

**INTEGRASI SISTEM PEMBELAJARAN MESIN UNTUK DETEKSI
INJEKSI SQL MENGGUNAKAN ALGORITMA RANDOM FOREST**

SKRIPSI



Disusun oleh:

Naufal Arieq Wira Pradana

20.83.0574

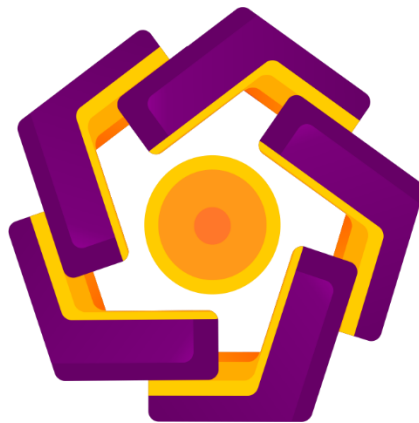
**PROGRAM SARJANA
PROGRAM STUDI TEKNIK KOMPUTER
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2024

**INTEGRASI SISTEM PEMBELAJARAN MESIN UNTUK DETEKSI
INJEKSI SQL MENGGUNAKAN ALGORITMA RANDOM FOREST**

SKRIPSI

Diajukan kepada Fakultas Ilmu Komputer Universitas AMIKOM Yogyakarta
untuk memenuhi salah satu syarat memperoleh gelar Sarjana Komputer
Pada Jenjang Program Sarjana – Program Studi Teknik Komputer



Disusun oleh:

Naufal Arieq Wira Pradana

20.83.0574

**PROGRAM SARJANA
PROGRAM STUDI TEKNIK KOMPUTER
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2024**

HALAMAN PERSETUJUAN

SKRIPSI

INTEGRASI SISTEM PEMBELAJARAN MESIN UNTUK DETEKSI INJEKSI SQL MENGGUNAKAN ALGORITMA RANDOM FOREST

yang disusun dan diajukan oleh

Naufal Arieq Wira Pradana
20.83.0574

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 22 Agustus 2024

Dosen Pembimbing,

Wahid Miftahul Ashari, S.Kom., M.T.
NIK. 190302452

HALAMAN PENGESAHAN

SKRIPSI

INTEGRASI SISTEM PEMBELAJARAN MESIN UNTUK DETEKSI INJEKSI SQL MENGGUNAKAN ALGORITMA RANDOM FOREST

yang disusun dan diajukan oleh

Naufal Arieq Wira Pradana

20.83.0574

Telah dipertahankan di depan Dewan Penguji
pada tanggal 22 Agustus 2024

Susunan Dewan Penguji

Nama Penguji

Banu Santoso, S.T., M.Eng
NIK. 190302327

Rina Pramitasari, S.Si., M.Cs
NIK. 190302335

Wahid Miftahul Ashari, S.Kom., M.T.
NIK. 190302452

Tanda Tangan

Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 22 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 : Naufal Arieq Wira Pradana

NIM : 20.83.0574

Menyatakan bahwa Skripsi dengan judul berikut:

INTEGRASI SISTEM PEMBELAJARAN MESIN UNTUK DETEKSI INJEKSI SQL MENGGUNAKAN ALGORITMA RANDOM FOREST

Dosen Pembimbing : Wahid Miftahul Ashari, S.Kom., M.T.

