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

*by* Alfin Ramdhani

**Submission date:** 24-Jul-2019 02:34PM (UTC+0700)
**Submission ID:** 1154575955
**File name:** BAB_1-6.pdf (2.06M)
**Word count:** 4482
**Character count:** 53853

# 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 m engklasifikasikan lalu lintas jaringan, untuk memonitor kapan lalu lintas sed ang sibuk dan ketika lalu lintas sedikit aktivitas. Lihat jenis kegiatan apa y ang sering dilakukan seseorang ketika terhubung ke internet dengan *ranking system* (*browsing*, *chatting*, *videocalling*, *streaming*, dll). Serta mengatasi *bo ttle-*

*necking* atau pengiriman data yang tertunda beberapa saat. Salah satu metod e 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 men ggunakan kombinasi metode *Feature Selection* dalam penyeleksian atribut p ada data trafik jaringan.

*Naive Bayes* (NB) adalah salah satu metode paling awal yang digunak an 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 probab ilitas posterior data yang di tes terhadap kelasnya (Schlosser et al., 2009). Sedangkan *Feature Selection* adalah metode tambahan yang terdiri dari *Entr opy* dan *Information Gain*. Kedua fitur tersebut memungkinkan penulis dala m proses penyeleksian data trafik yang digunakan seperti, pengurangan atrib ut yang tidak banyak mendukung dalam penelitian ini, mendapatkan nilai b obot dari setiap atribut, dan pemeringkatan atribut yang banyak muncul di data trafik.

Penelitian sebelumnya yang dilakukan oleh (Oklilas & Tasmi, 2017) menggunakan metode DPI dalam mengelompokkan jenis trafik, sedang penel itian yang dilakukan oleh (Bujlow, Carela-Español, & Barlet-Ros, 2015) membandingkan tool DPI dalam mengelompokkan jenis trafik i nternet.

Solusi dalam mengklasfikasi trafik pada network telah banyak dilakuka n dengan menghasilkan solusi yang aktif dan pasif sebagai solusi yang dita warkan, seperti penelitian yang dilakukan oleh (Molina et al., 2012) menyat akan bahwa management network sebagai media pendukung dalam kasus id entifikasi paket data dengan pendekatan *Operation, Administration, Maintena nce & 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 trafi k internet menyatakan, seleksi fitur dapat membantu memahami data, mengu rangi perhitungan, mengurangi efek *curse of dimensionality*, meningkatkan k inerja dan mengurangi waktu komputasi. Sedangkan Aliakbarian (2013) men yatakan 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 kl asfikasi data trafik internet dengan menggunakan metode *Naive Bayes* dan

*Feature Selection* (*entropy, information gain)* sebagai pemilihan atribut, sehi ngga dihasilkan klasifikasi data dimensi tinggi dengan akurasi yang baik.

## 1.3 Rumusan Masalah

Rumusan permasalahan yang diselesaikan dalam penelitian ini adalah s ebagai berikut:

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

## 1.4 Tujuan Penelitian

Tujuan penelitian ini adalah mengetahui nilai atribut yang banyak mun cul pada data trafik dengan *Confusion Matrix* sesuai dengan *class* yang tela h 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* sebag ai metode klasifikasi data trafik;
2. Mampu menerapkan teknik klasifikasi dan seleksi *Naive Bayes Cla ssification* dan *Feature Selection* pada klasifikasi data trafik;

3. Hasil penelitian dapat digunakan untuk referensi dalam penelitian l ainnya yang sejenis yang menggunakan seleksi fitur *entropy* dan *in formation gain* dalam metode *Naïve 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 Mat rix*.
4. Validasi data trafik menggunakan *10-fold Cross Validation*.

## 1.7 Sistematika Penulisan

Penyusunan skripsi ini disusun dengan sistematika penulisan sebagai b erikut:

**BAB I. PENDAHULUAN**

Bab ini membahas mengenai latar belakang, perumusan masalah, tuju an penelitian, manfaat penelitian, batasan masalah, dan sistematika pen ulisan.

**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.

# BAB II

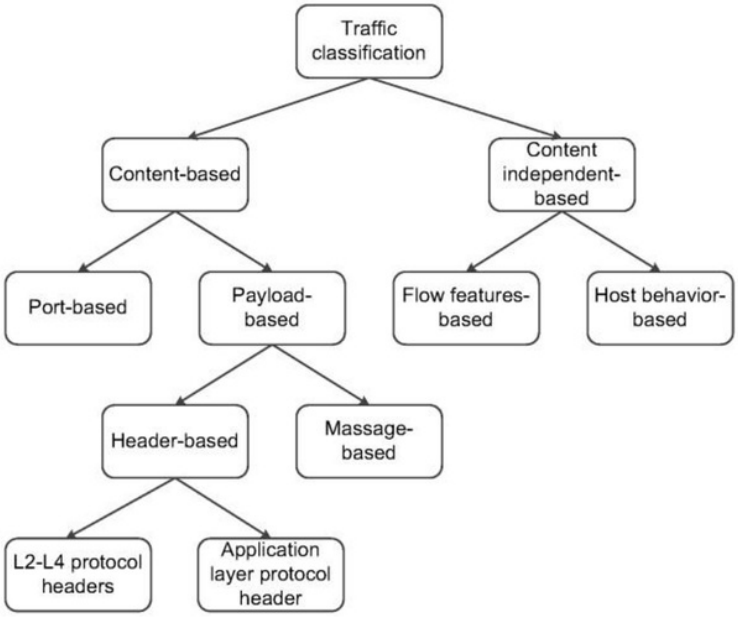
# TINJAUAN PUSTAKA

## 2.1 Pendahuluan

Pada bab ini berisi tentang teori-teori dasar yang berhubungan dengan penelitian. Dalam bab ini dijelaskan secara singkat mengenai klasifikasi trafik, *Naïve Bayes*, *Entropy* dan *Information Gain* dalam *Feature Selection*, *Confusion Matrix*, dan *Cross Validation*. Pada akhir bab ini disertakan penelitian terdahulu yang relevan dengan penelitian ini dan kesimpulan.

## 2.2 Klasifikasi Trafik

Penelitian sebelumnya oleh (Siqueira et al., 2007) menyatakan bahwa klasifikasi trafik adalah satu metode yang digunakan untuk mengoptimalkan *bandwidth* dengan tujuan koneksi internet yang handal dan stabil. Metode-metode klasifikasi trafik dapat dikelompokan dalam metode *Port-Based*, *Payload Based*, *Statistical Classification*. Pada gambar II-1 menunjukkan dalam proses klasifikasi trafik ada dua metode yaitu *content-based* dan *content independent-based*, dimana di bagian *content-based* terdiri dari dua metode yaitu (*port-based* dan *payload-based*) sedangkan dibagian *Content Independent-based* terdiri dari *Flow features-based* dan *Host behavior-based*.




Gambar II-1. Proses Klasifikasi Trafik (Siqueira et al., 2007)

### 2.2.1 Naïve Bayes

Menurut (Dhivya & P.Shanmugaraja, 2015) *Naïve Bayes* adalah al gortima yang berbasis *Bayesian theorem* serta menggunakan perhitunga n probabilitas dalam menentukan kelas. Algortima ini disebut juga seb agai mode fitur independen karena nilai dari atribut pada sebuah kelas tidak tergantung pada nilai atribut yang lain.

Persamaan dari Naïve Bayes sebagai berikut:

$$P(C_i|X) = \frac{P(X|C_i)P(C_i)}{P(X)} \quad \text{(II-1)}$$

Keterangan:

- X = Kriteria suatu kasus berdasarkan masukan
- Ci = Kelas solusi pola ke-
  i, dimana i adalah jumlah label kelas
- P(Ci|X) = Probabilitas kriteria masukan X dengan label ke las Ci
- P(Ci) = Probabilitas label kelas C

*Naïve bayes* merupakan salah satu pembelajaran *supervised* dalam *machine learning*, artinya dalam tahap pembelajaran dibutuhkan data tr aining untuk mengambil keputusan. Pada proses klasifikasi akan dihitu ng nilai probabilitas dari masing-
masing atribut dalam sebuah kelas. Atribut yang mempunyai nilai prob abilitas yang tinggi akan dijadikan label dari kelas tersebut.

Pada proses klasifikasi *Bayesian theorem* diperlukan beberapa penj elasan untuk menentukan kelas yang baik bagi sample yang akan dian alisis, oleh karena itu persamaan *Bayesian theorem* untuk klasifikasi di tampilkan pada persamaan (II-2).

$$p(C|F_1, \ldots, F_n) = \frac{p(C)p(F_1, \ldots, F_n|C)}{p(F_1, \ldots, F_n)} \quad \text{(II-2)}$$

Variabel C menjelaskan kelas, variabel *F1* ... *Fn* merupakan karakte ristik-
karakteristik petunjuk yang dibutuhkan untuk melakukan klasifikasi. Pe

rsamaan (II-

2) menjelaskan bahwa peluang masuknya sampel dengan karakteristik t ertentu dalam kelas C (*posterior*) adalah peluang munculnya kelas C ( *prior*) dikali dengan peluang kemunculan karakteristik-karakteristik sampel pada kelas C (*likelihood*), dibagi dengan peluang kemunculan karakteristik-karakteristik sampel secara global (*evidence*).

## 2.3 Feature Selection Untuk Pemilihan Atribut

Fungsi dari proses seleksi attribut ini adalah pertama untuk menentuk an rangking dari setiap attribut, yang kedua adalah untuk menghilangkan attr ibut yang tidak relevan. Pada penelitian oleh (Slocum, 2012) menyatakan ba hwa pada dasarnya *entropy* digunakan untuk menafsirkan ketidakpastian dari beberapa attribut dari sebuah dataset. Semakin tinggi *entropy* sebuah attribut maka nilai ketidakpastian semakin tinggi, sedangkan penelitian oleh (Novak ović, Strbac, & Bulatović, 2011) untuk mencari nilai entropy menggunakan persamaan (II-3) sebagai berikut:

$$(Y) = -\sum_{y \epsilon Y} (y)2(p(y)) \qquad \text{(II-3)}$$

Dimana, (*py*) adalah fungsi probabilitas marginal untuk nilai variabel a cak Y. Jika nilai-nilai Y yang diukur dalam dataset S dibagi dengan nilai-nilai fitur kedua X, dan *entropy* Y terhadap partisi dipengaruhi oleh X kura ng dari *entropy* Y sebelum proses partisi, maka ada relasi antar fitur Y dan X, maka persaman *entropy* Y setelah mengamati X ditunjukan pada persam aan II.4.

$$H(Y|X) = -\sum_{x\epsilon X} p(x) \sum_{y\epsilon Y} p(y|x) log_2 p((y|x))$$

(II-4)

*p(y|x)* adalah probabilitas bersyarat dari y terhadap x. Entropy sebagai kriteria tidak valid dalam *training dataset* S, maka dapat didefinisikan ukura n merupakan informasi tambahan tentang Y yang disediakan oleh X yang m ewakili jumlah dimana nilai entropy Y menurun. Bagian ini dikenal sebagai *Information Gain (IG)*.

$$IG = H(Y) - H(Y|X) = H(X) - H(X|Y) \quad \text{(II-5)}$$

Dari persamaan II-5, informasi yang diperoleh tentang Y setelah mengamati X sama dengan in formasi yang diperoleh tentang X setelah mengamati Y.

## 2.4 Cross Validation

Cross Validation merupakan teknik yang dapat digunakan untuk mela kukan validasi terhadap model klasifikasi. Teknik ini dapat digunakan untuk menguji performa dan mengukur seberapa akurat model klasifikasi yang tela h dibuat. Salah satu teknik *cross validation* adalah *k-fold cross validation*. *K-fold cross validation* adalah metode validasi di mana dataset dibagi menjadi k bagian, dimana standar yang biasa digunakan untuk memperoleh estimasi kesalahan terbaik adalah 10 bagian atau *ten-fold cross validation* (Gorunescu, 2011).

## 2.5 Confusion Matrix

*Classifier Accuracy Measures* (Han dan Kamber, 2006 : 360) adalah metode klasifikasi yang dilakukan berdasarkan tingkat akurasi model dalam melakukan prediksi. Hal ini dilakukan karena keakuratan dalam mengolah da ta merupakan salah satu hal yang penting.

