagai ancaman siber, terutama ancaman serangan Structure Query Language Injection (SQLI). Serangan SQL Injection bekerja dengan memasukkan kode yang telah disisipi kueri SQL yang berbahaya ke dalam aplikasi web untuk mendapatkan akses yang tidak sah ke basis data. Serangan ini dapat menyebabkan kerugian yang berkelanjutan mulai dari pencurian data hingga mendapatkan akses terhadap seluruh sistem.

Seiring perkembangan teknologi, muncul berbagai alternatif cara untuk mendeteksi SQL Injection, salah satunya dengan menggunakan machine learning. Dalam penelitian ini, penulis menggunakan pendekatan preprocessing Bag of Words dan algoritma Support Vector Machine (SVM) untuk membuat sistem deteksi SQL Injection. Hasil evaluasi menggunakan confusion matrix menunjukkan bahwa pendekatan Bag of Words pada algoritma SVM dengan kernel RBF menghasilkan performa yang sangat baik dengan nilai akurasi sebesar 0,995, precision 0,995, recall 0,992, dan F1 Score 0,993. Performa SVM meningkat setelah dilakukan hyperparameter tuning menggunakan grid search terhadap nilai gamma menjadikan nilai akurasi sebesar 0,996, precision 0,997, recall 0,990, dan F1 Score 0,994. Pada kasus ini, hyperparameter yang berdampak signifikan terhadap meningkatnya performa sistem adalah nilai gamma.

Dengan demikian, penelitian ini berhasil mengevaluasi sistem deteksi SQL Injection dengan pendekatan Bag of Words dan algoritma SVM. Pendekatan Bag of Words berdampak signifikan terhadap kinerja model machine learning bahkan sebelum dilakukan hyperparameter tuning. Hasil ini memiliki implikasi positif dalam evaluasi dan peningkatan performa SVM untuk mendeteksi SQL Injection dengan kinerja yang optimal.

Kata kunci: Basis data, Injeksi SQL, SVM, Confusion Matrix, Bag of Words

ABSTRACT

Web applications are commonly used to assist various activities in modern society. Databases are an essential part of web applications for storing various data, including sensitive information. Ensuring database security is a serious challenge in the face of various cyber threats, particularly the threat of SQL Injection (SQLI) attacks. SQL Injection attacks work by inserting malicious SQL query code into web applications to gain unauthorized access to the database. These attacks can lead to significant damages, ranging from data theft to gaining control of entire systems.

As technology advances, various alternatives for detecting SQL Injection have emerged, one of which is through the use of machine learning. In this study, the author uses a Bag of Words preprocessing approach and the Support Vector Machine (SVM) algorithm to create a SQL Injection detection system. The evaluation results using a confusion matrix show that the Bag of Words approach with the SVM algorithm using an RBF kernel produced excellent performance with an accuracy of 0.995, precision of 0.995, recall of 0.992, and an F1 Score of 0.993. The SVM's performance improved after hyperparameter tuning using grid search on the gamma value, resulting in an accuracy of 0.996, precision of 0.997, recall of 0.990, and an F1 Score of 0.994. In this case, the hyperparameter that significantly impacted the system's performance was the gamma value.

Thus, this study successfully evaluated the SQL Injection detection system using the Bag of Words approach and the SVM algorithm. The Bag of Words approach significantly impacted the machine learning model's performance even before hyperparameter tuning. These results have positive implications for the evaluation and performance improvement of SVM in detecting SQL Injection with optimal performance.

Keyword: Database, SQL Injection, SVM, Confusion Matrix, Bag of Words