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, 22 Agustus 2024

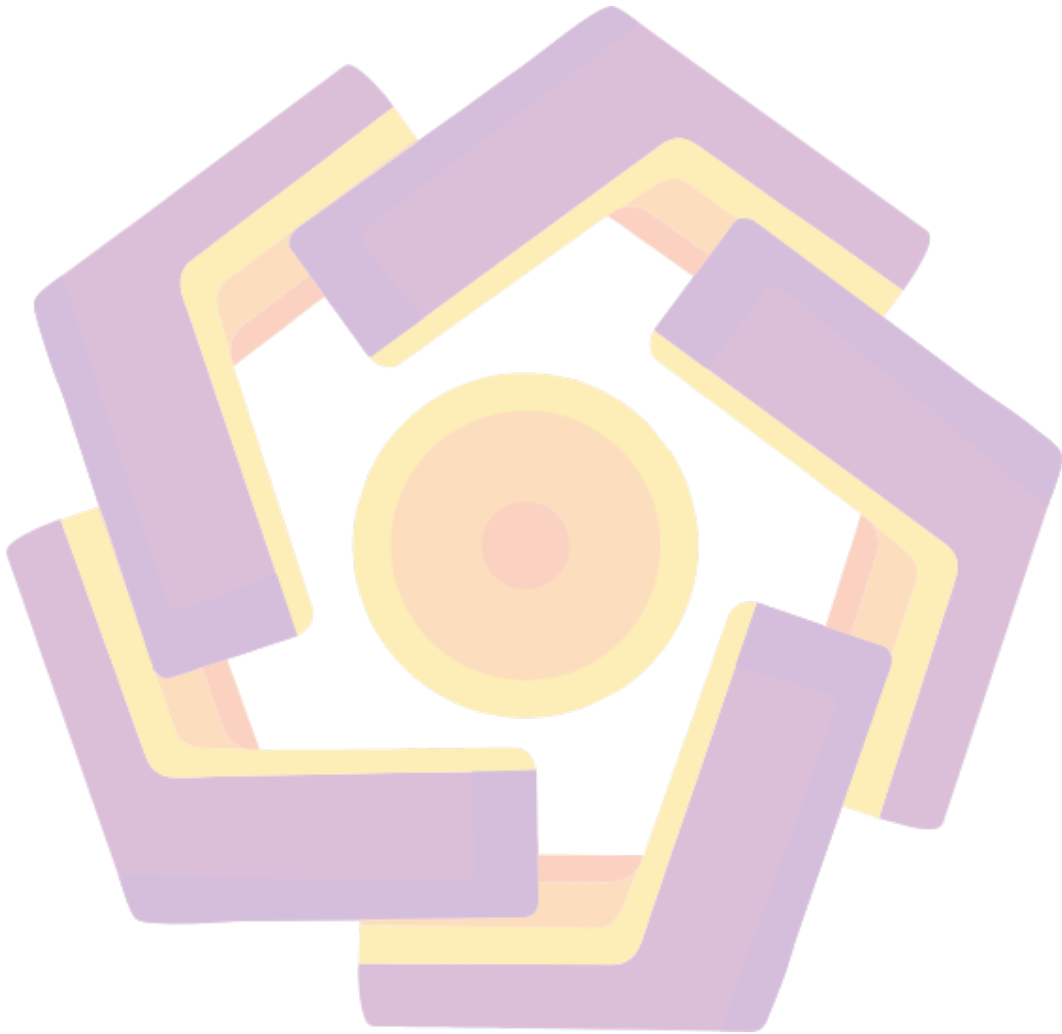
Yang Menyatakan,



Naufal Arieq Wira Pradana

HALAMAN MOTTO

"Dan bersabarlah. Sesungguhnya Allah beserta orang-orang yang sabar."
(QS: Al-Anfal ayat 46)



HALAMAN PERSEMBAHAN

Penelitian ini dibuat atas keinginan untuk mengeksplorasi lebih tentang implementasi dari model machine learning khususnya Random Forest serta, berawal dari topik pembicaraan tongkrongan rekan-rekan perkuliahan yang hendak maju lomba di bidang UI/UX namun, berbelok untuk melakukan penelitian bersama mendalami keilmuan tentang pembelajaran mesin. Saya sangat bersyukur atas hadirnya kawan-kawan yang senantiasa masih ingin belajar disela-sela kesibukan masing-masing dalam menjalani kehidupan. Semoga lewat pembahasan kali ini dapat digunakan untuk selalu berpikir kritis dan menumbuhkan kreativitas dari rekan-rekan yang sudah berjuang.



