

**KLASIFIKASI *TRAFFIC NETWORK* DENGAN  
MENGUNAKAN *NAÏVE BAYES* DAN *FEATURE SELECTION*  
DALAM PEMILIHAN ATRIBUT**

*Diajukan Sebagai Syarat Untuk Menyelesaikan  
Pendidikan Program Strata-1 Pada  
Jurusan Teknik Informatika Fakultas Ilmu Komputer UNSRI*



Oleh :

Alfin Ramdhani  
NIM : 09021381419104

**Jurusan Teknik Informatika  
FAKULTAS ILMU KOMPUTER UNIVERSITAS SRIWIJAYA  
2019**

## LEMBAR PENGESAHAN TUGAS AKHIR

Klasifikasi Traffic Network dengan Menggunakan Naive Bayes dan Feature Selection Dalam Pemilihan Atribut

Oleh :

ALFIN RAMDHANI  
NIM : 09021381419104

Palembang, Juli 2019

Pembimbing I,

Pembimbing II,



Rifkie Primartha, M.T.  
NIP. 197706012009121004



Yopy Sazaki, M.T.  
NIPUS. 197406062012101201

Mengetahui,  
Ketua Jurusan Teknik Informatika



Rifkie Primartha, M.T.  
NIP. 197706012009121004

## TANDA LULUS UJIAN SIDANG TUGAS AKHIR

Pada hari Kamis, 18 Juli 2019 telah dilaksanakan ujian sidang tugas akhir oleh Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

Nama : Alfin Ramdhani  
NIM : 09021381419104  
Judul : Klasifikasi Network Traffic dengan Menggunakan Naive Bayes dan Feature Selection Dalam Pemilihan Atribut

1. Pembimbing I

Rifkie Primartha, M.T.  
NIP. 197706012009121004

2. Pembimbing II

Yopy Sazaki, M.T.  
NIPUS. 197406062012101201

3. Penguji I

Hardini Novianti, M.T.  
NIP. 197911012014042002

4. Penguji II

Nabila Rizky Oktadini, S.SI, M.T.  
NIP. 199110102018032001

Mengetahui,  
Ketua Jurusan Teknik Informatika,



Rifkie Primartha, M.T.  
NIP. 197706012009121004

## HALAMAN PERNYATAAN

Yang bertanda tangan di bawah ini :

Nama : Alfin Ramdhani  
NIM : 09021381419104  
Program Studi : Teknik Informatika Bilingual  
Judul Skripsi : Klasifikasi *Traffic Network* dengan Menggunakan *Naive Bayes* dan *Feature Selection* Dalam Pemilihan Atribut  
Hasil Pengecekan Software *iThenticate/Turnitin* : 15%

Menyatakan bahwa Laporan Projek saya merupakan hasil karya sendiri dan bukan hasil penjiplakan/plagiat. Apabila ditemukan unsur penjiplakan/plagiat dalam laporan projek ini, maka saya bersedia menerima sanksi akademik dari Universitas Sriwijaya sesuai dengan ketentuan yang berlaku.

Demikian, pernyataan ini saya buat dengan sebenarnya dan tidak ada paksaan oleh siapapun.

Palembang, Juli 2019



(Alfin Ramdhani)

NIM. 09021381419104

## MOTTO DAN PERSEMBAHAN

“Ketika ada sesuatu yang cukup penting, anda harus melakukannya bahkan jika peluangnya tidak menguntungkanmu.”  
(Elon Musk, CEO Tesla & SpaceX)

Kupersembahkan karya tulis ini kepada :

- Allah SWT
- Kedua Orang Tua Tercinta
- Saudari Tercinta
- Keluarga Besar Tercinta
- Dosen Pembimbing dan Penguji
- Sahabat-sahabatku
- Almamaterku

# **CLASSIFICATION OF NETWORK TRAFFIC USING NAÏVE BAYES AND FEATURE SELECTION IN SELECTION OF ATTRIBUTES**

**By :**

**Alfin Ramdhani**

**09021381419104**

## **ABSTRACT**

Now the number of users of Internet services is increasing. So, the more traffic on the internet is also congested. How to get the utilization traffic pattern is through the traffic classification process. Because the volume of traffic log data is very large and always increasing rapidly, it requires an effective and simple method to be applied in the classification process. So the Naive Bayes method is chosen, which is pretty much applied to calculating the probability level. And using Feature Selection in the selection of attributes that appears. This study uses internet traffic data with 248 attributes. In this research, the results of the Naive Bayes classification based on class with the highest value of accuracy, precision, and recall in the WWW class are (0.994, 0.91, 0.994) and the results of the feature selection using the ranking feature method are the highest entropy value 0.86914856 for attribute 4. The highest value attribute validation on attribute 35 is 97.2942%. The results obtained from this study are expected to be useful for decision makers for managing the internet in the future.

