PENERAPAN *STREAMING ANALYTICS* PADA DATA *HONEYPOT* UNTUK PENDETEKSIAN POLA SERANGAN BERBASIS APACHE FLINK

WIDYA AMELLIA PUTRI
2210511052

PROGRAM STUDI S1 INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN" JAKARTA
JAKARTA
2026

PENERAPAN *STREAMING ANALYTICS* PADA DATA *HONEYPOT* UNTUK PENDETEKSIAN POLA SERANGAN BERBASIS APACHE FLINK

WIDYA AMELLIA PUTRI
2210511052

Skripsi

sebagai salah satu syarat untuk memperoleh gelar

Sarjana Komputer

PROGRAM STUDI S1 INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN" JAKARTA
JAKARTA
2026

PERNYATAAN MENGENAI SKRIPSI DAN SUMBER INFORMASI SERTA PELIMPAHAN HAK CIPTA

PERNYATAAN MENGENAI SKRIPSI DAN SUMBER INFORMASI SERTA PELIMPAHAN HAK CIPTA

Dengan ini saya menyatakan bahwa skripsi dengan judul "PENERAPAN *STREAMING ANALYTICS* PADA DATA *HONEYPOT* UNTUK PENDETEKSIAN POLA SERANGAN BERBASIS APACHE FLINK" adalah karya saya dengan arahan dari dosen pembimbing dan belum diajukan dalam bentuk apa pun kepada perguruan tinggi mana pun. Sumber informasi yang berasal atau dikutip dari karya yang diterbitkan maupun tidak diterbitkan dari penulis lain telah disebutkan dalam teks dan dicantumkan dalam Daftar Pustaka di bagian akhir proposal ini.

Dengan ini saya melimpahkan hak cipta dari karya tulis saya kepada Universitas Pembangunan Nasional "Veteran" Jakarta.

Jakarta, Juli 2026

Widya Amellia Putri
2210511052

PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS

Sebagai civitas akademika Universitas Pembangunan Nasional "Veteran" Jakarta, saya yang bertanda tangan di bawah ini:

Nama : Widya Amellia Putri

NIM : 2210511052

Fakultas : Ilmu Komputer

Program Studi : S1 Informatika

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Pembangunan Nasional Veteran Jakarta Hak Bebas Royalti Non eksklusif *(Non - exclusive Royalty Free Right)* atas skripsi saya yang berjudul:

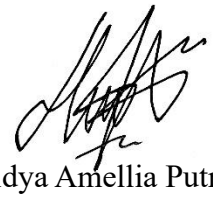
PENERAPAN *STREAMING ANALYTICS* PADA DATA *HONEYPOT* UNTUK PENDETEKSIAN POLA SERANGAN BERBASIS APACHE FLINK

Dengan Hak Bebas Royalti ini Universitas Pembangunan Nasional "Veteran" Jakarta berhak menyimpan, mengalih media atau memformatkan, mengelola dalam bentuk pangkalan data (basis data), merawat dan mempublikasikan skripsi saya selama tetap mencantumkan nama saya sebagai penulis dan sebagai pemilik Hak Cipta. Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di: Jakarta

Pada Tanggal: 17 Juli 2026

Yang menyatakan,



Widya Amellia Putri

LEMBAR PENGESAHAN

LEMBAR PENGESAHAN

Judul	:	PENERAPAN *STREAMING ANALYTICS* PADA DATA *HONEYPOT* UNTUK PENDETEKSIAN POLA SERANGAN BERBASIS APACHE FLINK
Nama	:	Widya Amellia Putri
NIM	:	2210511052
Program Studi	:	S1 Informatika

Disetujui oleh :

Penguji 1:
Dr. Indra Permana Solihin, S.Kom.,M.Kom.

Penguji 2:
Ichsan Mardani, S.Kom.,M.Sc.

Pembimbing 1:
I Wayan Widi Pradnyana, M.TI.

Pembimbing 2:
Neny Rosmawarni, S.Kom., M.Kom.

Diketahui oleh:

Koordinator Program Studi:
Dr. Ridwan Raafi'udin, S.Kom.,M.Kom.
NIP. 198805022021211001

Dekan Fakultas Ilmu Komputer:
Prof. Dr. Ir. Supriyanto, M.Sc., IPM.
NIP. 197605082003121002