Metode yang digunakan untuk menguji tingkat akurasi model klasifika si ini adalah *Confusion matrix*. Dalam metode ini memberikan keputusan ya ng diperoleh dalam *training* dan *testing*, *Confusion Matrix* memberikan penil aian *performance* klasifikasi berdasarkan objek dengan benar atau salah. *Con fusion matrix* berisi informasi aktual (*actual*) dan prediksi (*predicted*) pada si stem klasifikasi.

Tabel II-1. Contoh Tabel Confusion Matrix

| Classification | Predicted Class | | |
|---|---|---|---|
| Actual Class | | Class = Yes | Class = No |
| | Class = Yes | a(true positive-TP) | b(false negative-FN) |
| | Class = No | C(false positive-FP) | d(true negative-TN) |

Keterangan:

- True Positive (TP) = Data dengan kelas positif diklasifikasikan ke positi f.

- True Negative (TN) = Data dengan kelas negatif diklasifikasikan ke neg atif.
- False Positive (FP) = Data dengan kelas negatif diklasifikasikan ke posit if.
- False Negative (FN) = Data dengan kelas positif diklasifikasikan ke neg atif.

Dimana setelah itu akan dilakukan perhitungan akurasi, precision, dan recall sebagai berikut:

$$\text{akurasi} = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}} \quad \text{(II-6)}$$

$$Precision = \frac{TP}{TP+FN} \quad \text{(II-7)}$$

$$Recall = \frac{TP}{TP+FP} \quad \text{(II-8)}$$

## 2.6 Penelitian Lain Yang Relevan

Pada bagian ini dipaparkan beberapa penelitian yang telah dilakukan o leh beberapa peneliti lain. Dimaksudkan untuk memperkuat penalaran dan ra sionalitas keterlibatan sejumlah variabel pada penelitian ini. Selain itu juga difungsikan sebagai pendapat ilmilah yang dipadukan dengan hasil kajian pu staka untuk membangun kerangka berpikir peneliti dalam kaitannya dengan masalah yang sedang diteliti.

### 2.6.1. Li Jun, Zhang Shunyi, Lu Yanqing, Zhang Zailong (2007)

Li Jun, Zhang Shunyi, Lu Yanqing, dan Zhang Zailong melak ukan penelitian berjudul *Internet Traffic Classification Using Machine Learning*, Penelitian tersebut dilakukan pada Post and Telecommunicati ons, Nanjing University, Nanjing, China dan Zhejiang Wanli Universit y, Ningbo, China. Penelitian mengenai pengaruh teknik reduksi dimens i dan klasifikasi dilakukan oleh Jun et al. (2014) yang menerapkan te knik reduksi dimensi yaitu *genetic algorithm*(GA) pada metode klasifik asi Naïve Bayes, Naïve Bayes Tree, C4.5 Decision Tree, Random For ests, dan KNN. Penelitian dilakukan dua kali, di mana pertama hanya melakukan klasifikasi terhadap data full set dan kedua menggunakan data yang sudah menggunakan reduksi. Hasil penelitian tersebut reduks i fitur GA tidak terlalu mempengaruhi akurasi metode klasifikasi yan g ada, namun meningkatkan *modelling time* dan *training time*.

### 2.6.1. K. Keerthi Vasan dan B. Surendiran (2016)

K. Keerthi Vasan dan B. Surendiran melakukan penelitian tent ang *Dimensionality reduction using Principal Component Analysis for network intrusion dection*. Penelitian tersebut dilakukan di Departem ent of Computer Science, National Institute of Technology Puducherr y, Karaikal, India. Penerapan kombinasi *Principal Component Analy sis (PCA)* untuk teknik reduksi dimensi dan teknik klasifikasi data in

truksi jaringan, yang dilakukan oleh (Vasan & Surendiran 2016) dida patkan dimensi ideal yakni 10 dimensi pertama. Penelitian ini mengg unakan dua patokan kumpulan data jaringan, KDD Cup dan UNB IS CX dengan hasil rasio reduksi dimensi masing–

masing adalah 0,24 dan 0,36. Sedangkan akurasi klasifikasi sebesar 9 9,7% dan 98,8%, hampir sama dengan akurasi yang diperoleh denga n menggunakan data asli 41 fitur untuk KDD dan 28 fitur untuk IS CX. Hasil penelitian ini membuktikan teknik reduksi dimensi dapat d igunakan untuk membentuk jumlah dimensi ideal serta menghasilkan pengelompokkan data yang baik.

## 2.8 Kesimpulan

Penelitian ini akan menerapkan metode *Naïve Bayes* untuk klasifikasi data trafik internet, Entropy dan Information Gain sebagai fitur seleksi, dat a akan divalidasi menggunakan 10-

*fold cross validation*, dan dievaluasi menggunakan *Confusion Matrix*.

# BAB III

# METODOLOGI PENELITIAN

## 3.1 Pendahuluan

Bab ini akan menjelaskan unit dan tahapan penelitian yang diimplementasikan, metodologi penelitian serta manajemen proyek penelitian. Tahapan penelitian dijadikan sebagai acuan pada setiap fase pengembangan dan memberikan sebuah solusi untuk rumusan masalah dan mencapai tujuan penelitian.

## 3.2 Unit Penelitian

Dalam penelitian ini yang digunakan sebagai unit penelitian adalah situs *Computer Laboratory University of Cambridge (http://www.cl.cam.ac.uk/)* yang menyediakan layanan dataset untuk penelitian.

## 3.3 Data

### 3.3.1. Jenis dan Sumber Data

Data yang digunakan sebagai obyek penelitian adalah jenis data sekunder berupa data trafik internet, yang bersumber dari situs *Computer Laboratory University of Cambridge* (http://www.cl.cam.ac.uk/). Dataset yang digunakan terdiri 19384 data (*record*), 248 fitur, termasuk fitur kelas yang didefinisikan menggunakan teks (*String*), dimana nama-nama fitur dapat dilihat pada lampiran

I. Terdapat 10 kelas pada data trafik internet yaitu, Services, WWW, Attack, P2P, FTP-Pasv, Multimedia, FTP-Control, FTP-Data, dan Interactive.

### 3.3.2. Metode Pengumpulan Data

Data didapatkan secara manual dengan mengunduh data trafik internet di situs *Computer Laboratory University of Cambridge (http://www.cl.cam.ac.uk/)* dalam bentuk file .arff, kemudian menggunakan weka untuk mengubah tipe data menjadi .csv

## 3.3 Tahapan Penelitian

Untuk mendapatkan hasil penyeleksian atribut yang paling banyak muncul dalam data trafik yang digunakan dengan *Feature Selection* dan pengklasifikasian menggunakan *Naïve Bayes*, maka penelitian dengan tahapan-tahapan yang akan dijelaskan pada subbab 3.4.1 sampai dengan 3.4.6.

### 3.3.1. Menetapkan Kerangka Kerja

Data akan di klasifikasi dengan *Naïve Bayes* dan akan diseleksi dengan *Feature Selection.*

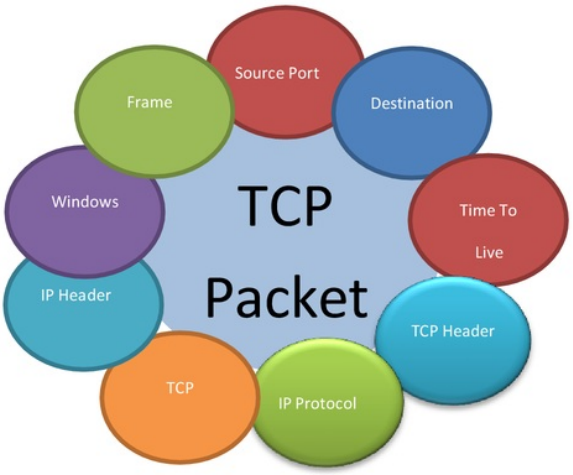
**a. Praproses**

Tahap praproses merupakan tahapan dalam mengelola data masukan. Tahapan pertama dalam praproses pada perangkat lunak ini adalah *Data*

*Cleaning*. Dalam penelitian ini, proses *Data Cleaning* akan membersihkan data yang tidak lengkap fiturnya (*missing value*) dengan cara mengganti nilai-nilai yang hilang. Nilai-nilai yang hilang jika bersifat kontinyu akan diganti dengan rata-rata fitur, dan jika nilai yang hilang bersifat *yes/no* akan diganti dengan modus dari fitur tersebut.

**b. Feature Extraction**

Pada tahap ini adalah proses bagaimana dapat mengambil informasi-informasi atau atribut-atribut dari *data traffic* yang dibutuhkan seperti ditunjukan pada gambar III-1. Pengambilan atau proses mengekstrak *file* mengunakan program weka. *Capture paket* sangat dibutuhkan dalam penelitian ini karena paket-paket yang lewat akan diproses. *Extract Packet* ini adalah proses untuk mendapatkan informasi-informasi dan nilai-nilai atribut dari data trafik.




Gambar III.1 Arsitektur attribut TCP Header (Ford et al., 2011)

**c. Klasifikasi Naïve Bayes dan Feature Selection**

Kelas atau label yang digunakan pada klasifikasi data trafik internet adalah Mail, Services, WWW, Attack, P2P, FTP-Pasv, Multimedia, FTP-Control, FTP-Data, dan Interactive. Proses klasifikasi dilakukan oleh metode *Naïve Bayes*, dan penyeleksian atribut dilakukan oleh *entropy* dan *information gain* yang akan menghasilkan keluaran perangkingan jenis atribut yang paling banyak keluar dari data trafik.

**d. Evaluasi dan validasi menggunakan *Confusion Matrix* dan *10-fold Cross Validation***

*Confusion Matrix* digunakan untuk mencari nilai *true positive, true negative, false positive,* dan *false negative*. Jika label yang digunakan adalah Mail, Services, WWW, Attack, P2P, FTP-Pasv, Multimedia, FTP-Control, FTP-Data, dan Interactive, maka pada *confusion matrix* dihitung nilai *true* Mail, *true* Services, *true* WWW, *true* Attack, *true* P2P, *true* FTP-pasv, *true* Multimedia, *true* FTP-control, *true* FTP-data, *true* Interactive, *false* Mail, *false* Services, *false* WWW, *false* Attack, *false* P2P, *false* FTP-pasv, *false* Multimedia, *false* FTP-control, *false* FTP-data, dan *false* Interactive. Nilai-nilai tersebut akan digunakan untuk menghitung akurasi, *precision,* dan *recall*. Model klasifikasi divalidasi menggunakan 10-*fold cross validation* dimana *dataset* dibagi menjadi 10 bagian dengan satu bagian digunakan sebagai data pengujian dan sisanya digunakan sebagai data pelatihan yang dilakukan sebanyak 10 kali dengan

bagian data pengujian yang berbeda-beda. Tiap pengujian diukur menggunakan *confusion matrix*.

### 3.3.2. Menetapkan Kriteria Pengujian

Pembahasan mengenai tahapan ini akan dijelaskan pada bab IV. Pada tahapan pengujian awal penelitian, data terlebih dahulu akan diklasifikasi dengan menggunakan metode *Naïve Bayes* dan diseleksi oleh *Entropy* dan *information gain* dengan validasi data menggunakan 10-*fold cross validation*. Proses klasifikasi akan dilanjutkan dengan perhitungan mencari nilai akurasi, *precision*, dan *recall*, serta hasil akhir berupa data per kelas dalam bentuk tabel oleh *Confusion Matrix*.

### 3.3.3. Menetapkan Format Data Pengujian

Hasil perhitungan dalam pengujian, akan dijelaskan sebagai berikut:

Hasil pengujian klasifikasi dengan tabel *confusion matrix* pada tiap bagian data akan digambarkan dalam tabel III-1

Tabel III-1. Rancangan Tabel *Confusion Matrix* Untuk Setiap Hasil Pengujian

| Cross validation | | Prediksi | | | |
|---|---|---|---|---|---|
| | Label | MAIL | SERVICE | … | INTERACTIVE |
| | MAIL | | | | |
| | SERVICE | | | | |
| | … | | | | |
| | INTERACTIVE | | | | |

Hasil pengujian klasifikasi akan digambarkan pada tabel III-2.