**Keywords :** *Feature Selection, Internet Traffic, Naive Bayes, Classification.*

# **KLASIFIKASI TRAFFIC NETWORK DENGAN MENGGUNAKAN NAIVE BAYES DAN FEATURE SELECTION DALAM PEMILIHAN ATRIBUT**

**Oleh :**

**Alfin Ramdhani**

**09021381419104**

## **ABSTRAK**

Semakin meningkat jumlah pengguna layanan Internet. Maka, semakin padat pula trafik pada internet. Untuk mendapatkan pola trafik pemanfaatan adalah melalui proses klasifikasi trafik. Karena volume data log trafik yang sangat besar dan selalu bertambah dengan cepat, maka dibutuhkan metode yang efektif dan sederhana untuk diterapkan dalam proses klasifikasi. Sehingga dipilih metode Naive Bayes, yang cukup banyak diterapkan untuk menghitung tingkat probabilitas. Serta menggunakan Feature Selection dalam pemilihan atribut yang banyak muncul. Penelitian ini menggunakan data trafik internet dengan 248 atribut. Pada penelitian ini didapat hasil dari klasifikasi Naive Bayes berdasarkan class dengan nilai akurasi, *precision*, dan *recall* tertinggi pada class WWW yaitu (0,994, 0,91, 0,994) dan hasil dari *feature selection* dengan menggunakan metode *feature ranking* didapatkan nilai *entropy* tertinggi adalah 0.86914856 untuk atribut 4. Untuk validasi atribut nilai tertinggi pada attribute 35 sebesar 97.2942 %. Hasil yang didapatkan dari penelitian ini diharapkan dapat berguna bagi para pengambil keputusan untuk pengelolaan internet pada masa yang akan datang.

Kata kunci: *Feature Selection*, trafik internet, *Naive Bayes*, klasifikasi.

## KATA PENGANTAR



Puji syukur kepada Allah SWT atas berkat dan rahmat-Nya yang telah diberikan kepada Penulis sehingga dapat menyelesaikan Tugas Akhir ini dengan baik. Tugas akhir ini disusun untuk memenuhi salah satu syarat guna menyelesaikan pendidikan program Strata-1 pada Fakultas Ilmu Komputer Program Studi Teknik Informatika di Universitas Sriwijaya.

Dalam menyelesaikan Tugas Akhir ini banyak pihak yang telah memberikan bantuan dan dukungan baik secara langsung maupun secara tidak langsung. Penulis ingin menyampaikan rasa terima kasih kepada:

1. Orang tuaku, H. Anwar Wan Saman (alm) dan Hj. Lisnurleni, S.Pd.SD., Saudari peremuanku, Astri Mareta S.Pd, dan seluruh keluarga besarku yang selalu mendoakan, motivasi, menasehati, serta memberikan dukungan baik moril maupun materil.
2. Bapak Jaidan Jauhari, S.Pd., M.T. selaku Dekan Fakultas Ilmu Komputer Universitas Sriwijaya.
3. Bapak Rifkie Primartha, M.T. selaku Ketua Jurusan Teknik Informatika Universitas Sriwijaya.
4. Bapak Rifkie Primartha, M.T. selaku dosen pembimbing I dan Bapak Yoppy Sazaki, M.T. selaku dosen pembimbing II yang telah membimbing, mengarahkan dan memberikan motivasi dalam proses perkuliahan dan pengerjaan Tugas Akhir.