KATA PENGANTAR

Puji syukur peneliti panjatkan kepada hadirat Allah Swt. yang telah memberikan nikmat dan rezeki berupa kesehatan serta, kebutuhan material sehingga skripsi dengan judul “Integrasi Sistem Pembelajaran Mesin untuk Mendeteksi Serangan SQL Injection” dapat diselesaikan. Tak dapat disangkal bahwa dalam penyusunan skripsi ini telah menghadapi segala tantangan dan kesulitan. Hal tersebut tidak akan terselesaikan apabila peneliti tidak mendapatkan dukungan dari banyak pihak. Oleh karena itu, izinkan peneliti untuk menyampaikan rasa terima kasih kepada:

1. Allah Swt. atas limpahan nikmat lahir dan batin yang peneliti syukuri sehingga peneliti dapat menjalani hidup ini dengan lancar dan istiqomah dalam menghadapi tantangan pahit dan manis. Terima kasih telah menuliskan skenario takdir yang begitu bervariasi;
2. Kedua orang tua peneliti, Bondan Wirawan, S.E. dan Rafika Nilasari yang telah mendukung segala pilihan dalam hidup peneliti tanpa sedikit penyesalan, selalu berjuang demi kebutuhan peneliti apapun jenis dan bentuknya. Terima kasih telah menciptakan keluarga kecil yang penuh arti;
3. Keluarga besar peneliti, yang senantiasa memberikan dukungan akan kesuksesan peneliti dalam meniti perkuliahan hingga memasuki tahap akhir.
4. Dosen pembimbing, Wahid Miftahul Ashari, S.Kom., M.T., yang telah membantu dan mengarahkan penelitian ini sampai selesai. Terima kasih atas kesempatan diskusi dan pertukaran perspektif dalam setiap kesempatan bimbingan;
5. Magfira Dewinta Humairah, yang telah membersamai peneliti untuk berproses bersama dukungan dan tekanan yang selalu diberikan.