Tanggal Ujian Tugas Akhir:
06 Juli 2026

ABSTRAK

Serangan siber terhadap layanan jaringan terus meningkat, sementara mekanisme pemantauan *honeypot* Cowrie di Laboratorium *Cyber Security and Networking* Fakultas Ilmu Komputer UPN "Veteran" Jakarta masih dilakukan secara *batch* dan manual, sehingga menimbulkan keterlambatan dalam mendeteksi dan merespons pola serangan yang sedang berlangsung. Kondisi ini mendorong kebutuhan akan sistem deteksi yang mampu memproses data *honeypot* secara kontinu dan otomatis. Penelitian ini bertujuan merancang dan mengimplementasikan sistem *streaming analytics* berbasis Apache Flink yang mengintegrasikan model klasifikasi *k-Nearest Neighbors* (kNN) dan *Logistic Regression* untuk mendeteksi aktivitas normal dan anomali pada data *honeypot* Cowrie, serta mengevaluasi efektivitas *(accuracy, precision, recall, F1-score)* dan efisiensi (latensi) kedua model tersebut. Penelitian menggunakan metode *experimental research* dengan pendekatan kuantitatif. Data *honeypot* Cowrie dikumpulkan dari basis data MongoDB laboratorium, kemudian melalui tahap *preprocessing* meliputi pelabelan, ekstraksi 19 fitur, penanganan ketidakseimbangan kelas dengan SMOTE, serta normalisasi. Kedua model dilatih dengan data *training* dan diuji dengan data *testing* (rasio 80:20), lalu di implementasikan melalui *pipeline* PyFlink yang terhubung dengan PostgreSQL sebagai penyimpanan hasil. Pengujian hipotesis dilakukan menggunakan analisis deskriptif terhadap latensi klasifikasi per *event*, dengan hasil dianalisis dan didiskusikan bersama dosen pembimbing untuk memastikan kesesuaian dengan kaidah ilmiah.

Kata Kunci : Apache Flink, *Honeypot* Cowrie, *k-Nearest Neighbors, Logistic Regression, streaming analytics*

ABSTRACT

Cyberattacks against network services continue to increase, while the monitoring mechanism for the Cowrie honeypot at the Cyber Security and Networking Laboratory of the Faculty of Computer Science, UPN "Veteran" Jakarta, is still performed through batch and manual processing, causing delays in detecting and responding to ongoing attack patterns. This condition drives the need for a detection system capable of processing honeypot data continuously and automatically. This research aims to design and implement an Apache Flink-based streaming analytics system that integrates k-Nearest Neighbors (kNN) and Logistic Regression classification models to detect normal and anomalous activities in Cowrie honeypot data, as well as to evaluate the effectiveness (accuracy, precision, recall, F1-score) and efficiency (latency) of both models. This research employs the experimental research method with a quantitative approach. Cowrie honeypot data is collected from the laboratory's MongoDB database, then processed through preprocessing stages including labeling, extraction of 19 features, class imbalance handling with SMOTE, and normalization. Both models were trained with training data and tested with testing data (80:20 ratio), then implemented through a PyFlink pipeline connected to PostgreSQL as the result storage. Hypothesis testing is conducted using descriptive analysis of per-event classification latency, with results analyzed and discussed together with the supervisor to ensure scientific validity.

Keywords: Apache Flink, Cowrie honeypot, k-Nearest Neighbors, Logistic Regression, streaming analytics

KATA PENGANTAR

Puji syukur penulis panjatkan ke hadirat Allah SWT atas segala limpahan rahmat, karunia, dan hidayah-Nya sehingga penulis dapat menyelesaikan penyusunan tugas akhir dalam bentuk skripsi yang berjudul "PENERAPAN *STREAMING ANALYTICS* PADA DATA *HONEYPOT* UNTUK PENDETEKSIAN POLA SERANGAN BERBASIS APACHE FLINK" sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer pada Program Studi S1-Informatika Fakultas Ilmu Komputer Universitas Pembangunan Nasional "Veteran" Jakarta. Pelaksanaan penelitian ini berlangsung sejak bulan November 2025 hingga April 2026. Penyusunan skripsi ini tentunya tidak terlepas dari dukungan, bantuan, serta doa dari berbagai pihak. Oleh karena itu, penulis dengan tulus menyampaikan ucapan terima kasih dan penghargaan setinggi-tingginya kepada:

1. Allah SWT, Tuhan Yang Maha Esa, atas segala limpahan rahmat, karunia, dan kekuatan yang diberikan selama proses penulisan skripsi ini. Segala hal yang penulis capai tidak akan mungkin terwujud tanpa kehendak dan pertolongan-Nya. Dalam setiap kesulitan, Allah memberikan jalan, dalam setiap kelelahan, Allah memberikan ketabahan. Segala puji bagi-Mu, yang selalu mendengarkan setiap doa dan harapan hamba-Mu.
2. Bapak Prof. Dr. Ir. Supriyanto, ST., M.Sc., IPM, selaku Dekan Fakultas Ilmu Komputer Universitas Pembangunan Nasional Veteran Jakarta. Terima kasih atas kepemimpinan yang bijaksana dan dorongan untuk selalu berinovasi di tengah tantangan zaman.
3. Bapak Dr. Ridwan Raafi'udin, S.Kom, M.Kom. selaku Ketua Program Studi S1 Informatika. Terima kasih atas dedikasi dan komitmen Bapak dalam mengembangkan kualitas pendidikan serta perhatian Bapak terhadap kemajuan setiap mahasiswa.
4. Bapak I Wayan Widi Pradnyana, M.TI. selaku Dosen Pembimbing I, yang telah meluangkan waktu, tenaga, dan pikiran untuk membimbing penulis dengan penuh kesabaran dan ketelitian. Terima kasih atas kritik membangun, motivasi yang tak pernah putus, serta arahan yang telah membawa penulis untuk terus berkembang dan memperbaiki diri dalam setiap proses penyusunan skripsi ini.
5. Ibu Neny Rosmawarni, S.Kom., M.Kom. selaku Dosen Pembimbing II, yang telah memberikan dukungan tambahan, masukan ilmiah, serta sudut pandang yang luas dalam menyempurnakan penelitian ini. Kontribusi Ibu sangat berarti dalam memperkaya hasil akhir Tugas Akhir ini.
6. Mamah Eka dan Ayah Fauzi, orang tua tercinta yang menjadi pilar hidup penulis. Terima kasih atas cinta yang tak terhingga, doa yang tak pernah terputus, dan pengorbanan tanpa pamrih yang telah mengiringi setiap langkah penulis sejak kecil hingga saat ini. Ayah dan Ibu adalah sumber

kekuatan terbesar dalam hidup penulis. Segala capaian ini adalah berkat kasih sayang dan keikhlasan kalian dalam mendidik dan membesarkan penulis dengan penuh cinta dan harapan.

7. Fatin Tsarwah Damia, Zea Nayla Khanza, dan Almh. Yurika Jasmien Nova, adik-adik yang luar biasa, terima kasih telah menjadi sosok yang selalu menghadirkan semangat dan energi positif di tengah kepenatan dan tekanan.
8. Teman-teman seperjuangan di Informatika UPN Veteran Jakarta angkatan 2022, yang telah menjadi bagian penting dalam perjalanan akademik ini. Terima kasih atas kebersamaan, kerja sama, tawa, dan dukungan yang tak ternilai selama masa perkuliahan.
9. Kedua sahabat terdekat, Firda Azzahra dan Efita Candra, yang selalu menjadi tempat berbagi keluh kesah dan penyemangat saat semangat mulai redup. Terima kasih telah hadir dengan segala keunikan dan ketulusan hati. Kebersamaan kita menjadi warna tersendiri dalam proses panjang yang penulis jalani. Kata-kata semangat, telinga yang selalu mendengar, serta waktu yang kalian berikan sangat berarti. Tanpa kalian, proses ini pasti terasa jauh lebih berat.
10. Untuk diri saya sendiri, terima kasih telah bertahan sejauh ini. Terima kasih karena tidak menyerah meski jalan terasa berat, karena terus melangkah meski kadang ragu, dan karena berani menghadapi setiap tantangan yang datang. Terima kasih telah percaya bahwa segala perjuangan ini akan membuahkan hasil. Setiap malam begadang, setiap lembar revisi, dan setiap titik lelah adalah bukti bahwa aku mampu melewati masa-masa sulit dan tumbuh menjadi pribadi yang lebih kuat. Semoga pencapaian ini menjadi awal dari langkah- langkah besar berikutnya dalam hidupku.

Penulis menyadari bahwa skripsi ini masih jauh dari kesempurnaan, sehingga penulis membuka diri terhadap segala saran dan kritik yang membangun demi penyempurnaan ke depannya. Semoga karya tulis ini dapat memberikan manfaat serta menjadi kontribusi positif dalam pengembangan ilmu pengetahuan, khususnya di bidang teknologi informasi.

Jakarta, 26 Juni 2026

Widya Amellia Putri
2210511052

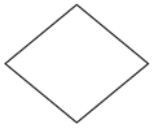
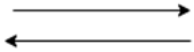
DAFTAR ISI

DAFTAR GAMBAR

DAFTAR TABEL

DAFTAR RUMUS

DAFTAR SIMBOL

No	Simbol	Nama	Penjelasan
1		*Terminator*	Simbol yang menandai titik awal proses maupun titik akhir proses dalam suatu alur kerja
2		*Process*	Simbol yang merepresentasikan suatu kegiatan yang dilaksanakan sebagai bagian dari alur kerja
3		*Decision*	Simbol yang melambangkan percabangan logika yang menghasilkan pilihan jalur berdasarkan kondisi tertentu.
4		*Flow*	Simbol yang menjelaskan arah hubungan yang menghubungkan satu simbol dengan simbol berikutnya dalam alur proses
5		*Input-Output*	Simbol yang menyatakan proses *input* atau *output*

DAFTAR LAMPIRAN