5. Dr. Syed Othmawi Bin Abd. Rahman, selaku supervisor yang telah membimbing penulis dalam menyelesaikan Projek Sarjana Muda (PSM 1) di Universiti Teknologi Malaysia.
6. Bapak Tasmi Salim, S.Si., M.Kom. selaku dosen yang memberi saran topik penelitian dibidang networking yang dipilih oleh penulis serta membantu saat pemahaman garis besar teori dan hal-hal lain yang sangat berguna dari penelitian ini.
7. Ibu Hardini Novianti, M.T. selaku dosen penguji I dan Ibu Nabila Rizky Oktadini, S.Si., M.T. selaku dosen penguji II yang telah memberikan masukan dan ilmu pengetahuan kepada penulis.
8. Seluruh Bapak dan Ibu dosen Program Studi Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.
9. Seluruh staf Tata Usaha yang telah membantu dalam kelancaran proses administrasi dan akademik selama masa perkuliahan.
10. Teman-teman IF Reguler dan Bilingual, yang selalu berjuang bersama dalam menempuh ilmu, terutama Ihsan Muhdlari, M Imam Nurrahman, M Aznen Hamdiyah, Fahrur Nabil Arighi, Ricardo, M Satrio Wijaya, Joko Prawibowo, Yoga Anugrah Pratama, Margono Saftian, Deo Wicaksono, Sugi Pratama, Abdul Hamid Zulni, Redha Bayu Anggara, Yesica Viania Variska, Winda Agusthia, Shofi Salsabila, Puri Indah Lestari, Anggita Dewintiara Laupati, serta Lia Septiani.
11. Septrianto Nugroho, M Wahyu Sanjaya, Arief Yanda Budiman, Ilham Akbar, Andila Tiara Putri, Amrina Ilma Rosyada, Umi Madestia Lugan, Tiara Windri Apriani, Tri Anggun Setya, Filda Kurnia Barasky, serta

Elsya R. Gustiarini. Sebagai senior di jurusan Informatika dan teman dekat dari penulis yang telah menemani kehidupan penulis, tempat berbagi cerita, serta menjadi pendengar setia dalam keluh kesah penulis.

12. Semua pihak yang tidak dapat penulis sebutkan satu-persatu yang telah banyak membantu dan berperan bagi penulis terutama dalam penyelesaian tugas akhir ini, terima kasih banyak atas semuanya.

Penulis menyadari dalam penyusunan Tugas Akhir ini masih terdapat banyak kekurangan disebabkan keterbatasan pengetahuan dan pengalaman, oleh karena itu kritik dan saran yang membangun sangat diharapkan untuk kemajuan penelitian selanjutnya. Akhir kata semoga Tugas Akhir ini dapat berguna dan bermanfaat bagi kita semua.

Palembang, Juli 2019

Penulis

## DAFTAR ISI

	<b>Halaman</b>
<b>HALAMAN JUDUL .....</b>	<b>i</b>
<b>HALAMAN PENGESAHAN.....</b>	<b>ii</b>
<b>HALAMAN TANDA LULUS UJIAN SIDANG TUGAS AKHIR.....</b>	<b>iii</b>
<b>HALAMAN PERNYATAAN.....</b>	<b>iv</b>
<b>MOTTO DAN PERSEMBAHAN.....</b>	<b>v</b>
<b>ABSTRACT .....</b>	<b>vi</b>
<b>ABSTRAK .....</b>	<b>vii</b>
<b>KATA PENGANTAR.....</b>	<b>viii</b>
<b>DAFTAR ISI .....</b>	<b>xi</b>
<b>DAFTAR GAMBAR .....</b>	<b>xv</b>
<b>DAFTAR TABEL .....</b>	<b>xvii</b>
<b>DAFTAR LAMPIRAN .....</b>	<b>xix</b>

### **BAB 1 PENDAHULUAN**

1.1 Pendahuluan .....	I-1
1.2 Latar Belakang .....	I-1
1.3 Rumusan Masalah .....	I-4
1.4 Tujuan Penelitian .....	I-4
1.5 Manfaat Penelitian .....	I-4
1.6 Batasan Masalah .....	I-5
1.7 Sistematika Penulisan .....	I-6
1.8 Kesimpulan .....	I-7

## **BAB II TINJAUAN PUSTAKA**

2.1	Pendahuluan .....	II-1
2.2	Klasifikasi Trafik .....	II-1
2.2.1	Naive Bayes .....	II-2
2.3	Feature Selection Untuk Pemilihan Atribut .....	II-4
2.4	<i>Cross Validation</i> .....	II-5
2.5	Confusion Matrix .....	II-5
2.6	Penelitian Lain Yang Relevan .....	II-7
2.6.1	Li Jun, Zhang Shunyi, Lu Yanqing, Zhang Zailong (2007).....	II-7
2.6.2	K. Keerthi Vasan dan B. Surendiran (2016) .....	II-8
2.7	Kesimpulan .....	II-8