Tabel III-2. Rancangan Tabel Hasil Pengujian Untuk Setiap Hasil Klasifikasi

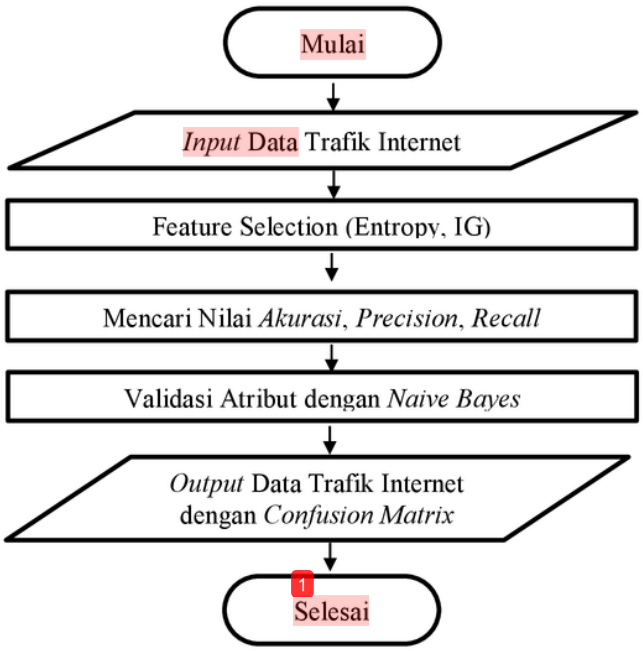
| Cross Validation | Akurasi | Precision | Recall |
|---|---|---|---|
| 1 | | | |
| ….. | | | |
| 10 | | | |
| Rata-rata | | | |

### 3.3.4 Menentukan Alat yang Digunakan dalam Pelaksanaan Penelitian

Untuk melaksanakan penelitian mengenai klasifikasi data trafik menggunakan Naïve Bayes dan Feature Selection dalam pemilihan atribut

dibutuhkan alat penelitian. Oleh karena itu, penulis akan mengembangkan sebuah perangkat lunak yang dapat melakukan proses klasifikasi data trafik internet.

### 3.3.5 Melakukan Pengujian Penelitian




Gambar III-2. Tahapan Pengujian Penelitian

### 3.3.6 Melakukan Analisa Hasil Pengujian dan Membuat Kesimpulan

Untuk mengetahui hasil penelitian mengenai klasifikasi data trafik menggunakan *Naïve Bayes* dan *Feature Selection* dalam pemilihan atribut, maka cara yang dilakukan adalah dengan membandingkan hasil pengujian akurasi, *precision*, dan *recall* dari Metode *Naïve Bayes dan Feature Selection*. Analisis

hasil pengujian dapat dilihat pada tabel III-3. Setelah mendapatkan hasil analisis pengujian penelitian, maka langkah selanjutnya adalah membuat kesimpulan penelitian yang akan dijelaskan pada bab V.

### 3.4 Metode Pengembangan Perangkat Lunak

Metodologi yang diterapkan dalam pengembangan perangkat lunak sebagai alat penelitian tugas akhir ini berorientasi pada objek menggunakan metode *Rational Unified Process* (RUP). Secara umum, langkah-langkah yang akan dilakukan pada pengembangan perangkat lunak adalah fase insepsi, elaborasi, konstruksi, dan transisi.

### 3.5 Manajemen Proyek Penelitian

Manajemen proyek merupakan perencanaan aktivitas penelitian dari tahap inisialisasi masalah sampai dengan pada tahap kesimpulan dari penelitian. Adapun kegiatan-kegiatan yang berlangsung selama penelitian dapat dilihat dalam *Work Breakdown Structure* (WBS) pada lampiran IV.

### 3.6 Kesimpulan

Penelitian ini menggunakan data trafik internet dari situs *Computer Laboratory University of Cambridge* (http://www.cl.cam.ac.uk/). Data tersebut akan melalui tahap praproses sebelum di proses. Kemudian data terlebih dahulu akan diklasifikasi dengan metode *Naïve Bayes*. Lalu data diseleksi oleh *Feature Selection* dengan *Entropy* dan *Information Gain*, sehingga akan diperoleh hasilnya dengan membuat peringkat setiap atribut yang banyak muncul.

# BAB IV

# PENGEMBANGAN PERANGKAT LUNAK

## 4.1 Pendahuluan

Pada bab III disebutkan bahwa diperlukan sebuah alat berupa perang kat lunak yang digunakan dalam pelaksanaan penelitian ini, maka penulis mengembangkan perangkat lunak dengan metode pemrograman berorientasi obyek berdasarkan panduan *Rational Unified Process*. Di dalam *Rational U nified Process* terdapat empat fase pengembangan perangkat lunak yaitu fase insepsi, elaborasi, konstruksi, dan transisi. Setiap fase terdiri dari pemodela n bisnis, kebutuhan, analisis dan desain, implementasi, dan pengujian. Pada bab ini dibahas proses pengembangan perangkat lunak yang digunakan seba gai alat penelitian.

## 4.2 Fase Insepsi

Tahapan pertama dalam pengembangan perangkat lunak ialah melaku kan identifikasi terhadap sistem yang dikembangkan. Aktivitas yang dilakuk an pada fase ini meliputi penentuan *user requirement* dan fungsionalitas per angkat lunak pada permodelan bisnis, mengumpulkan data penelitian pada k ebutuhan, membuat *use-case diagram* pada analisis dan desain, mendokumentasikan *user requiremen t*, fungsionalitas perangkat lunak, dan *use-*

*case diagram* pada implementasi, serta memastikan *user requirement* dan fu ngsionalitas perangkat lunak valid pada pengujian.

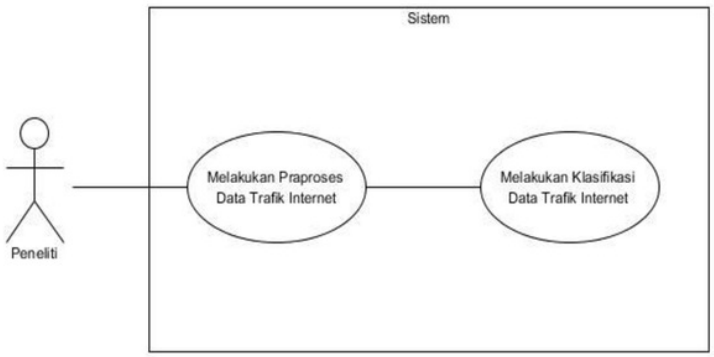
### 4.2.1 Permodelan Bisnis

Andersen dan Feamster (2006) menyatakan trafik internet menghasilk an sejumlah besar data yang membantu pengelola jaringan untuk mengelola dan mempelajari berbagai karateristik trafik internet. Untuk mempermudah mengidentifikasi data trafik internet, trafik internet dapat dikelompokkan ber dasarkan isi trafik internet tersebut. Namun pengelompokan secara manual menimbulkan permasalahan sebagai berikut:

1. Dibutuhkan ketelitian yang tinggi untuk menganalisis satu pers atu yang melewati trafik;
2. Membutuhkan waktu yang lama untuk mengelompokkan;
3. Dibutuhkan banyak sumber daya manusia.

Dari permasalahan tersebut, dibutuhkan pengelompokan otomati s terhadap data trafik internet dengan melakukan klasifikasi. Pada gambar IV–

1 menunjukkan proses pengelompokkan data trafik internet menggu nakan klasifikasi.




Gambar IV – 1. Diagram *Use Case Current Existing*

Namun, jika jumlah data yang akan diklasifikasi berjumlah sangat banyak, maka algoritma klasifikasi biasa tidak mampu menghasilkan performa klasifikasi yang optimal karena jumlah data yang banyak akan menghasilkan fitur data yang sangat banyak. Tidak semua fitur relevan digunakan pada proses klasifikasi dan juga dapat menurunkan performa klasifikasi. Untuk itu, dibutuhkan teknik reduksi dimensi untuk memaksimalkan kinerja algoritma klasifikasi.

Perangkat lunak yang dibangun merupakan perangkat lunak berbasis desktop yang mampu melakukan klasifikasi data trafik internet secara otomatis. Masukan pada perangkat lunak yang dikembangkan berupa berkas bertipe *.xls* yang berisikan konten data trafik internet. Keluaran yang dihasilkan berupa tabel berisi data trafik internet yang telah terkelompok berdasarkan kelasnya.

### 4.2.2 Kebutuhan Sistem

Berdasarkan permodelan bisnis yang telah dilakukan, spesifikasi kebu tuhan perangkat lunak yang akan dikembangkan memiliki fitur prapengolaha n dan klasifikasi.

#### 4.2.2.1 Fitur Prapengolahan

Perangkat lunak dilengkapi fitur prapengolahan yang digunakan oleh pengguna untuk menyiapkan data trafik internet agar dapat di reduksi dan d i klasifikasi. Prapengolahan yang akan dilakukan adalah *replace missing val ue*.

#### 4.2.2.2 Fitur Klasifikasi

Fitur berikutnya adalah klasifikasi data trafik internet. Selanjutnya pr oses perhitungan akan berlanjut pada tahapan pengklasifikasian trafik interne t menggunakan *Naïve Bayes* dari data trafik.

#### 4.2.2.3 Fitur Seleksi

Fitur ini adalah fitur yang digunakan untuk menseleksi atribut data tr afik yang banyak dengan atribut yang hanya berpengaruh dalam melakukan penelitian ini karena tidak semua atribut dari dataset yang dipakai cocok de ngan topik dari penelitian.

#### 4.2.2.4 Fitur Evaluasi dan Validasi

Fitur ini merupakan fitur yang digunakan untuk memvalidasi dataset menggunakan metode 10-*fold cross validation* dan mengukur akurasi, *precision,* dan recall mengguna kan *confusion matrix.*

Untuk dapat merealisasikan fitur-fitur tersebut, perangkat lunak harus memenuhi kebutuhan fungsional dan no n-fungsional. Kebutuhan fungsional menjelaskan kebutuhan atau fasilitas utama perangkat lunak yang dibangun pada tabel IV–1. Sedangkan, kebutuhan non fungsional menjelaskan kebutuhan atau fasilita s yang tidak wajib dimiliki oleh perangkat lunak, dalam artian hanya merup akan pelengkap agar perangkat lunak lebih baik kinerjanya yang dapat dilih at pada tabel IV–2.

Tabel IV – 1 Kebutuhan Fungsional

| No. | Kebutuhan |
|---|---|
| 1 | Perangkat lunak dapat memuat data trafik internet |
| 2 | Perangkat lunak dapat melakukan praproses data trafik internet |
| 3 | Perangkat lunak dapat melakukan proses klasifikasi data trafik internet |

Tabel IV – 2 Kebutuhan Non Fungsional

| No. | Kebutuhan |
|---|---|
| 1 | Perangkat lunak dapat menampilkan pesan kesalahan jika terdapa t aksi pengguna yang salah dan konfirmasi tindakan pengguna. |
| 2 | Perangkat lunak memiliki antarmuka yang mudah dimengerti dan digunakan oleh pengguna. |

### 4.2.3 Analisis dan Desain

Kegiatan yang dilakukan pada tahapan ini adalah menganalisis kebut uhan perangkat lunak, data, prapengolahan, klasifikasi dengan *Naive Bayes*, dan evaluasi menggunakan *confusion matrix*.

#### 4.2.3.1 Analisis Kebutuhan Perangkat Lunak

Dari pemodelan bisnis yang telah dijabarkan, untuk menyelesaikan pe rmasalahan yang terjadi diperlukan perangkat lunak yang mampu mengelom pokkan data trafik internet secara otomatis. Untuk itu, perangkat lunak haru s memiliki kemampuan sebagai berikut:

1. Melakukan prapengolahan berupa *replace missing value* terhadap data trafik internet;
2. Melakukan klasifikasi data trafik internet secara otomatis menggu nakan *Naive Bayes* dan *Feature Selection*;

3. Melakukan validasi terhadap data trafik internet menggunakan *10-fold cross validation*;
4. Melakukan evaluasi terhadap hasil klasifikasi menggunakan *confus ion matrix*.