Yogyakarta, 15 Mei 2024

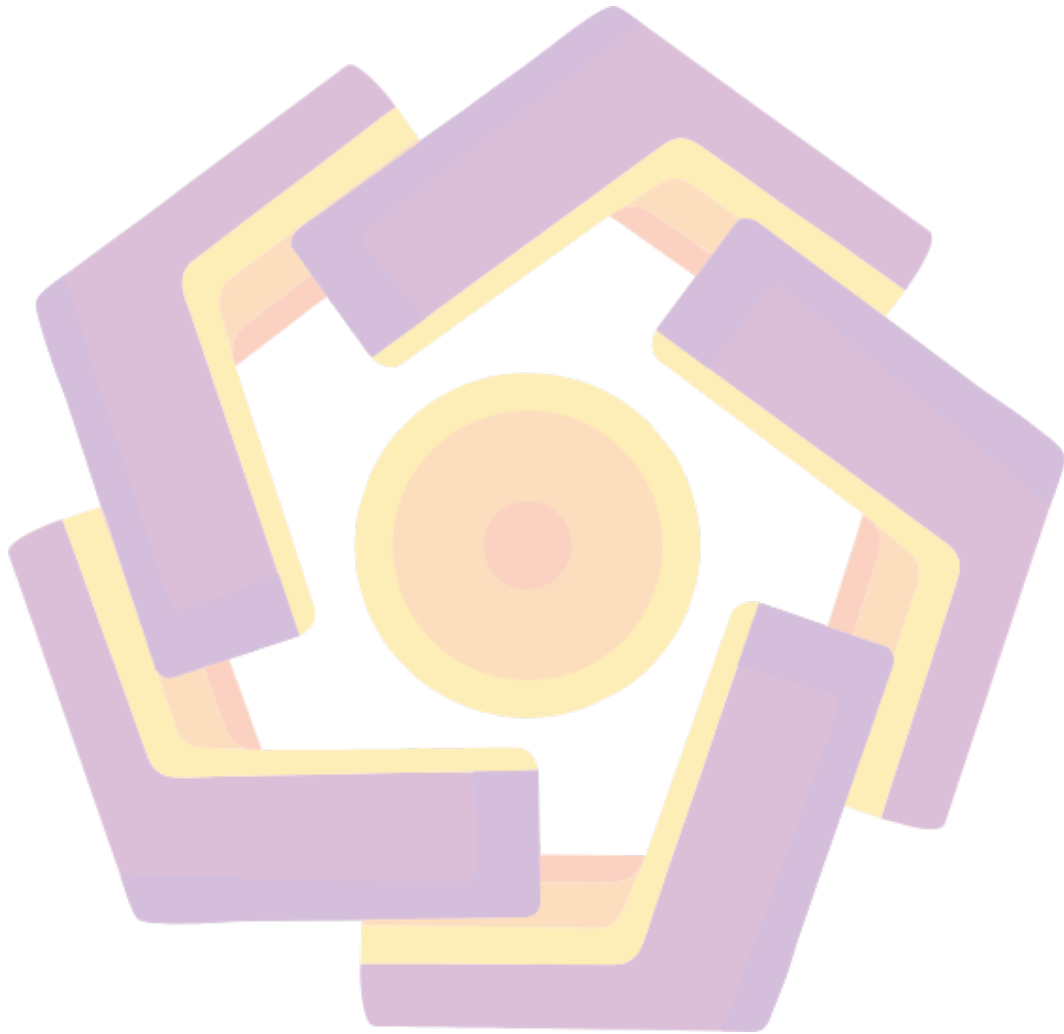
Penulis

DAFTAR ISI

HALAMAN JUDUL.....	i
HALAMAN PERSETUJUAN.....	ii
HALAMAN PENGESAHAN.....	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	iv
HALAMAN MOTTO	v
HALAMAN PERSEMBAHAN	vi
KATA PENGANTAR	vii
DAFTAR ISI.....	viii
DAFTAR TABEL.....	ix
DAFTAR GAMBAR	x
INTISARI.....	xi
ABSTRACT.....	xii
BAB I PENDAHULUAN	1
1.1 Latar Belakang Masalah	1
1.2 Rumusan Masalah.....	2
1.3 Batasan Masalah	2
1.4 Tujuan Penelitian	3
1.5 Sistematika Penulisan	3
BAB II LANDASAN TEORI	5
2.1. SQL Injection.....	5
2.2. Pengembangan Sistem Deteksi SQL Injection dengan Pembelajaran Mesin	7
2.3. Algoritma Random Forest dengan Pendekatan <i>Bag of Words</i>	10
BAB III METODOLOGI PENELITIAN.....	14
3.1. Metodologi Penelitian.....	14
3.2. Metode Penelitian	14
3.3. Metode Pengambilan Data.....	15
3.4. Metode Observasi	15
3.5. Metode Analisis	15
3.6. Metode Perancangan.....	16
BAB IV PEMBAHASAN.....	17
4.1 Perancangan	17
4.2 Implementasi Model	25
4.3 Pengujian Model	29
BAB V PENUTUP.....	34
5.1 Kesimpulan	34
5.2 Saran	35
DAFTAR PUSTAKA	36

DAFTAR TABEL

Tabel 4. 1. Tabel Ekstraksi Fitur.....	22
Tabel 4. 2. Lanjutan	23
Tabel 4. 3. Tabel Hasil Perhitungan Mesin.....	30
Tabel 4. 4. Tabel Hasil Percobaan setelah Modifikasi Parameter.....	30
Tabel 4. 5. Rangkuman Perbandingan Performa Model	31



DAFTAR GAMBAR

Gambar 3. 1. Tabel Alur Penelitian	14
Gambar 4. 1. Flowchart Proses Perancangan	17
Gambar 4. 2. Struktur dan tipe data pada kolom dataset	18
Gambar 4. 3. Kode Lowercasing.....	19
Gambar 4. 4. Kode Menghapus Null Values	19
Gambar 4. 5. Kode Menghapus Duplikasi Data	19
Gambar 4. 6. Sebelum dilakukan preprocessing awal	20
Gambar 4. 7. Setelah dilakukan preprocessing awal.....	20
Gambar 4. 8. Kata yang terindikasi serangan SQL Injection.....	21
Gambar 4. 9. Kata Normal	21
Gambar 4. 10. Persebaran fitur yang telah diekstraksi.....	24
Gambar 4. 11. Library Penting.....	25
Gambar 4. 12. Splitting dan Penyimpanan Data dengan Fungsi Pickle	26
Gambar 4. 13. Pemrosesan Kolom Sentence dengan BoW Unigram	26
Gambar 4. 14. Penyimpanan data yang telah diolah dengan BoW Unigram.....	27
Gambar 4. 15. Kode Proses Encoding no_single_qts	27
Gambar 4. 16. Kode Implementasi Model Random Forest	28
Gambar 4. 17. Confusion Matrix Random Forest.....	29
Gambar 4. 18. Hasil Confusion Matrix dari Percobaan Hyperparameter Tuning Percobaan Ketiga	31
Gambar 4. 19. Grafik Percobaan Cross validation	32