## **BAB III METODOLOGI PENELITIAN**

3.1	Pendahuluan .....	III-1
3.2	Unit Penelitian .....	III-1
3.3	Data .....	III-1
3.3.1	Jenis dan Sumber Data .....	III-1
3.3.2	Metode Pengumpulan Data .....	III-2
3.4	Tahapan Penelitian .....	III-2
3.4.1	Menetapkan Kerangka Kerja .....	III-2
3.4.2	Menetapkan Kriteria Pengujian .....	III-5
3.4.3	Menetapkan Format Data Pengujian .....	III-5
3.4.4	Menentukan Alat yang Digunakan Dalam Pelaksanaan Penelitian .....	III-6
3.4.5	Melakukan Pengujian Penelitian .....	III-7
3.4.6	Melakukan Analisa Hasil Pengujian Dan Membuat Kesimpulan .....	III-8
3.5	Metode Pengembangan Perangkat Lunak .....	III-8
3.5.1	<i>Rational Unified Process</i> .....	III-8
3.5.2	Fase Insepsi .....	III-10
3.5.3	Fase Elaborasi.....	III-10
3.5.4	Fase Konstruksi .....	III-11
3.5.5	Fase Transisi .....	III-11
3.6	Manajemen Proyek Penelitian.....	III-12
3.7	Kesimpulan .....	III-23

## BAB IV PENGEMBANGAN PERANGKAT LUNAK

4.1	Pendahuluan .....	IV-1
4.2	Fase Insepsi .....	IV-1
4.2.1	Pemodelan Bisnis .....	IV-2
4.2.2	Kebutuhan Sistem .....	IV-3
4.2.2.1	Fitur Prapengolahan.....	IV-3
4.2.2.2	Fitur Klasifikasi .....	IV-4
4.2.2.3	Fitur Seleksi .....	IV-4
4.2.2.4	Fitur Evaluasi dan Validasi .....	IV-4
4.2.3	Analisis dan Desain .....	IV-5
4.2.3.1	Analisis Kebutuhan Perangkat Lunak .....	IV-6
4.2.3.2	Analisis Prapengolahan .....	IV-7
4.2.3.3	Desain Perangkat Lunak .....	IV-8
	1. Model <i>Use Case</i> .....	IV-8
	2. Diagram Aktivitas .....	IV-13
4.3	Fase Elaborasi .....	IV-15
4.3.1	Pemodelan Bisnis .....	IV-16
4.3.1.1	Perancangan Data .....	IV-16
4.3.1.2	Perancangan Antarmuka .....	IV-16
4.3.2	Kebutuhan Sistem .....	IV-18
4.3.3	Diagram <i>Sequence</i> .....	IV-19
4.4	Fase Konstruksi .....	IV-22
4.4.1	Kebutuhan Sistem .....	IV-22
4.4.2	Diagram Kelas .....	IV-22
4.4.3	Implementasi .....	IV-23
4.4.3.1	Implementasi Kelas .....	IV-23
4.4.3.2	Implementasi Antarmuka .....	IV-25
4.5	Fase Transisi .....	IV-27
4.5.1	Permodelan Bisnis .....	IV-27
4.5.2	Kebutuhan Sistem .....	IV-27
4.5.3	Rencana Pengujian .....	IV-28
4.5.3.1	Rencana Use Case Melakukan Praproses Data Trafik Internet .....	IV-28
4.5.3.2	Rencana Melakukan Klasifikasi Dengan Naive Bayes .....	IV-29
4.5.4	Implementasi .....	IV-30
4.5.4.1	Pengujian <i>Use Case</i> Memasukan Data Trafik.....	IV-30
4.5.4.2	Pengujian Use Case Melakukan Klasifikasi Dengan <i>Naive Bayes</i> .....	IV-31

4.6 Kesimpulan .....	IV-33
----------------------	-------

## **BAB V HASIL DAN ANALISIS PENELITIAN**

5.1 Pendahuluan .....	V-1
5.2 Percobaan Penelitian .....	V-1
5.3 Hasil Feature Extraction.....	V-2
5.4 Hasil Feature Selection.....	V-2
5.5 Hasil Klasifikasi Dengan Naive Bayes .....	V-6
5.6 Analisis Hasil Penelitian .....	V-8
5.7 Kesimpulan .....	V-8