Pengembangan perangkat lunak dimulai dengan mengumpulkan data t rafik internet yang diunduh dari situs *Computer Laboratory University of C ambridge* (http://www.cl.cam.ac.uk). Selanjutnya file data trafik internet disi mpan dalam bentuk .xls. Selanjutnya berkas tersebut diimpor ke perangkat l unak lalu akan dilakukan prapengolahan berupa *replace missing value*. Taha p selanjutnya ialah melakukan klasifikasi *Naive Bayes.*

#### 4.2.3.2 Analisis Prapengolahan

Tahap prapengolahan yang dilakukan pada data trafik internet yaitu *replace missing value*. Contoh data trafik internet yang digunakan sebanyak 7 buah, masing-masing memiliki 4 fitur.

Tabel IV – 3 Contoh Data Trafik Internet

| F1 | F2 | F3 | F4 | KELAS |
|---|---|---|---|---|
| 55 | 1 | 1 | 3 | MAIL |
| 57 | 1 | 1 | 2 | MAIL |
| 49 | ? | 1 | 3 | MAIL |

| | | | | |
|---|---|---|---|---|
| 49 | 1 | ? | 3 | MAIL |
| 5 | ? | 1 | 0 | WWW |
| 5 | 1 | 0 | 1 | WWW |
| 6 | ? | 1 | 0 | WWW |

Berikut langkah-langkah prapengolahan yang dilakukan pada ke-tujuh data dapat dilihat dibawah ini:

1. Dari data yang telah dimasukkan, diperiksa tiap fiturnya;
2. Ditemukan, fitur F2 terdapat 3 missing value dan F3 terdapat 1 *missing value* dari 7 data trafik internet;
3. Hitung rata-

   rata fitur yang terdapat missing value kemudian ubah missing va lue tersebut dengan rata-rata fitur tersebut.

Tabel IV – 4 Hasil Prapengolahan

| F1 | F2 | F3 | F4 | KELAS |
|---|---|---|---|---|
| 55 | 1 | 1 | 3 | MAIL |
| 57 | 1 | 1 | 2 | MAIL |
| 49 | 1 | 1 | 3 | MAIL |

| | | | | |
|---|---|---|---|---|
| 49 | 1 | 0.83 | 3 | MAIL |
| 5 | 1 | 1 | 0 | WWW |
| 5 | 1 | 0 | 1 | WWW |
| 6 | 1 | 1 | 0 | WWW |

#### 4.2.3.3 Desain Perangkat Lunak

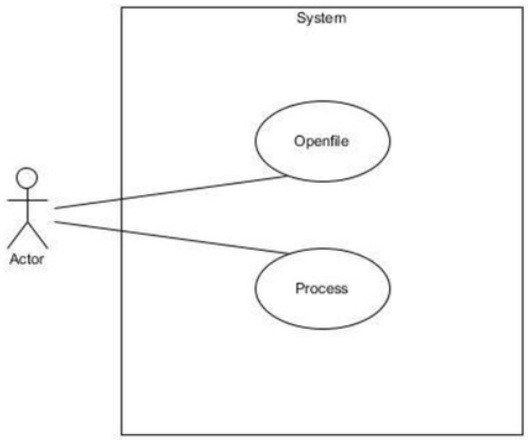
Desain perangkat lunak dideskripsikan dengan diagram *use c ase* dan diagram aktivitas.

**1. *Use Case***

Pada subbab ini dijelaskan gambaran fungsionalitas perangk at lunak yang dibangun dengan menggunakan pemodelan *Use Case*.

a) Diagram *Use Case*

Diagram *Use Case* menjelaskan secara umum kegiatan yan g dilakukan oleh aktor terhadap perangkat lunak yang dapat dilihat pada gambar IV–2.




Gambar IV–2. Diagram *Use Case*

b) Tabel Definisi Aktor

Dalam penelitian ini, seluruh pengguna menjadi aktor yang dijelaskan pada tabel IV –

5. Definisi dari aktor dapat dilihat pada tabel IV-5.

Tabel IV–5 Definisi Aktor Use Case

| Nomor | Aktor | Definisi |
|---|---|---|
| 1 | Peneliti | Peneliti adalah orang yang berhubungan dengan perangkat lunak aplikasi untuk memasukkan data, melatih algoritma NB, melakukan seleksi fitur menggunakan Feature Selection, melakukan klasifikasi, dan menampilkan hasil dari pengujian yang dilakukan. |

c) Tabel Definisi Use Case

Definisi *Use case* yang dijelaskan pada tabel IV–6 merupakan penjelasan dari kerja perangkat lunak secara spesifik pada gambar IV–2 disimbolkan dengan sebuah notasi lingkaran yang agak lonjong atau berbentuk oval.

Tabel IV – 6 Definisi *Use Case*

| Nomor | Use Case | Deskripsi |
|---|---|---|
| 1 | Melakukan Praproses Data Trafik Internet | Aktivitas memasukkan data trafik internet dalam format .xls yang kemudian dilakukan praproses *replace missing value*. |
| 2 | Melakukan Klasifikasi dengan NB | Kegiatan ini digunakan untuk melakukan klasifikasi data trafik internet yang telah dipraproses menggunakan metode NB. |

d) Skenario *Use Case*

Skenario adalah merupakan urutan spesifik dari aksi dan interaksi antara aktor dan sistem. Berikut ini adalah skenario dari *use case* yang telah didefinisikan sebelumnya.

1. Skenario *Use Case* Melakukan Klasifikasi Dengan *Naive Bayes*

   No : 001

   Nama Use Case: Melakukan Klasifikasi Dengan *Naive Bayes* dan *Feature Selection*

   Aktor : Pengguna

   Tujuan : Mengetahui kelompok data trafik internet

   Deskripsi: *Use Case* ini digunakan untuk melakukan klasifikasi dengan metode *Naive Bayes* dan *Feature Selection*.

   Kondisi Awal: Terdapat data trafik internet yang siap diproses.

   Kondisi Akhir: Hasil klasifikasi data internet, nilai akurasi, hasil seleksi atribut data.

Skenario *use case* menambahkan data trafik internet dijelaskan pada tabel IV-7

Tabel IV-

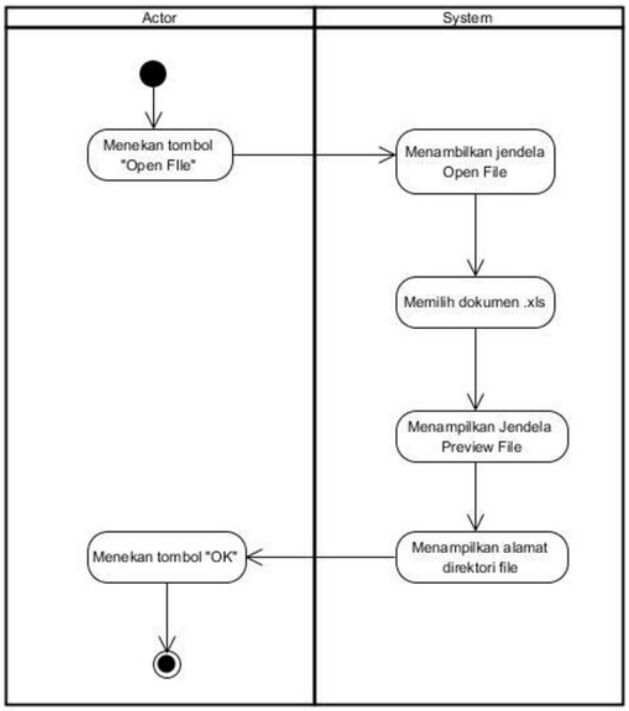
7 Skenario *Use Case* Melakukan Klasifikasi dengan *Naive Baye s* dan *Feature Selection*.

| Aktor | Sistem |
|---|---|
| 1. Pengguna menek an tombol "Ope n File" | |
| | 2. Menampilkan kotak dialog yang menu njukkan direktori penyimpanan data trafi k internet |
| 3. Pengguna memilih data yang memiliki format .xls untuk d iproses | |
| | 4 Menampilkan jendela yang menunjukka n preview file yang akan diklasifikasi |
| 6. Pengguna meneka n tombol "Process" | |
| | 7. Melakukan *10 fold Cross Validation* |
| | 8. Klasifikasi menggunakan NB |

| | |
|---|---|
| | 9. Penyeleksian atribut dengan *Entropy* dan *IG* |
| | 10. Menampilkan hasil klasifikasi dan nilai akurasi, *precison*, *recall*, serta nilai hasil seleksi atribut dari *Confusion Matrix* dalam tabel |

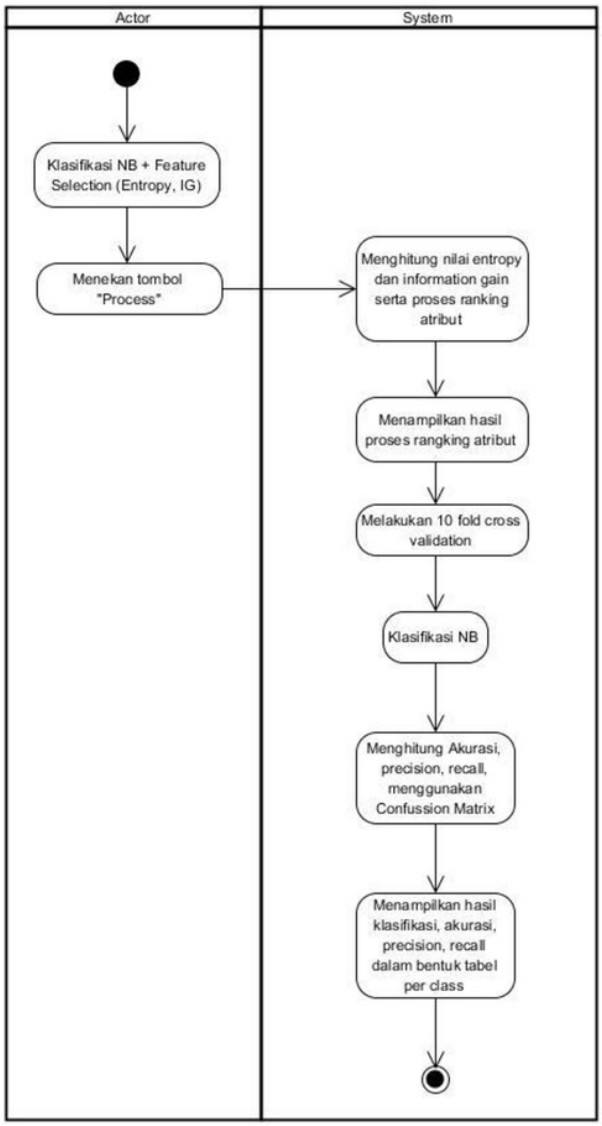
**2. Diagram Aktivitas**

Diagram aktivitas menggambarkan aliran kerja atau aktivitas dari sebuah sistem atau proses bisnis. Diagram aktivitas melakukan praproses dapat dilihat pada gambar IV-3, diagram aktivitas melakukan klasifikasi dengan Naive Bayes dan Feature Selection dapat dilihat pada gambar IV-4.




Gambar IV-

3. Diagram Aktivitas *Use Case* melakukan Praproses Data Tra fik Internet




Gambar IV-

4. Diagram Aktivitas Use Case Melakukan Klasifikasi dengan *Naive Bayes dan Feature Selection*

## 4.3 Fase Elaborasi

Tahapan kedua dalam pengembangan perangkat lunak adalah melakukan identifikasi terhadap sistem yang dikembangkan. Aktivitas yang dilakukan mencakup perancangan data, perancangan antarmuka, identifikasi kebutuhan, perumusan kebutuhan pengujian, pemodelan diagram sequence, dan pembuatan dokumentasi.

### 4.3.1 Permodelan Bisnis

Pada subbab ini akan dibahas mengenai perancangan perangkat lunak yang akan dibangun. Perancangan dilakukan berdasarkan hasil analisis yang telah dilakukan pada fase insepsi. Perancangan yang dibahas pada subbab ini meliputi perancangan data dan perancangan antar muka.

