

KLASIFIKASI LALU LINTAS JARINGAN MENGGUNAKAN *MACHINE LEARNING* UNTUK ANALISIS TREND PENGGUNAAN INTERNET DI IPB

DHIANITA SHAFIA



**PROGRAM SARJANA ILMU KOMPUTER
SEKOLAH SAINS DATA, MATEMATIKA, DAN INFORMATIKA
INSTITUT PERTANIAN BOGOR
BOGOR
2025**



@Hak cipta milik IPB University

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



PERNYATAAN MENGENAI SKRIPSI DAN SUMBER INFORMASI SERTA PELIMPAHAN HAK CIPTA

Dengan ini saya menyatakan bahwa laporan akhir dengan judul “Klasifikasi Lalu Lintas Jaringan Menggunakan *Machine Learning* untuk Analisis Tren Penggunaan Internet di IPB” 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 skripsi ini.

Dengan ini saya melimpahkan hak cipta dari karya tulis saya kepada Institut Pertanian Bogor.

Bogor, Agustus 2025

Dhianita Shafa
G6401211016



ABSTRAK

DHIANITA SHAFa. Klasifikasi Lalu Lintas Jaringan Menggunakan *Machine Learning* untuk Analisis Tren Penggunaan Internet di IPB. Dibimbing oleh HERU SUKOCO dan MUSHTHOFA.

Internet merupakan bagian penting dalam mendukung aktivitas akademik dan non-akademik di lingkungan kampus. Namun, penggunaan internet yang berlebihan untuk aktivitas non-akademik dapat memengaruhi performa akademik. Penelitian ini bertujuan untuk mengklasifikasikan lalu lintas jaringan IPB menggunakan algoritma XGBoost serta menganalisis tren penggunaannya berdasarkan hasil klasifikasi tersebut. Data yang digunakan berupa log *firewall* Palo Alto yang dikumpulkan selama dua pekan pada jam sibuk. Data diproses melalui pembersihan, pengategorian ke dalam tujuh kategori trafik, rekayasa fitur, serta pembagian data untuk pelatihan, validasi, serta pengujian. Model XGBoost dibangun dengan *tuning* lima parameter utama, menghasilkan akurasi 94,89% dan F1-score 94,72%. Hasil klasifikasi menunjukkan bahwa trafik didominasi oleh kategori *Social Network* yang mencapai 39,04% dengan total volume 8 TB, diikuti oleh *Infrastructure* sebesar 24,14% atau sekitar 4,9 TB. Sementara itu, kategori *Education & Productivity* hanya menyumbang 6,33% dengan total volume 1,2 TB. Temuan ini menunjukkan adanya ketidakseimbangan penggunaan internet di lingkungan kampus, sehingga diperlukan pengelolaan *bandwidth* yang lebih terarah untuk mendukung prioritas akademik. Penelitian ini juga memberikan dasar bagi pengembangan sistem pemantauan jaringan berbasis *machine learning*.

Kata Kunci: klasifikasi lalu lintas, pengelolaan *bandwidth*, penggunaan internet, XGBoost

ABSTRACT

DHIANITA SHAFa. Network Traffic Classification Using Machine Learning for Internet Usage Trend Analysis at IPB University. Supervised by HERU SUKOCO and MUSHTHOFA.

The internet plays a vital role in supporting both academic and non-academic activities on campus. However, excessive use of the internet for non-academic purposes can negatively impact academic performance. This study aims to classify network traffic at IPB using the XGBoost algorithm and analyze usage trends based on the classification results. The data used consists of Palo Alto firewall logs collected over two weeks during peak hours. The data underwent cleaning, categorization into seven traffic categories, feature engineering, partitioning for training, validation, and testing. The XGBoost model was built by tuning five main parameters, resulting in an accuracy of 94.89% and an F1-score of 94.72%. The classification results show that traffic is dominated by the Social Network category, which accounts for 39.04% with a total volume of 8 TB, followed by Infrastructure at 24.14% or approximately 4.9 TB. Meanwhile, the Education & Productivity category only contributed 6.33% with a total volume of 1.2 TB. These findings indicate an imbalance in internet usage on campus, highlighting the need for more

targeted bandwidth management to support academic priorities. This study also provides a foundation for the development of machine learning-based network monitoring systems.