## **BAB VI KESIMPULAN DAN SARAN**

6.1 Pendahuluan .....	VI-1
6.2 Kesimpulan .....	VI-1
6.3 Saran .....	VI-2

<b>DAFTAR PUSTAKA .....</b>	<b>xx</b>
-----------------------------	-----------

## DAFTAR GAMBAR

	<b>Halaman</b>
Gambar II-1. Proses Klasifikasi Trafik .....	II-2
Gambar III-1. Rancangan Arsitektur atribut TCP Header. ....	III-3
Gambar III-2. Tahapan Pengujian Penelitian.....	III-7
Gambar III-3. Struktur Proses RUP .....	III-9
Gambar IV-1. Diagram <i>Use Case Current Existing</i> .....	IV-2
Gambar IV-2. Diagram <i>Use Case</i> .....	IV-9
Gambar IV-3. Diagram Aktivitas <i>Use Case</i> Melakukan Praproses Data Trafik Internet .....	IV-14
Gambar IV-4. Diagram Aktivitas <i>Use Case</i> Melakukan Klasifikasi dengan Naive Bayes dan Feature Selection.....	IV-15
Gambar IV-5. Rancangan Antarmuka Menu Utama .....	IV-17
Gambar IV-6. Rancangan Antarmuka <i>File Preview</i> .....	IV-17
Gambar IV-7. Rancangan Antarmuka <i>Output View</i> .....	IV-18
Gambar IV-8. <i>Sequence</i> Diagram Melakukan Praproses Data Trafik .....	IV-20
Gambar IV-9. <i>Sequence</i> Diagram Melakukan Klasifikasi Dengan NB .....	IV-21
Gambar IV-10. Diagram Kelas Perangkat Lunak .....	IV-23
Gambar IV-11. Antarmuka Halaman Menu Utama .....	IV-26
Gambar IV-12. Antarmuka <i>File Preview</i> .....	IV-26
Gambar IV-12. Antarmuka <i>Output View</i> .....	IV-27
Gambar V-1. Tampilan <i>raw data</i> berekstensi <i>.arff</i> .....	V-2

Gambar V-2. Tampilan data trafik berekstensi <i>.xls</i> .....	V-3
Gambar V-3. Diagram Garis Hasil Ranging dan Nilai Entropy Atribut.....	V-5
Gambar V-4. Diagram Garis Hasil Validasi Atribut dengan <i>Naive Bayes</i> .....	V-6
Gambar V-5. Gambar Confusion Matrix Hasil Penelitian.....	V-9



## DAFTAR TABEL

	<b>Halaman</b>
Tabel II-1. Contoh Tabel Confusion Matrix .....	II-6
Tabel III-1. Rancangan Tabel Confusion Matrix Setiap Hasil Pengujian ...	III-14
Tabel III-2. Rancangan Tabel Hasil Pengujian Setiap Hasil Klasifikasi .....	III-20
Tabel III-4. Tabel Penjadwalan Penelitian dalam Bentuk WBS .....	III-12
Tabel IV-1. Kebutuhan Fungsional .....	IV-5
Tabel IV-2. Kebutuhan Non Fungsional .....	IV-5
Tabel IV-3. Contoh Data Trafik Internet .....	IV-7
Tabel IV-4. Hasil Prapengolahan .....	IV-8
Tabel IV-5. Definisi Aktor <i>Use Case</i> .....	IV-10
Tabel IV-6. Definisi <i>Use Case</i> .....	IV-11
Tabel IV-7. Skenario <i>Use Case</i> Melakukan Klasifikasi dengan <i>Naive Bayes</i> dan <i>Feature Selection</i> . .....	IV-12
Tabel IV-8. Implementasi Kelas .....	IV-24
Tabel IV-9. Rencana Pengujian <i>Use Case</i> Melakukan Praproses Data. ....	IV-29
Tabel IV-10. Rencana Pengujian <i>Use Case</i> Melakukan Klasifikasi dengan Naive Bayes. ....	IV-29
Tabel IV-11. Pengujian <i>Use Case</i> Melakukan Praproses Data Internet.....	IV-30
Tabel IV-12. Pengujian <i>Use Case</i> Melakukan Klasifikasi dengan Naive Bayes dan <i>Feature Selection</i> .....	IV-38

Tabel V-1. Hasil Ranging Atribut dengan menggunakan IG.....	V-3
Tabel V-2. Hasil Training dan Testing .....	V-5
Tabel V-3. Tabel Hasil Klasifikasi.....	V-5

## DAFTAR LAMPIRAN

### Halaman

LAMPIRAN 1 Nama Atribut .....	L1-1
LAMPIRAN 2 Perhitungan Manual <i>Naive Bayes</i> .....	L2-1
LAMPIRAN 3 Kode Program .....	L3-1

# **BAB I**

## **PENDAHULUAN**

### **1.1 Pendahuluan**

Bab ini membahas mengenai latar belakang penelitian Klasifikasi *Traffic Network* dengan Menggunakan *Naïve Bayes* dan *Feature Selection* dalam pemilihan atribut yang akan dibahas secara umum dan singkat pada latar belakang.