#### 4.3.1.1 Perancangan Data

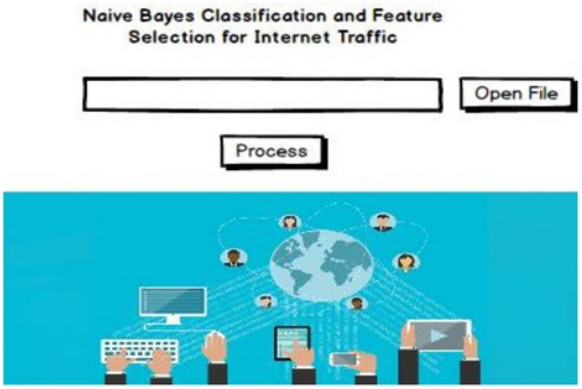
Perangkat lunak yang akan dibangun memiliki kemampuan klasifikasi terhadap data. Adapun data yang akan melalui proses klasifikasi adalah data trafik internet yang disimpan dalam file berformat .xls.

#### 4.3.1.2 Perancangan Antar Muka

Pada subbab ini membahas tentang perancangan antar muka dari perangkat lunak yang dibangun. Adapun rancangan antarmuka Menu Utama digambarkan pada gambar IV–5, rancangan antarmuka *File Preview* digambarkan pada gamb

ar IV-

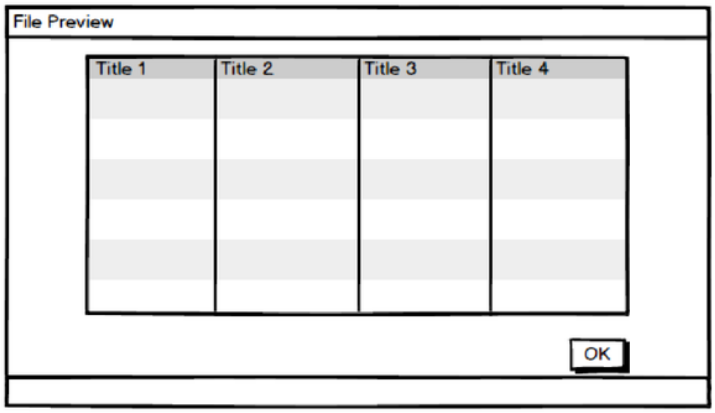
6, dan rancangan antarmuka *Output View* pada gambar IV-7.




Gambar IV-5. Rancangan Antarmuka Menu Utama

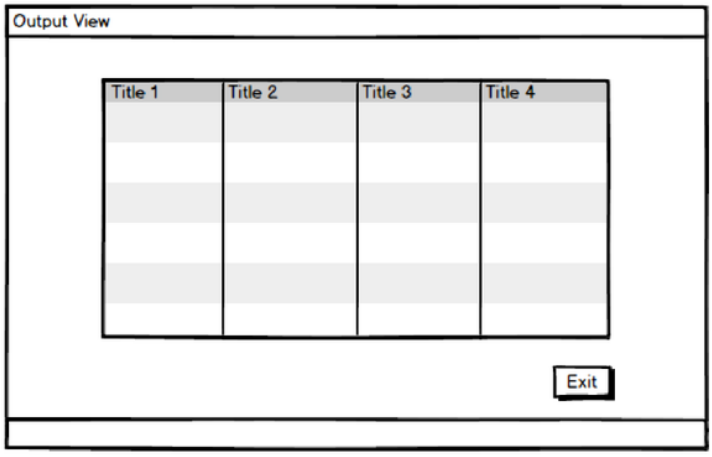
Gambar IV-

5 merupakan halaman utama yang berfungsi untuk melakukan klasifikasi data trafik internet. Terdapat empat bagian pada ant armuka ini yaitu tombol Open File untuk membuka file data . xls, *text field* untuk menampilkan direktori data trafik, tombol Process untuk memproses klasifikasi data trafik, dan tabel ya ng menampilkan output data per kelas.




Gambar IV-6. Rancangan antarmuka *File Preview*

Gambar IV-6 merupakan tampilan yang berfungsi untuk menampilkan isi file .xls yang akan diproses. Terdapat dua bagian yaitu tabel yang akan menampilkan isi file tersebut dan tombol "OK" untuk mengkonfirmasi apakah file yang diuji sudah sesuai.




Gambar IV-7. Rancangan antarmuka *Output View*

Gambar IV-7 merupakan tampilan yang berfungsi untuk menampilkan nilai output yang telah diproses. Terdapat dua bagian yaitu tabel yang akan menampilkan nilai tersebut dan tombol "Exit" untuk keluar dari jendela *Output View*.

### 4.3.2 Kebutuhan Sistem

Pada subbab ini dibahas mengenai kebutuhan sistem dari perangkat lunak yang dibangun berdasarkan hasil analisis dan perancangan pada tahap selanjutnya. Untuk membangun perangkat lunak dalam penelitian ini dibutuhkan perangkat keras (*hardware*), perangkat lunak (*software*) dan bahasa pemrograman. Bahasa pemrograman yang digunakan untuk implementasi perangkat lunak adalah Java, Perangkat keras yang digunakan pada tahap pengembangan dan penelitian ini adalah laptop dengan spesifikasi sebagai berikut:

1. Laptop merk ASUS GL553VE;
2. Processor Intel(R) Core(TM) i7-7700HQ CPU @ 2.80GHz;
3. RAM 8 GB;
4. Hard Disk 1 TB.

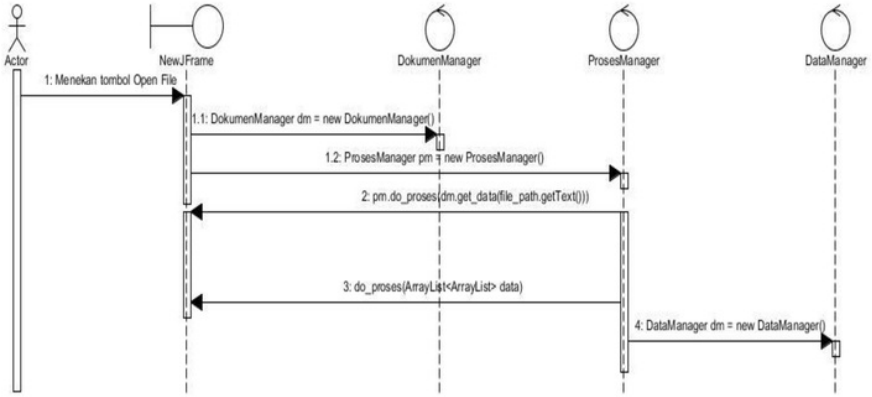
Sedangkan perangkat lunak yang digunakan untuk implementasi yaitu :

1. Windows 10 Home Single Language 64-bit;
2. Compiler Netbeans IDE 8.2;

3. Visual Paradigm UML Enterprise Edition V8.0.

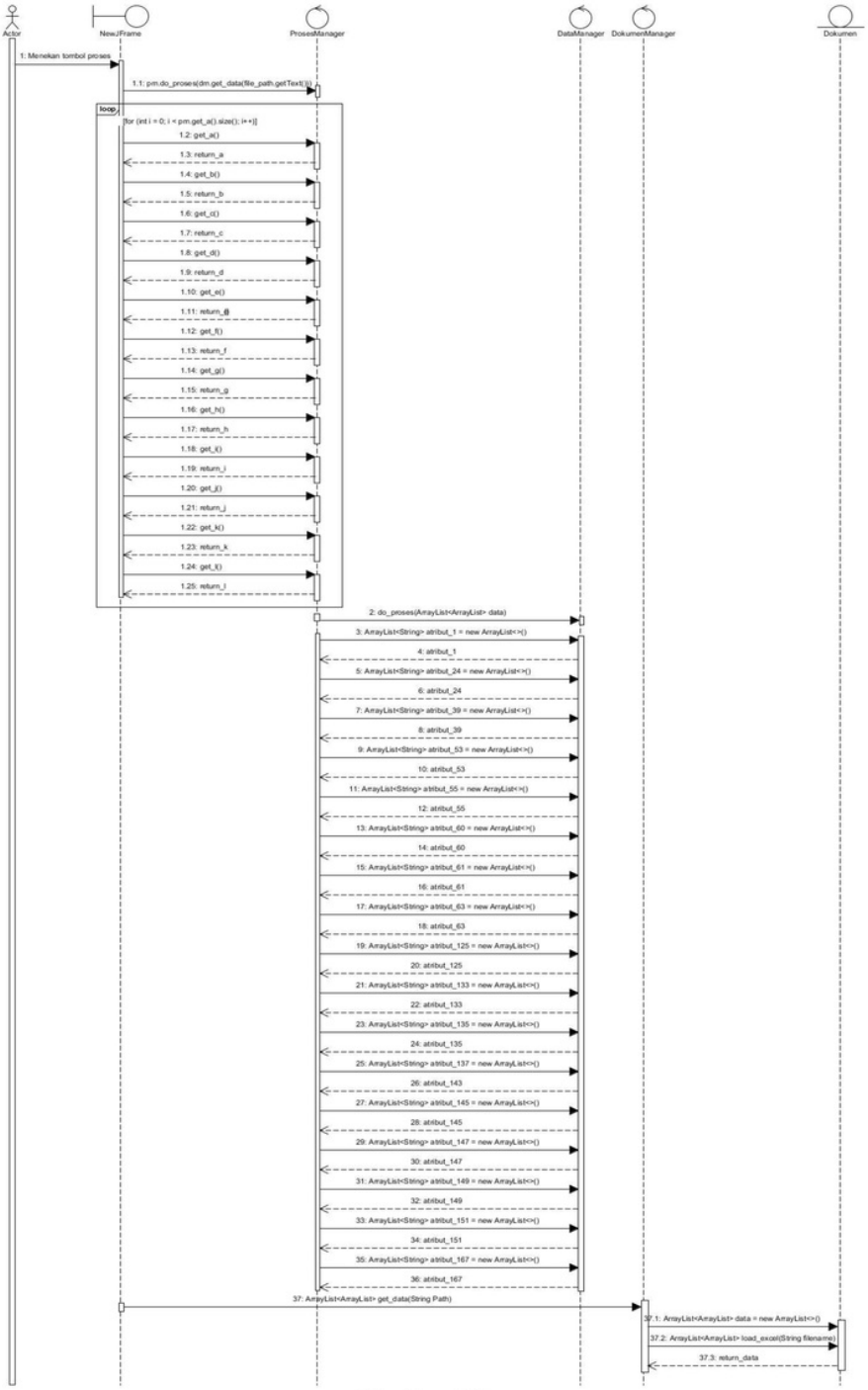
### 4.3.3 Diagram Sequence

Diagram sequence adalah diagram yang menggambarkan kolab orasi dinamis antara sebuah objek. Berdasarkan use case yang dibent uk, diagram sequence yang dibentuk pada pengembangan perangkat l unak ini berjumlah 2 buah. Diagram *sequence* melakukan praproses Data Trafik Internet dapat dilihat pada gambar IV-
8, diagram sequence melakukan klasifikasi dengan Naive Bayes dan Feature Selection dapat dilihat pada gambar IV-9.