Keywords: bandwidth management, internet usage, traffic classification, XGBoost

@Hak cipta milik IPB University

IPB University





Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

© Hak Cipta milik IPB, tahun 2025
Hak Cipta dilindungi Undang-Undang

Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan atau menyebutkan sumbernya. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik, atau tinjauan suatu masalah, dan pengutipan tersebut tidak merugikan kepentingan IPB.

Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apa pun tanpa izin IPB.



@Hak cipta milik IPB University

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

KLASIFIKASI LALU LINTAS JARINGAN MENGGUNAKAN *MACHINE LEARNING* UNTUK ANALISIS TREND PENGUNAAN INTERNET DI IPB

DHIANITA SHAFIA

Skripsi
sebagai salah satu syarat untuk memperoleh gelar
Sarjana pada
Program Studi Ilmu Komputer

**PROGRAM SARJANA ILMU KOMPUTER
SEKOLAH SAINS DATA, MATEMATIKA, DAN INFORMATIKA
INSTITUT PERTANIAN BOGOR
BOGOR
2025**



@Hak cipta milik IPB University

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



Judul Skripsi : Klasifikasi Lalu Lintas Jaringan Menggunakan *Machine Learning*
untuk Analisis Tren Penggunaan Internet di IPB

Nama : Dhianita Shafa
NIM : G6401211016

Disetujui oleh

Pembimbing 1:
Dr. Heru Sukoco, S.Si., M.T.

Pembimbing 2:
Dr. Mushthofa, S.Kom., M.Sc.

Diketahui oleh

Ketua Program Sarjana Ilmu Komputer:
Dr. Sony Hartono Wijaya, S.Kom., M.Kom.
19810809 200812 1 002

Tanggal Ujian:
31 Juli 2025

Tanggal Lulus:

PRAKATA

Puji dan syukur penulis panjatkan kepada Allah subhanaahu wa ta'ala atas segala karunia-Nya sehingga karya ilmiah ini berhasil diselesaikan. Tema yang dipilih dalam penelitian yang dilaksanakan sejak bulan Desember 2024 sampai bulan Juli 2025 ini ialah karakteristik lalu lintas internet dengan judul “Klasifikasi Lalu Lintas Jaringan Menggunakan *Machine Learning* untuk Analisis Tren Penggunaan Internet di IPB”.

Penelitian ini dapat terselesaikan dengan baik berkat bantuan dan dukungan dari berbagai pihak. Untuk itu, penulis menyampaikan ucapan terima kasih yang sebesar-besarnya kepada:

- a Ayah, Ibu, serta kedua kakak penulis atas segala doa, cinta, dukungan, dan kasih sayang yang senantiasa diberikan selama masa perkuliahan hingga proses penelitian dan penulisan skripsi,
- b Bapak Dr. Heru Sukoco, S.Si., M.T. dan Bapak Dr. Mushthofa, S.Kom., M.Sc. selaku dosen pembimbing skripsi yang telah memberikan bimbingan, arahan, serta motivasi kepada penulis dalam menyelesaikan penelitian ini,
- c Mas Khafi dan Mas Imam selaku administrator jaringan IPB yang telah membantu penulis dalam proses pengambilan data,
- d Althaf, Eva, Nabila, Novia, Sanur, Illinia, dan Khansa yang telah menjadi teman seperjuangan dan senantiasa menemani penulis selama masa perkuliahan hingga penyusunan skripsi,
- e Fathiya, Zahra, Mirza, Dharma, Naufal, Bang Jundi, tim Big Brain Kidz, serta seluruh teman-teman penulis yang tidak dapat disebutkan satu per satu, atas segala semangat, bantuan, dan motivasi yang diberikan,

Penulis berharap karya ilmiah ini dapat memberikan manfaat bagi semua pihak yang membutuhkan serta turut berkontribusi dalam pengembangan ilmu pengetahuan.