### **1.2 Latar Belakang**

Kebutuhan koneksi internet semakin hari semakin meningkat, trafik internet pun meningkat. Dengan trafik yang semakin tinggi, maka akses/koneksi internet akan semakin berat/lambat. Sehingga perlu diketahui bagaimana pola trafik internet yang ada selama ini. Pola tersebut berguna untuk dijadikan dasar kebijakan manajemen koneksi internet untuk saat sekarang dan diwaktu yang akan datang, bermanfaat juga untuk mengetahui ada tidaknya pola yang tidak wajar yang bisa jadi mengarah ke serangan dari luar yang semakin membebani jaringan dalam aspek keamanan (Manshaei et al., 2013). Selain itu, pola yang didapatkan bisa menunjukkan aktifitas pengguna sehari-hari seperti apa, yaitu aplikasi internet apa saja yang mayoritas dimanfaatkan oleh pengguna selama ini. Hal tersebut berkaitan dengan tujuan utama dan prioritas dari ketersediaan internet. Sehingga jangan sampai, internet lebih banyak dimanfaatkan untuk hal-hal di luar tujuan utamanya (Sigit, 2016).

Jadi, dengan adanya deskripsi permasalahan diatas, maka kita harus mengklasifikasikan lalu lintas jaringan, untuk memonitor kapan lalu lintas sedang sibuk dan ketika lalu lintas sedikit aktivitas. Lihat jenis kegiatan apa yang sering dilakukan seseorang ketika terhubung ke internet dengan *ranking system* (*browsing, chatting, videocalling, streaming, dll*). Serta mengatasi *bottle-necking* atau pengiriman data yang tertunda beberapa saat. Salah satu metode yang akan digunakan dalam pengklasifikasian adalah metode *Naive Bayes*. *Naive Bayes* adalah bagian dari Jaringan Syaraf Tiruan. Klasifikasi *Naive Bayes* akan mengklasifikasikan jumlah catatan trafik jaringan, dan juga menggunakan kombinasi metode *Feature Selection* dalam penyeleksian atribut pada data trafik jaringan.

*Naive Bayes* (NB) adalah salah satu metode paling awal yang digunakan untuk klasifikasi trafik internet, yang sederhana dan cukup efektif untuk mengklasifikasi peluang (Sigit, 2016). Serta karena performa dan kecepatan yang tinggi dalam proses klasifikasi, dan mudah untuk menghasilkan probabilitas posterior data yang di tes terhadap kelasnya (Schlosser et al., 2009). Sedangkan *Feature Selection* adalah metode tambahan yang terdiri dari *Entropy* dan *Information Gain*. Kedua fitur tersebut memungkinkan penulis dalam proses penyeleksian data trafik yang digunakan seperti, pengurangan atribut yang tidak banyak mendukung dalam penelitian ini, mendapatkan nilai bobot dari setiap atribut, dan pemeringkatan atribut yang banyak muncul di data trafik.

Penelitian sebelumnya yang dilakukan oleh (Okililas & Tasmi, 2017) menggunakan metode DPI dalam mengelompokkan jenis trafik, sedang penelitian

yang dilakukan oleh (Bujlow, Carela-Español, & Barlet-Ros, 2015) membandingkan tool DPI dalam mengelompokkan jenis trafik internet.

Solusi dalam mengklasifikasi trafik pada network telah banyak dilakukan dengan menghasilkan solusi yang aktif dan pasif sebagai solusi yang ditawarkan, seperti penelitian yang dilakukan oleh (Molina et al., 2012) menyatakan bahwa management network sebagai media pendukung dalam kasus identifikasi paket data dengan pendekatan *Operation, Administration, Maintenance & Provisioning* dan juga penelitian yang dilakukan oleh (Zhang et al., 2016) mereka berhasil mengenali pola-pola paket dengan baik, namun sistem masih bersifat pasif sehingga tidak ada control trafik yang keluar masuk.