Gambar IV-

8. *Sequence Diagram* Melakukan Praproses Data Trafik Internet


Actor
NewJFrame
ProsesManager
DataManager
DokumenManager
Dokumen
1: Menekan tombol proses
1.1: pm.do_proses(dm.get_data(file_path.getText()))
loop
[for (int i = 0; i < pm.get_a().size(); i++)]
1.2: get_a()
1.3: return_a
1.4: get_b()
1.5: return_b
1.6: get_c()
1.7: return_c
1.8: get_d()
1.9: return_d
1.10: get_e()
1.11: return_e
1.12: get_f()
1.13: return_f
1.14: get_g()
1.15: return_g
1.16: get_h()
1.17: return_h
1.18: get_i()
1.19: return_i
1.20: get_j()
1.21: return_j
1.22: get_k()
1.23: return_k
1.24: get_l()
1.25: return_l
2: do_proses(ArrayList<ArrayList> data)
3: ArrayList<String> atribut_1 = new ArrayList<>()
4: atribut_1
5: ArrayList<String> atribut_24 = new ArrayList<>()
6: atribut_24
7: ArrayList<String> atribut_39 = new ArrayList<>()
8: atribut_39
9: ArrayList<String> atribut_53 = new ArrayList<>()
10: atribut_53
11: ArrayList<String> atribut_55 = new ArrayList<>()
12: atribut_55
13: ArrayList<String> atribut_60 = new ArrayList<>()
14: atribut_60
15: ArrayList<String> atribut_61 = new ArrayList<>()
16: atribut_61
17: ArrayList<String> atribut_63 = new ArrayList<>()
18: atribut_63
19: ArrayList<String> atribut_125 = new ArrayList<>()
20: atribut_125
21: ArrayList<String> atribut_133 = new ArrayList<>()
22: atribut_133
23: ArrayList<String> atribut_135 = new ArrayList<>()
24: atribut_135
25: ArrayList<String> atribut_137 = new ArrayList<>()
26: atribut_143
27: ArrayList<String> atribut_145 = new ArrayList<>()
28: atribut_145
29: ArrayList<String> atribut_147 = new ArrayList<>()
30: atribut_147
31: ArrayList<String> atribut_149 = new ArrayList<>()
32: atribut_149
33: ArrayList<String> atribut_151 = new ArrayList<>()
34: atribut_151
35: ArrayList<String> atribut_167 = new ArrayList<>()
36: atribut_167
37: ArrayList<ArrayList> get_data(String Path)
37.1: ArrayList<ArrayList> data = new ArrayList<>()
37.2: ArrayList<ArrayList> load_excel(String filename)
37.3: return_data

Gambar IV-

9. *Sequence Diagram* Melakukan Klasifikasi dengan *Naïve Bayes*

## 4.4 Fase Konstruksi

Fase konstruksi berfokus pada pengembangan perangkat lunak baik komponen utama maupun fitur-fitur pendukung dengan melakukan sederet iterasi. Disetiap iterasi terdapat proses analisis, desain, implementasi, dan pengujian. Dalam proses pengembangannya dapat menggunakan konstruksi paralel agar mempercepat hasil perangkat lunak. Hasil yang diharapkan dari fase ini adalah sebuah produk perangkat lunak yang siap digunakan oleh end-user, yaitu sebuah produk perangkat lunak yang dapat digunakan sebagai alat penelitian.

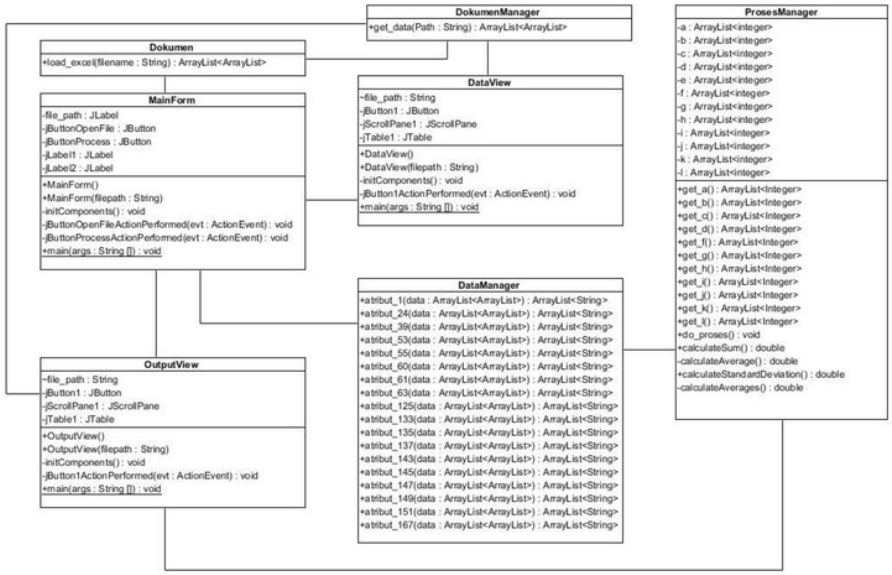
### 4.4.1 Kebutuhan Sistem

Dalam pengembangan perangkat lunak ini, penulis menggunakan beberapa library. Diantaranya adalah weka, dan *Jxl*. *Library jxl* digunakan untuk memproses data dari file .xls, metode klasifikasi Naïve Bayes menggunakan weka.

### 4.4.2 Diagram Kelas

Diagram kelas adalah diagram UML yang menggambarkan kelas-kelas dalam sebuah sistem dan hubungannya antara satu dengan yang lain. Terdapat serta dimasukkan pula atribut dan operasi. Terdapat 7 kelas yang terdiri dari 3 kelas *boundary* (kelas MainForm, DataVie

w, OutputView), 3 kelas *control* (kelas DataManager, DokumenMana ger, ProsesManager), dan 1 kelas *entity* (kelas Dokumen).




Gambar IV-10. *Class Diagram*

### 4.4.3 Implementasi

Fase implementasi dalam konstruksi adalah mengembangkan perangka t lunak berdasarkan diagram kelas dan rancangan antarmuka yang tel ah dibuat dalam fase sebelumnya.

#### 4.4.3.1 Implementasi Kelas

Kelas-

kelas yang telah dirancang pada diagram kelas diimple mentasikan dalam bahasa pemrograman Java. Tabel (IV

-

8) menunjukkan implementasi kelas dalam bahasa Java.

Tabel IV-8. Implementasi Kelas

| No | Nama Kelas | Nama File | Keterangan |
|---|---|---|---|
| 1 | MainForm | MainForm.java | Kelas MainForm adalah kelas *boundary* yang merupakan form utama dan menyediakan akses untuk melakukan masukkan data dan memproses klasifikasi. |
| 2 | DataView | DataView.java | Kelas DataView adalah kelas *boundary* yang bertujuan untuk menampilkan preview dari data (file .xls) yang akan diuji. |
| 3 | OutputView | OutputView.java | Kelas DataView adalah kelas *boundary* yang bertujuan untuk menampilkan *output* dari nilai atribut yang paling banyak muncul sesuai *class* yang ditetapkan |
| 4 | DokumenManager | DokumenManager.java | Kelas DokumenManager |

| | | | |
|---|---|---|---|
| | | | merupakan kelas *control* yang menangani proses pengambilan data dari file excel. |
| 5 | Dokumen | Dokumen.java | Kelas Dokumen merupakan kelas *entity* yang bertujuan untuk menarik data mentah/data training. |
| 6 | ProsesManager | ProsesManager.java | Kelas DokumenManager merupakan kelas *control* yang menangani proses perhitungan klasifikasi. |
| 7 | DataManager | DataManager.java | Kelas DataManager merupakan kelas *control* yang menangani pemisahan data training dan data testing. |

#### 4.4.3.2 Implementasi Antarmuka

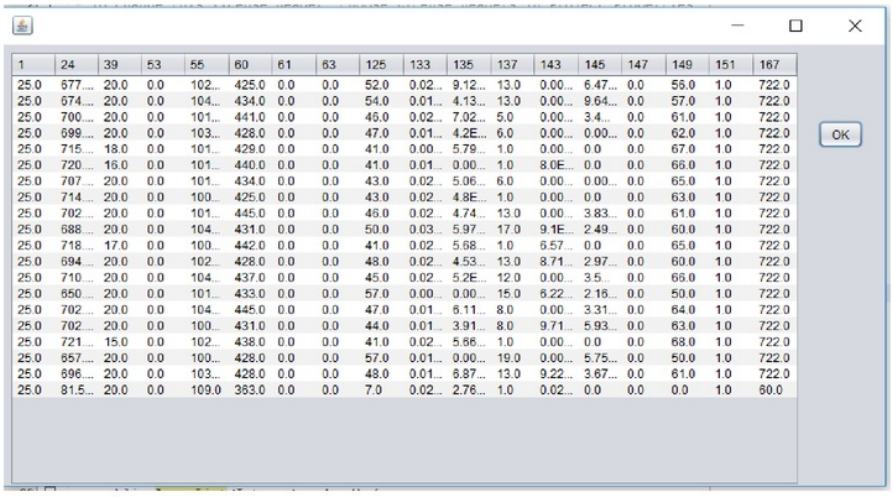
Implementasi antarmuka dilakukan berdasarkan perancangan antarmuka pada fase elaborasi. Gambar IV-

11 merupakan antarmuka menu utama utama perangkat lunak, gambar IV-

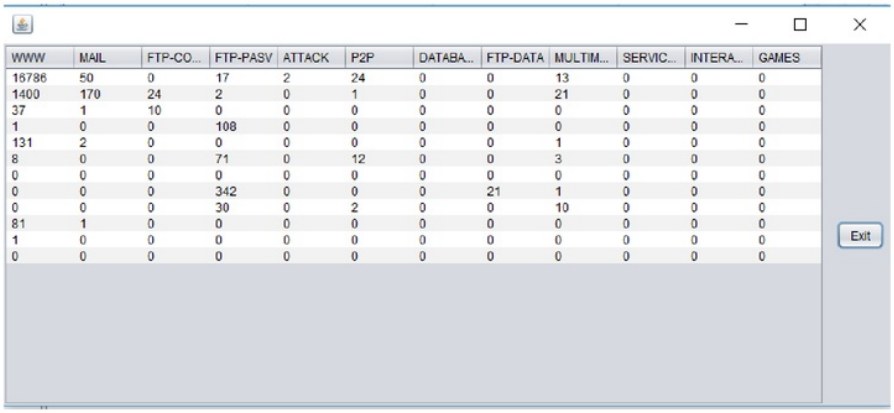
12 merupakan antarmuka *File Preview*, gambar IV-

13 merupakan antarmuka *OutputView*.


Gambar IV-11. Antarmuka Menu Utama Perangkat Lunak




Gambar IV-12. Antarmuka *File Preview*




Gambar IV-13. Antarmuka *OutputView*

## 4.5 Fase Transisi

Pada subbab ini dibahas mengenai pengujian dari perangkat lu nak klasifikasi data trafik internet yang telah dibangun. Pengujian dilakukan berdasarkan perangkat lunak hasil pengembangan difas e konstruksi.

### 4.5.1 Permodelan Bisnis

Pengujian perangkat lunak secara *blackbox* dan *whitebo x* dengan terlebih dahulu membuat rencana pengujian berdasar kan use case yang dibuat pada fase insepsi.

### 4.5.2 Kebutuhan Sistem

Lingkungan pengujian yang digunakan pada fase transis i adalah perangkat keras yang sama saat membangun perangk

at lunak klasifikasi data trafik internet dengan spesifikasi seba gai berikut:

1. Laptop merk ASUS GL553VE;
2. Processor Intel(R) Core(TM) i7-7700HQ CPU @ 2.80GHz;
3. RAM 8 GB;
4. Hard Disk 1 TB.

Sedangkan perangkat lunak yang digunakan untuk implementas i yaitu:

1. Windows 10 Home Single Language 64-bit;
2. Compiler Netbeans IDE 8.2;
3. Visual Paradigm UML Enterprise Edition V8.0.

### 4.5.3 Rencana Pengujian

Rencana pengujian pada perangkat lunak klasifikasi dat a trafik internet digambarkan dalam tabel-tabel. Kolom pada tabel meliputi identifikasi, pengujian, jenis pengujian, serta tingkat pengujian.

#### 4.5.3.1 Rencana Pengujian Use Case Melakukan Praprose s Data Trafik Internet

Tabel IV-9 menerangkan rencana pengujian memasukkan data perangkat lunak berdasarkan *Use Case*.

Tabel IV-

9. Rencana Pengujian *Use Case* Melakukan Praproses Data

| No | Identifikasi | Pengujian | Jenis Pengujian | Tingkat Pengujian |
|---|---|---|---|---|
| 1 | U-1-101 | Masukkan folder yang berisi file data trafik Internet berekstensi .xls. | *Black Box* | Pengujian Unit |
| 2 | U-1-102 | Melakukan keseluruhan prapengolahan data trafik Internet. | *White Box* | Pengujian Unit |

#### 4.5.3.2 Rencana Pengujian Use Case Melakukan Klasifikasi dengan Naïve Bayes

Tabel IV-

10 menerangkan rencana pengujian melakukan klasifikasi dengan Naïve Bayes pada perangkat lunak berdasarkan Use Case.