Bogor, Agustus 2025

Dhianita Shafa



DAFTAR ISI

DAFTAR TABEL	xi
DAFTAR GAMBAR	xi
DAFTAR LAMPIRAN	xi
PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Tujuan	2
1.4 Manfaat Penelitian	3
1.5 Ruang Lingkup	3
TINJAUAN PUSTAKA	4
2.1 Paket Jaringan	4
2.2 Log Lalu Lintas Jaringan	5
2.3 Klasifikasi Lalu Lintas Jaringan	6
2.4 Algoritma XGBoost	9
METODE PENELITIAN	11
3.1 Tahapan Penelitian	11
3.2 Data Penelitian	19
3.3 Lingkungan Pengembangan	20
HASIL DAN PEMBAHASAN	21
4.1 Segmentasi IP Jaringan	21
4.2 Penentuan Jam Sibuk	22
4.3 Akuisisi Data	23
4.4 Pembersihan Data	24
4.5 Kategorisasi	24
4.6 Pengambilan Sampel	26
4.7 Rekayasa Fitur	27
4.8 Pembagian Data	27
4.9 Seleksi Fitur	28
4.10 <i>Hyperparameter Tuning</i>	29
4.11 Evaluasi Model	30
4.12 Analisis Interpretabilitas Model	31
4.13 Analisis Tren Penggunaan Internet	35
SIMPULAN DAN SARAN	42
5.1 Simpulan	42
5.2 Saran	42
DAFTAR PUSTAKA	43
LAMPIRAN	45
RIWAYAT HIDUP	48



DAFTAR TABEL

1	Kategori aplikasi berdasarkan nomor <i>port</i>	7
2	Parameter dan nilai grid yang diujikan	17
3	Fitur log lalu lintas jaringan Palo Alto	19
4	Segmentasi alamat IP pada jaringan IPB	22
5	Rata-rata <i>throughput</i> jaringan per jam selama periode pengamatan	22
6	Atribut log jaringan yang digunakan dalam pelatihan model	24
7	Kategorisasi konten berdasarkan aplikasi	25
8	Fitur baru yang ditambahkan	27
9	Jumlah sesi lalu lintas berdasarkan jenis pengguna	35
10	Jumlah sesi lalu lintas berdasarkan unit kerja	36
11	Volume lalu lintas berdasarkan jenis pengguna	38
12	Volume lalu lintas berdasarkan unit kerja	39

DAFTAR GAMBAR

1	Struktur <i>field</i> pada header IPv4 (sumber RFC 791)	4
2	Tahapan penelitian	11
3	Diagram alir penentuan metode pengukuran intensitas trafik (Sumber rekomendasi ITU-T E.500)	12
4	Topologi jaringan IPB secara umum (sumber IPB 2025)	21
5	Diagram lingkaran distribusi kelas pada data sampel	26
6	Peringkat fitur berdasarkan kontribusi terhadap model	28
7	Jumlah fitur optimal berdasarkan <i>f1-score</i> dan kompleksitas	29
8	<i>Confusion matrix</i> yang telah dinormalisasi	31
9	Plot KDE untuk distribusi fitur <i>average packet size</i>	31
10	Plot KDE untuk distribusi fitur <i>packets sent</i>	32
11	Plot KDE untuk distribusi fitur <i>packets</i>	33
12	Diagram batang untuk distribusi fitur IP <i>protocol</i>	34
13	<i>Heatmap</i> untuk fitur <i>destination port</i> dengan distribusi terbanyak	34
14	Diagram lingkaran distribusi jumlah sesi per kelompok pengguna	36
15	Diagram lingkaran distribusi jumlah sesi per kelompok unit kerja	37
16	Diagram batang total sesi penggunaan internet di IPB	38
17	Diagram lingkaran distribusi volume trafik per kelompok pengguna	39
18	Diagram lingkaran distribusi volume trafik per kelompok unit kerja	40
19	Diagram batang total volume penggunaan internet di IPB	41

DAFTAR LAMPIRAN

1	Tren <i>throughput</i> jaringan IPB tanggal 7-13 Februari 2025	46
2	Tren <i>throughput</i> jaringan IPB tanggal 2-8 Mei 2025	47