Dhote (2015) pada penelitian survei seleksi fitur untuk klasifikasi trafik internet menyatakan, seleksi fitur dapat membantu memahami data, mengurangi perhitungan, mengurangi efek *curse of dimensionality*, meningkatkan kinerja dan mengurangi waktu komputasi. Sedangkan Aliakbarian (2013) menyatakan ekstraksi mampu membuat data ekstrak baru dengan menghilangkan korelasi yang menghasilkan klasifikasi trafik tersebut optimal. Berdasarkan hal tersebut, maka penelitian ini akan membangun perangkat lunak untuk klasifikasi data trafik internet dengan menggunakan metode *Naive Bayes* dan *Feature Selection (entropy, information gain)* sebagai pemilihan atribut, sehingga dihasilkan klasifikasi data dimensi tinggi dengan akurasi yang baik.

### 1.3 Rumusan Masalah

Rumusan permasalahan yang diselesaikan dalam penelitian ini adalah sebagai berikut:

1. Bagaimana mekanisme *Naive Bayes* dan *Feature Selection* untuk klasifikasi data?
2. Bagaimana hasil dari implementasi *Naive Bayes* dan *Feature Selection* pada klasifikasi data trafik internet?

### 1.4 Tujuan Penelitian

Tujuan penelitian ini adalah mengetahui nilai atribut yang banyak muncul pada data trafik dengan *Confusion Matrix* sesuai dengan *class* yang telah ditetapkan pada *Naive Bayes* dan *Feature Selection*.

### 1.5 Manfaat Penelitian

Manfaat dari penelitian ini adalah sebagai berikut:

1. Memahami *Naive Bayes Classification* dan *Feature Selection* sebagai metode klasifikasi data trafik;
2. Mampu menerapkan teknik klasifikasi dan seleksi *Naive Bayes Classification* dan *Feature Selection* pada klasifikasi data trafik;
3. Hasil penelitian dapat digunakan untuk referensi dalam penelitian lainnya yang sejenis yang menggunakan seleksi fitur *entropy* dan *information gain* dalam metode *Naive Bayes*.

## 1.6 Batasan Masalah

Batasan masalah pada penelitian ini adalah sebagai berikut:

1. Data yang digunakan berupa data trafik internet yang diunduh dari situs *Computer Laboratory University of Cambridge* (<http://www.cl.cam.ac.uk/>) dalam bentuk .xls.
2. Metode klasifikasi yang digunakan adalah *Naïve Bayes*.
3. Evaluasi kualitas pengklasifikasian dilakukan dengan *Confusion Matrix*.
4. Validasi data trafik menggunakan *10-fold Cross Validation*.

## 1.7 Sistematika Penulisan

Penyusunan skripsi ini disusun dengan sistematika penulisan sebagai berikut:

### **BAB I. PENDAHULUAN**

Bab ini membahas mengenai latar belakang, perumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah, dan sistematika penulisan.

### **BAB II. KAJIAN LITERATUR**

Pada bab ini akan membahas dasar-dasar teori yang akan digunakan dalam penelitian, seperti pengetahuan dasar tentang klasifikasi dan metode yang akan digunakan dalam proses klasifikasi data trafik internet.

### **BAB III. METODOLOGI PENELITIAN**

Pada bab ini akan dibahas mengenai unit penelitian, tahapan yang akan dilaksanakan pada penelitian ini, tahapan proses secara umum, metode



pengembangan perangkat lunak, teknik pengujian dan manajemen proyek penelitian.

#### **BAB IV. PENGEMBANGAN PERANGKAT LUNAK**

Pada bab ini membahas mengenai analisis dan perancangan perangkat lunak yang akan digunakan sebagai alat penelitian. Dimulai dari pengumpulan dan analisis kebutuhan, rancangan dan konstruksi perangkat lunak serta pengujian untuk memastikan semua kebutuhan pengembangan perangkat lunak sesuai dengan dengan kebutuhan. Penyusunan pada bab ini memiliki kerangka penulisan dengan fase-fase dan elemen-elemen pengembangan perangkat lunak bersifat berorientasi objek.

#### **BAB V. HASIL DAN ANALISA PENELITIAN**

Pada bab ini diuraikan hasil pengujian berdasarkan langkah-langkah yang telah direncanakan. Tabel hasil pengujian serta analisisnya disajikan sebagai basis dari kesimpulan yang akan diambil dalam penelitian ini.