Tabel IV-

10. Rencana Pengujian Use Case Melakukan Klasifikasi dengan Naïve Bayes

| No | Identifikasi | Pengujian | Jenis Pengujian | Tingkat Pengujian |
|---|---|---|---|---|
| 1 | U-2-101 | Menekan tombol "Process" | *Black Box* | Pengujian Unit |
| 2 | U-2-102 | Melakukan klasifikasi data trafik internet dengan metode Naïve Bayes dan Feature Selection. | *White Box* | Pengujian Unit |

### 4.5.4 Implementasi

Berikut ini adalah kasus uji yang dilakukan terhadap perangkat lunak yang dibangun. Kasus uji dilakukan berdasarkan rencana uji yang telah dipaparkan sebelumnya.

#### 4.5.4.1 Pengujian Use Case memasukan Data Trafik Internet

Tabel IV-11 menerangkan pengujian memasukkan Data Trafik Internet pada perangkat lunak berdasarkan Use Case.

Tabel IV-

11. Pengujian Use Case Melakukan Praproses Data Trafik Internet

| Identifikasi | Deskripsi | Prosedur Pengujian | Masukan | Keluaran yang Diharapkan | Hasil yang Didapat | Kesimpulan |
|---|---|---|---|---|---|---|
| U-1-101 | Masukkan file data trafik internet berekstensi .xls. | Menekan tombol " Open File " | Tidak ada | Data yang sudah diproproses oleh sistem. | List data trafik internet dari file .xls. | Terpenuhi |
| U-1-102 | Masukan file data trafik internet bukan berekstensi .xls. | Menekan tombol " Open File " | Tidak ada | Data tidak berhasil dimuat. Kemudian, sistem menampilkan " File yang dimasukkan harus berekstensi .xls". | Data tidak berhasil dimuat. Kemudian, sistem menampilkan "File yang dimasukkan harus berekstensi .xls". | Terpenuhi |

#### 4.5.4.2 Pengujian Use Case Melakukan Klasifikasi dengan Naïve Bayes

Tabel IV-12 menerangkan pengujian klasifikasi Data Trafik Internet dengan C4.5 pada perangkat lunak berdasarkan Use Case.

Tabel IV-12. Rencana Pengujian Use Case Melakukan Klasifikasi dengan Naïve Bayes dan Feature Selection

| Identifikasi | Deskripsi | Prosedur Pengujian | Masukan | Keluaran yang Diharapkan | Hasil yang Didapat | Kesimpulan |
|---|---|---|---|---|---|---|
| U-2-101 | Menekan tombol "Process". | Menekan tombol "Process". | Trigger untuk mengaktifkan tombol | Tombol dapat ditekan dan mengaktifkan perhitungan klasifikasi | Tombol dapat ditekan dan mengaktifkan perhitungan klasifikasi | Terpenuhi |
| U-2-102 | Melakukan perhitungan klasifikasi Data Trafik Internet dengan metode Naïve Bayes dan Feature Selection | Menekan tombol "Process" | Masukan berupa hasil prapengolahan | Hasil klasifikasi sebanyak jumlah kelas yang sudah diketahui di awal | Hasil klasifikasi sebanyak jumlah kelas yang sudah diketahui di awal | Terpenuhi |

## 4.6 Kesimpulan

Dari proses pengembangan perangkat lunak, telah jelas diuraikan taha pan pengembangan alat penelitian yang akan membantu proses penelitian kl asifikasi data trafik internet. Pengembangan perangkat lunak ini telah disesu aikan dengan kebutuhan penelitian, yaitu pemrosesan data trafik internet, ma mpu melakukan *replace missing value*, melakukan proses klasifikasi dengan *Naïve Bayes* dan *Feature Selection* dalam pemilihan atribut, menghitung ni lai akurasi , *precision*, *recall* klasifikasi. Dijelaskan pula skenario dan alur pengembangan dan pengujian perangkat lunak sehingga dihasilkan perangkat lunak yang sesuai dengan kebutuhann penelitian. Selanjutnya setelah perang kat lunak selesai, dilakukan proses peengujian terhadap data uji serta melak ukan analisa dari hasil yang dihasilkan oleh alat penelitian.

# BAB V

# HASIL DAN ANALISIS PENELITIAN

## 5.1 Pendahuluan

Pada bab IV telah dilakukan pengembangan perangkat lunak yang menjadi alat penelitian klasifikasi *Network Traffic* menggunakan metode *Naive Bayes* dan *Feature Selection* dalam pemilihan atribut. Untuk membuktikan hasil klasifikasi, maka digunakan perhitungan akurasi, *precision,* dan *recall* untuk mengevaluasi hasil klasifikasi yang diperlukan untuk menyelasaikan proses klasifikasi. Penjelasan dan metode perhitungan akurasi, *precision,* dan *recall* dapat dilihat pada bab II.

## 5.2 Percobaan Penelitian

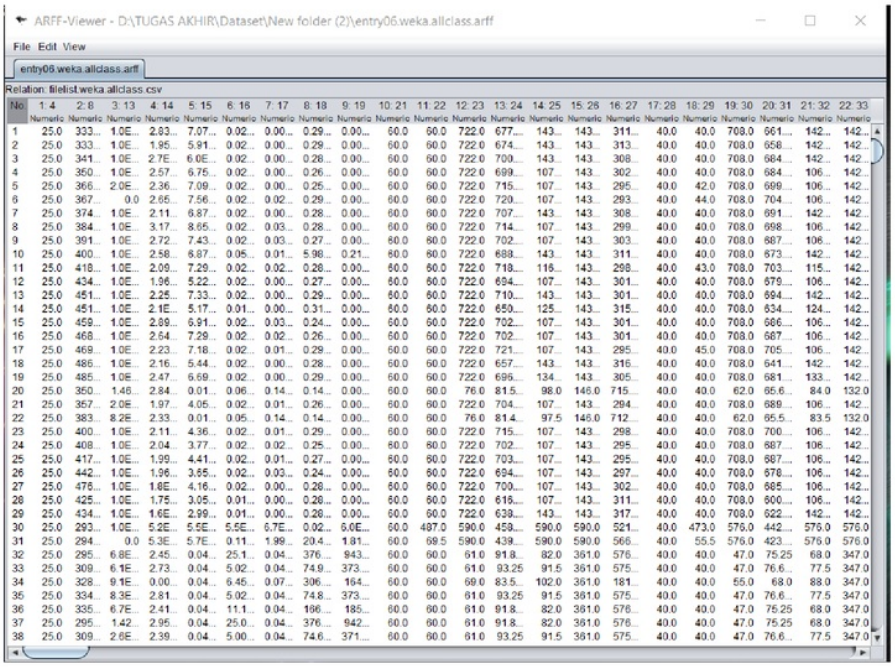
Pengujian dilakukan dengan menggunakan data uji berupa dokumen jurnal sebanyak 19384 buah dengan 248 atribut, data diunduh dari situs *Computer Laboratory University of Cambridge* (*http://www.cl.cam.ac.uk*) yang kemudian disalin dan disimpan dalam dokumen bereksistensi *.xls*. Proses pengujian dilakukan sesuai dengan arsitektur perangkat lunak yang telah dijabarkan pada subbab 4.2.3.1 hingga 4.2.3.3, yakni analisis kebutuhan perangkat lunak prapengolahan, dan desain perangkat lunak.

Setelah didapatkan hasil klasifikasi dari metode klasifikasi Naive Bayes, tahapan penelitian dilanjutkan dengan menganalisa hasil klasifikasi yang telah didapatkan.

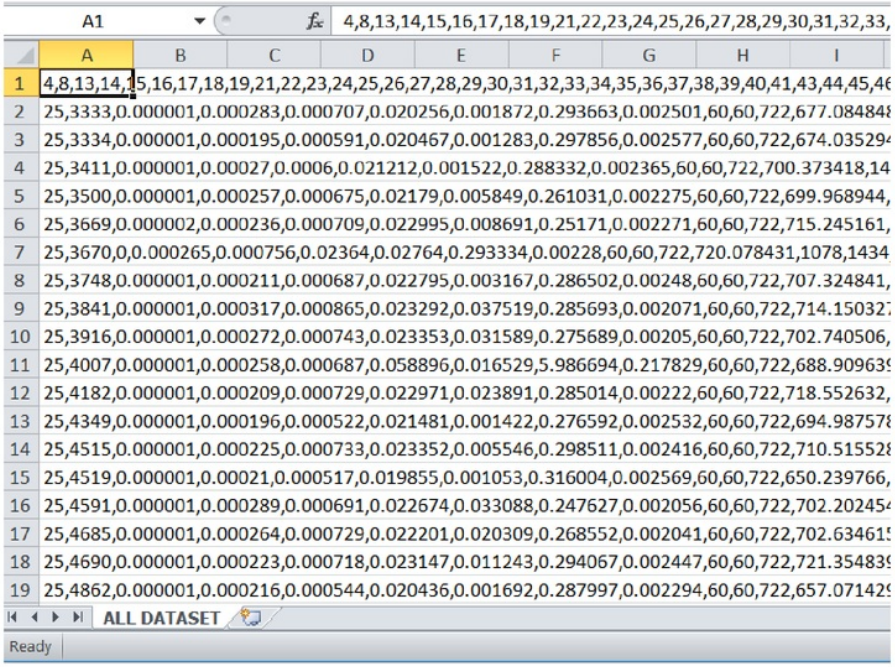
## 5.3 Hasil *Feature Extraction*

Bagian ini akan membaca atribut-atribut yang dihasilkan dari data trafik, dimana file data tersebut merupakan *raw data* yang sulit dibaca oleh manusia dikarenakan memiliki susunan yang unik dan juga adanya proses *encapsulated packet data.* Oleh karena itu untuk membaca data dari *raw data* dilakukan disebuah aplikasi yang bernama Weka, dengan tujuan untuk mendapat nilai-nilai dari semua atribut serta hasil *features extraction* diubah dari berekstensi *.arff* ke dalam bentuk file *.xls* yang akan digunakan untuk proses training karena dengan tipe file ini akan lebih mudah dalam proses *training.*

Pada gambar V-1 menunjukkan tampilan *raw data* saat masih berekstensi *.arff* dan gambar V-2 menunjukkan tampilan data trafik berekstensi *.xls.*




Gambar V-1. Tampilan *raw data* berekstensi *.arff*



Gambar V-2. Tampilan data trafik berekstensi *.xls*

## 5.4 Hasil *Feature Selection*

Pada bagian ini adalah pekerjaan utama yang dilakukan yaitu menentukan atribut-atribut yang relevan dari hasil data trafik yang didapat dari situs *Computer Laboratory University of Cambridge* (*http://www.cl.cam.ac.uk*) dengan menggunakan metode *feature rangking* yaitu *information Gain* (IG) untuk mencari rangking yang akan divalidasi dengan menggunakan metode klasifikasi *Naive Bayes*. Hasil dari klasifikasi akan dianalisis berdasarkan tingkat akurasi dari setiap atribut.

Bagian pertama yang dilakukan adalah *pre-processing* data yang didapat. Pada tahap ini adalah proses pembersihan data yang digunakan untuk menghilangkan data-data yang *error* sehingga didapat hasil akurasi yang baik. Tahap selanjutnya adalah membagi dataset menjadi dua bagian yaitu *training* sebesar 70% dan *testing* sebesar 30%, hal ini digunakan untuk proses pembelajaran untuk mendapat prediksi atribut-atribut yang kuat dalam menentukan pola sebuah paket data maaupun sebuah serangan. Hasil rangking atribut akan dijelaskan pada tabel V-1.