INTISARI

Seiring dengan meningkatnya kasus pembobolan data melalui teknik injeksi SQL (SQL Injection), puncaknya pada akhir Juni 2024 diduga data milik Kementerian Komunikasi dan Informatika telah bocor dan dibandrol dengan harga sekitar Rp. 1,9 miliar yang berisikan data pribadi meliputi NIK dan detail rekening perbankan yang hingga saat ini, belum ada langkah tegas dari Kominfo terkait permasalahan ini. Dalam upaya serangan SQL Injection ini, peneliti mengembangkan sistem deteksi SQL Injection menggunakan model Random Forest dengan preprocessing menggunakan *Bag of Words* serta *hyperparameter tuning*. Model ini dipilih karena kemampuannya dalam menangani data yang kompleks dan menghasilkan model yang kuat. Penelitian ini melibatkan serangkaian percobaan untuk menguji performa Random Forest dalam mendeteksi *query* SQL Injection dan *query* normal. Hasil penelitian menunjukkan bahwa Random Forest efektif dalam mendeteksi SQL Injection, dengan performa yang cukup baik setelah dilakukan modifikasi dari beberapa aspek *hyperparameter* seperti *n_estimators*, *max_depth*, *min_samples_split*, *min_samples_leaf*, *max_features*, dan *random_state* dengan menghasilkan akurasi sebesar 0.969, Presisi sebesar 1.0, F1-Score sebesar 0.952, dan *Recall* sebesar 0.908. Penyesuaian *hyperparameter* ini terbukti mampu mencegah *overfitting* dan meningkatkan akurasi model secara keseluruhan. Dengan demikian, dapat disimpulkan bahwa Random Forest dapat digunakan secara efektif untuk mendeteksi serangan SQL Injection melalui *query* yang muncul. Semoga dari penelitian ini dapat dikembangkan lagi dengan beberapa pendekatan-pendekatan canggih lainnya.

Kata kunci: Random Forest, Machine Learning, Query, SQL Injection, *Hyperparameter*

ABSTRACT

Along with the increasing cases of data breaches using SQL injection techniques, the peak at the end of June 2024 is that it is suspected that data belonging to the Ministry of Communication and Information has been leaked and is priced at around Rp. 1.9 billion containing personal data including NIK and banking account details. Until now, there has been no firm action from the Ministry of Communication and Information regarding this issue. In this SQL Injection attack effort, researchers developed a SQL Injection detection system using the Random Forest model with preprocessing using Bag of Words and hyperparameter tuning. This model was chosen because of its ability to handle complex data and produce a robust model. This research involves a series of experiments to test the performance of Random Forest in detecting SQL Injection queries and normal queries.. The research results show that Random Forest is effective in detecting SQL Injection, with quite good performance after modifying several hyperparameter aspects such as *n_estimators*, *max_depth*, *min_samples_split*, *min_samples_leaf*, *max_features*, and *random_state*, producing an accuracy of 0.969, Precision of 1.0, F1-Score of 0.952, and Recall of 0.908. In addition, this hyperparameter adjustment has been proven to be able to prevent overfitting and increase overall model accuracy. Thus, in this research it is concluded that Random Forest can be used effectively to detect SQL Injection attacks through the queries that appear. Hopefully this research can be further developed with several other sophisticated approaches.

Keywords: *Random Forest, Machine Learning, Query, SQL Injection, Hyperparameter*