#### **BAB VI. KESIMPULAN DAN SARAN**

Pada bab ini berisi kesimpulan dari semua uraian-uraian pada bab-bab sebelumnya dan juga saran-saran yang diharapkan berguna untuk pengembangan selanjutnya.

## **1.8 Kesimpulan**

Penelitian mengenai pengklasifikasian data trafik jaringan akan dilakukan dengan metode *Naïve Bayes* dan *Feature Selection* dalam pemilihan atribut. Tujuannya adalah untuk mengembangkan perangkat lunak yang mampu mengklasifikasi data trafik internet.

## DAFTAR PUSTAKA

- Addison, J. F. D., Wermter, S., & Arevian, G. Z. (2003). A Comparison of Feature Extraction and Selection Techniques. *Proceedings of the International Conference on Artificial Neural Networks*, 212–215.
- Aliakbarian, M. S., Fanian, A., Saleh, F. S., & Gulliver, T. A. (2013). Optimal supervised feature extraction in internet traffic classification. *IEEE Pacific RIM Conference on Communications, Computers, and Signal Processing - Proceedings*, 102–107.
- Andersen, D., & Feamster, N. (2006). Challenges and opportunities in Internet data mining. *Data Laboratory, Carnegie Mellon*.
- Dhivya, & P.Shanmugaraja. (2015). STUDY OF INTERNET TRAFFIC CLASSIFICATION USING MACHINE LEARNING ALGORITHMS. *International Research Journal of Engineering and Technology (IRJET)*, 2(1).
- Dhote, Y., Agrawal, S., & Deen, A. J. (2015). A Survey on Feature Selection Techniques for Internet Traffic Classification. *2015 International Conference on Computational Intelligence and Communication Networks (CICN)*, 1375–1380.
- Gorunescu, F. (2011). *Data Mining: Concepts and Techniques*. Elsevier (Vol. 12).
- Han, J. and M. Kamber (2007). "Conception and Technology of Data Mining."
- Han, J., M. Kamber and J. Pei (2012). "Data Mining Concept and Techniques (3rd Edition)."
- Kruchten, P. (2000). *The rational unified process: An introduction*. Reading, MA: Addison-Wesley.
- Kumar, V. (2014). Feature Selection: A literature Review. *The Smart Computing Review*, 4(3).
- Manshaei, Mohammad Hossein. et. al. (2013). *Game Theory Meets Network Security and Privacy*. ACM, Inc.
- Molina, M., Paredes-oliva, I., Routly, W., & Barlet-ros, P. (2012). Operational experiences with anomaly detection in backbone networks. *Computers & Security*, 31(3), 273–285. <https://doi.org/10.1016/j.cose.2012.01.009>

- Novaković, J., Strbac, P., & Bulatović, D. (2011). Toward optimal feature selection using ranking methods and classification algorithms. *Yugoslav Journal of Operations Research*, 21(1), 2334–6043. <https://doi.org/10.2298/YJOR1101119N>
- Okililas, A. F., & Tasmi. (2017). Monitoring and Identification Packet in Wireless With Deep Packet Inspection Method. *International Conference on Recent Trends in Physics 2016 (ICRTP2016) IAES International Conference on Electrical Engineering, Computer Science and Informatics*, 365, 11001. <https://doi.org/10.1088/1742-6596/365/1/011001>
- Siqueira, I. G., Ruiz, L. B., & Loureiro, a. a. F. (2007). Coverage area management for wireless sensor networks. *International Journal of Network Management*, (October 2005), 17–31. <https://doi.org/10.1002/nem>
- Slocum, M. (2012). DECISION MAKING USING ID3 ALGORITHM, 8(2), 1–12.
- STIKI Malang, Lppm & Riyadi, Sigit. (2016). Penerapan Metode Naive Bayes dalam Pengklasifikasi Trafik Jaringan. *SMATIKA Journal*. 06. 29-36.
- Tajunisha, & Saravanan. (2010). Performance analysis of k-means with different initialization methods for high dimensional data. *International Journal of Artificial Intelligence & Applications*, 1(4), 44–52.
- Zhang, H., Yao, D., Ramakrishnan, N., & Zhang, Z. (2016). Causality reasoning about network events for detecting stealthy malware activities. *Computers and Security*, 58, 180–198. <https://doi.org/10.1016/j.cose.2016.01.002>