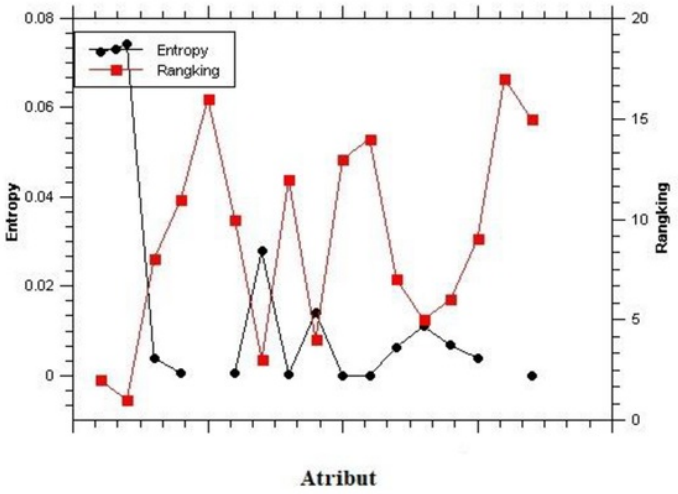
Tabel V-1. Hasil Rangking Atribut dengan menggunakan IG

| No | Atribut | Nomor Data | IG | |
|---|---|---|---|---|
| | | | Entropy | Rangking |
| 1 | 4 | 1 | 0,869 | 1 |

| | | | | |
|---|---|---|---|---|
| 2 | 35 | 24 | 0,496 | 2 |
| 3 | 180 | 167 | 0,470 | 3 |
| 4 | 137 | 125 | 0,387 | 4 |
| 5 | 145 | 133 | 0,361 | 5 |
| 6 | 149 | 137 | 0,329 | 6 |
| 7 | 73 | 61 | 0,316 | 7 |
| 8 | 75 | 63 | 0,293 | 8 |
| 9 | 65 | 53 | 0,283 | 9 |
| 10 | 67 | 55 | 0,283 | 10 |
| 11 | 163 | 151 | 0,280 | 11 |
| 12 | 51 | 39 | 0,275 | 12 |
| 13 | 157 | 145 | 0,246 | 13 |
| 14 | 155 | 143 | 0,245 | 14 |
| 15 | 161 | 149 | 0,244 | 15 |
| 16 | 72 | 60 | 0,241 | 16 |
| 17 | 159 | 147 | 0,237 | 17 |
| 18 | 147 | 135 | 0,231 | 18 |

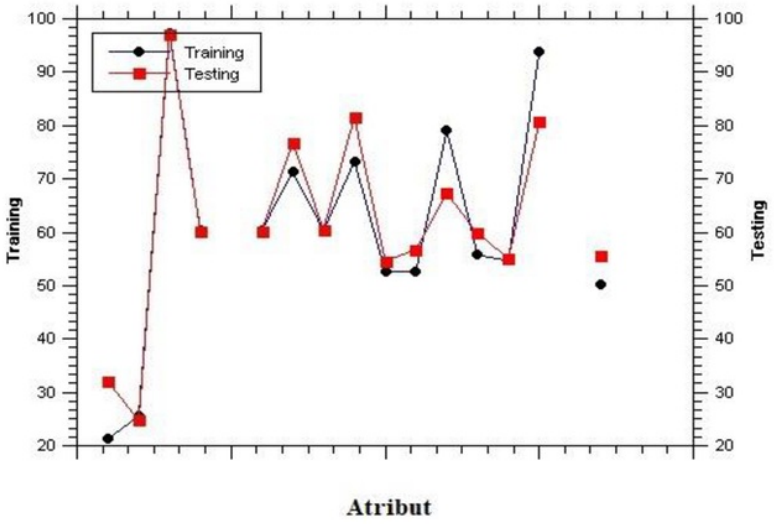
Tabel V-

1 menampilkan hasil *entropy* dan rangking dari setiap atribut dengan mengg unakan metode *information Gain* (IG). Hasil rangking didapatkan bahwa atri but 4 merupakan rangking pertama ini artinya atribut ini mempunyai kontri busi yang besar terhadap atribut yang dijadikan *class*, sedangkan tiga atribut yaitu 72, 159, dan 147 merupakan tiga rangking terendah, ini artinya atrib ut-atribut ini tidak mempunyai kontribusi.




Gambar V-3. Hasil Rangking dan Nilai *Entropy* Atribut

Tahap selanjutnya adalah melakukan proses validasi atribut dengan menggunakan metode klasifikasi *Naïve Bayes*. Proses pengujian ini dengan cara membagi *raw data* menjadi 2 bagian yaitu : 70 % untuk data Training dan 30 % untuk data Testing. Total data yang digunakan sebanyak 19384, maka untuk data *training* didapatkan sebanyak 13568 sedang data untuk testing sebanyak 5816. Tabel V-2 merupakan hasil pengujian *training* dan *testing* akan diuji untuk mendapatkan nilai validasi setiap atribut yang nantinya akan digunakan sebagai pengetahuan baru (*knowledge*) pada proses selanjutnya.




Gambar V-4. Hasil Validasi Atribut dengan *Naive Bayes*

Tabel V-2. Hasil *Training* dan *Testing*

| No | Atribut | IG | |
|---|---|---|---|
| | | Training | Testing |
| 1 | 147 | 0,0000 | 0,0000 |
| 2 | 161 | 21.3208 | 32.0856 |
| 3 | 155 | 25.4973 | 24.7892 |
| 4 | 4 | 97.2942 | 97.0525 |
| 5 | 73 | 60.4554 | 60.265 |
| 6 | 159 | 0,0000 | 0,0000 |
| 7 | 75 | 60.4554 | 60.265 |
| 8 | 137 | 71.2501 | 76.8213 |
| 9 | 149 | 60.3045 | 60.5209 |

| | | | |
|---|---|---|---|
| 10 | 180 | 73.2959 | 81.5143 |
| 11 | 51 | 52.6728 | 54.6164 |
| 12 | 67 | 52.6728 | 56.7231 |
| 13 | 145 | 79.0799 | 67.2691 |
| 14 | 65 | 55.7368 | 59.7512 |
| 15 | 163 | 54.662 | 55.1056 |
| 16 | 35 | 93.8767 | 80.6785 |
| 17 | 72 | 0,0000 | 0,0000 |
| 18 | 157 | 50.2074 | 55.7061 |

### 5.5 Hasil Klasifikasi Dengan Naive Bayes

Tabel V-3 menunjukkan hasil klasifikasi pada metode Naive Bayes. Proses pengujian dilakukan sebanyak 10 kali yang kemudian didapatkan nilai akurasi, precisi on, dan recall.

Tabel V-3. Tabel Hasil Klasifikasi

| | TP Rate | FP Rate | Precision | Recall | F-Measure | MCC | ROC Area | PRC Area | Class |
|---|---|---|---|---|---|---|---|---|---|
| | 0,994 | 0,666 | 0,91 | 0,994 | 0,95 | 0,511 | 0,950 | 0,987 | WWW |
| | 0,105 | 0,003 | 0,759 | 0,105 | 0,185 | 0,264 | 0,950 | 0,783 | MAIL |
| | 0,208 | 0,001 | 0,294 | 0,208 | 0,244 | 0,246 | 0,943 | 0,127 | FTP-CONTROL |
| | 0,991 | 0,024 | 0,189 | 0,991 | 0,318 | 0,428 | 0,996 | 0,676 | FTP-PASV |
| | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 | -0,001 | 0,552 | 0,008 | ATTACK |
| | 0,128 | 0,001 | 0,308 | 0,128 | 0,128 | 0,196 | 0,936 | 0,375 | P2P |
| | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | DATABASE |
| | 0,058 | 0 | 1 | 0,058 | 0,058 | 0,238 | 0,997 | 0,965 | FTP-DATA |
| | 0,238 | 0,002 | 0,204 | 0,238 | 0,238 | 0,219 | 0,997 | 0,323 | MULTIMEDIA |
| | 0 | 0 | 0 | 0 | 0 | 0 | 0,948 | 0,057 | SERVICES |
| | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0,333 | INTERACTIVE |
| | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | GAMES |

## 5.6 Analisis Hasil Penelitian

Tabel V-3 menunjukkan hasil dari klasifikasi *Naive Bayes*, terlihat bahwa *class* yang nilai atributnya banyak muncul adalah WWW. Sedangkan *class* DATABASE, INTERACTIVE, dan GAMES nilai atributnya yang banyak muncul cenderung sedikit dan bahkan tidak ada. Maka dapat dipastikan bahwa pada data trafik yang dipakai oleh penulis dua dari ketiga *class* tersebut tidak ada aktivitas yang terjadi di trafik yang dalam hal ini adalah *Computer Laboratory University of Cambridge* (*http://www.cl.cam.ac.uk*). Untuk penjelasan lebih detail dalam *Confusion Matrix* akan ditampilkan pada gambar V-3.

```
=== Confusion Matrix ===

     a     b     c     d     e     f     g     h     i     j     k     l   <-- classified as
 16786    50     0    17     2    24     0     0    13     0     0     0 |     a = WWW
  1400   170    24     2     0     1     0     0    21     0     0     0 |     b = MAIL
    37     1    10     0     0     0     0     0     0     0     0     0 |     c = FTP-CONTROL
     1     0     0   108     0     0     0     0     0     0     0     0 |     d = FTP-PASV
   131     2     0     0     0     0     0     0     1     0     0     0 |     e = ATTACK
     8     0     0    71     0    12     0     0     3     0     0     0 |     f = P2P
     0     0     0     0     0     0     0     0     0     0     0     0 |     g = DATABASE
     0     0     0   342     0     0     0    21     1     0     0     0 |     h = FTP-DATA
     0     0     0    30     0     2     0     0    10     0     0     0 |     i = MULTIMEDIA
    81     1     0     0     0     0     0     0     0     0     0     0 |     j = SERVICES
     1     0     0     0     0     0     0     0     0     0     0     0 |     k = INTERACTIVE
     0     0     0     0     0     0     0     0     0     0     0     0 |     l = GAMES
```

Gambar V-5. Gambar *Confusion Matrix* Hasil Penelitian

## 5.7 Kesimpulan

Dari hasil pengujian yang telah didapatkan dan dilakukaan tahapan analisis hasil klasifikasi, telah jelas diuraikan hasil masing-masing pengujian dari proses klasifikasi dan pemilihan atribut dengan *Naive Bayes* dan *Feature Selection*. Disimpulkan, banyaknya data sangat berpengaruh dalam besarnya nilai akurasi pengklasifikasian, semakin banyak data maka tingkat akurasi akan lebih baik saat pengujian.

# BAB VI

# KESIMPULAN DAN SARAN

## 6.1 Pendahuluan

Pada bab ini penulis memberikan kesimpulan dan saran yang diharapkan dapat menjadi acuan bagi penelitian selanjutnya di bidang yang sama.

## 6.2 Kesimpulan

Berdasarkan analisis dan pembahasan maka dapat ditarik beberapa kesimpulan yaitu sebagai berikut:

1. Di dalam dataset trafik internet yang digunakan oleh penulis nilai atribut yang banyak muncul pada *class* SERVICES, DATABASE, INTERACTIVE, dan GAMES cenderung sedikit dan bahkan tidak ada.
2. Hasil *feature selection* dengan menggunakan metode *feature rangking* didapatkan nilai entropy tertinggi adalah 0.86914856 untuk atribut 4. Untuk validasi atribut nilai tertinggi pada atribute 35 sebesar 97.2942 %.
3. Hasil klasifikasi *Naive Bayes* berdasarkan *class* dihasilkan nilai akurasi, *precision* dan *recall* tertinggi pada *class* WWW (0.994, 0.91, 0.994).

## 6.3 Saran

Saran yang diajukan oleh penulis, adalah:

1. Pada riset selanjutnya diharapkan dapat melakukan proses visualisasi pemrosesan data secara real-time agar lebih banyak lagi fitur-fitur yang bisa dideteksi.
2. Dapat menemukan solusi untuk meningkatkan kualitas klasifikasi dengan *Naive Bayes* dan *Feature Selection*.
3. Melakukan percobaan dengan metode klasifikasi dan teknik-teknik lain untuk data trafik internet.

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

ORIGINALITY REPORT

| 15% | 8% | 0% | 16% |
|---|---|---|---|
| SIMILARITY INDEX | INTERNET SOURCES | PUBLICATIONS | STUDENT PAPERS |

PRIMARY SOURCES

| | | |
|---|---|---|
| 1 | Submitted to Sriwijaya University<br>Student Paper | 14% |
| 2 | chemnitzer.linux-tage.de<br>Internet Source | 1% |

| | | | |
|---|---|---|---|
| Exclude quotes | On | Exclude matches | < 1% |
| Exclude bibliography | On | | |