

**PERBANDINGAN REDUKSI DIMENSI *SINGULAR VALUE DECOMPOSITION* DAN *PRINCIPAL COMPONENT ANALYSIS*
PADA KLASIFIKASI DATA TRAFIK INTERNET**

Diajukan Sebagai Syarat Untuk Menyelesaikan
Pendidikan Program Strata-1 Pada
Jurusan Teknik Informatika



Oleh:

Shofi Salsabila
NIM : 09021381419068

**Jurusan Teknik Informatika
FAKULTAS ILMU KOMPUTER UNIVERSITAS SRIWIJAYA
2018**

LEMBAR PENGESAHAN TUGAS AKHIR

**PERBANDINGAN PENGARUH REDUKSI DIMENSI
SINGULAR VALUE DECOMPOSITION DAN PRINCIPAL
COMPONENT ANALYSIS PADA KLASIFIKASI DATA TRAFIK
INTERNET**

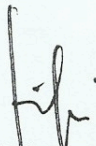
Oleh :

Shofi Salsabila
NIM : 09021381419068

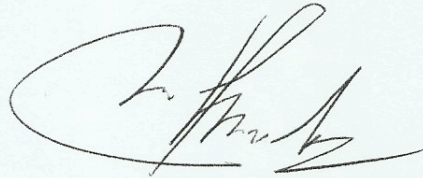
Palembang, 13 Desember 2018

Pembimbing 1

Pembimbing 2,

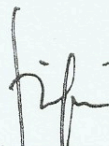


Rifkie Primartha, M.T.
NIP. 197706012009121004



Ir. Muhammad Ihsan Jambak, M.Sc.
NIPUS. 196804052015109101

Mengetahui,
Ketua Jurusan Teknik Informatika



Rifkie Primartha, M.T.
NIP. 197706012009121004

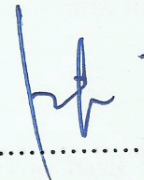
TANDA LULUS UJIAN SIDANG TUGAS AKHIR

Pada hari Jum'at tanggal 13 Desember 2018 telah dilaksanakan ujian sidang tugas akhir oleh Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

Nama : Shofi Salsabila
NIM : 09021381419068
Judul : Perbandingan Pengaruh Reduksi Dimensi *Singular Value Decomposition* dan *Principal Component Analysis* pada Klasifikasi Data Trafik Internet

1. Pembimbing I

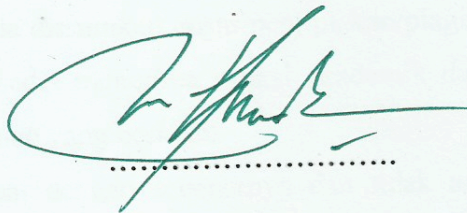
Rifkie Primartha, M. T.
NIP. 197706012009121004



.....

2. Pembimbing II

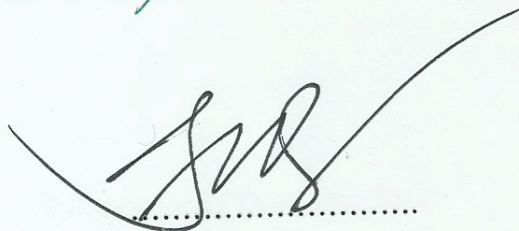
Ir. Muhammad Ihsan Jambak, M.Sc.
NIPUS. 196804052015109101



.....

3. Penguji I

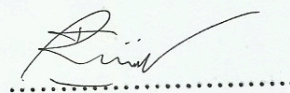
Rusdi Effendi, M. Kom.
NIPUS 198201022015109000



.....

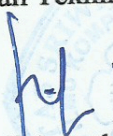
4. Penguji II

Mastura Diana Marieska, S.T., M. T.
NIP. 198603212018032001



.....

Mengetahui,
Ketua Jurusan Teknik Informatika



Rifkie Primartha, M. T.
NIP. 197706012009121004

HALAMAN PERNYATAAN

Yang bertanda tangan di bawah ini :

Nama : Shofi Salsabila
NIM : 09021381419068
Program Studi : Teknik Informatika (Bilingual)
Judul Skripsi : Perbandingan Pengaruh Reduksi Dimensi *Singular Value Decomposition* dan *Principal Component Analysis* pada Klasifikasi Data Trafik Internet

Hasil Pengecekan Software *iThenticate/Turnitin* : 16 %

Menyatakan bahwa Laporan Proyek saya merupakan hasil karya sendiri dan bukan hasil penjiplakan/plagiat. Apabila ditemukan unsur penjiplakan/plagiat dalam laporan proyek 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, 13 Desember 2018



Shofi Salsabila
NIM. 09021381419068

**IF THERE IS NO STRUGGLE,
THERE IS NO PROGRESS.**

Karya tulis ini di dedikasikan untuk:

- Orang tuaku tersayang,
- Pembimbing, dan
- Sahabat-sahabatku.

COMPARISON OF THE EFFECT OF DIMENSIONAL REDUCTION USING SVD
AND PCA FOR INTERNET TRAFFIC DATA CLASSIFICATION

by:

Shofi Salsabila
09021381419068

ABSTRACT

High-dimensional data is a data that has many attributes, one of them is internet traffic data. This research used internet traffic data with 248 attributes. If the internet traffic data is going to be classified, a dimensional reduction technique is needed, because conventional classification algorithms work better in handling low dimensional data. Dimension reduction techniques are classified into 2 types, feature selection and feature extraction. This study will compare the implementation of the Singular Value Decomposition (SVD) algorithm as a feature selection and Principal Component Analysis (PCA) technique as an feature extraction technique for C4.5 classification algorithm. The results obtained by ANOVA shows insignificant differences on the value of accuracy, precision, and recall. However, in terms of computation time, the combination of PCA and C4.5 is proven to be slower than the other two methods.

Keywords: high dimensional data, *Internet traffic*, dimension reduction, *singular value decomposition*, *principal component analysis*, *C4.5 classification*.

PERBANDINGAN PENGARUH REDUKSI DIMENSI SVD DAN PCA PADA
KLASIFIKASI DATA TRAFIK INTERNET

Oleh:

Shofi Salsabila
09021381419068

ABSTRAK

Data berdimensi tinggi adalah data yang memiliki banyak atribut. Pada penelitian ini, data trafik internet yang digunakan memiliki 248 atribut, maka dari itu data tersebut tergolong data berdimensi tinggi. Apabila ingin dilakukan pengelompokan terhadap data trafik internet tersebut, dibutuhkan teknik reduksi dimensi, karena algoritma pengelompokan konvensional bekerja lebih baik dalam menangani data berdimensi rendah. Teknik reduksi dimensi tergolong menjadi 2 jenis, seleksi fitur dan ekstraksi fitur. Penelitian ini akan membandingkan penerapan algoritma Singular Value Decomposition (SVD dan Principal Component Analysis (PCA) sebagai teknik reduksi dimensi terhadap algoritma klasifikasi C4.5. Hasil yang didapatkan melalui pengujian ANOVA menunjukkan tidak ada perbedaan signifikan pada nilai akurasi, precision, dan recall. Namun, jika dilihat dari waktu komputasi, klasifikasi C4.5 yang didahului oleh PCA terbukti lebih lambat 10 detik dibandingkan kedua metode pengujian lainnya.

Kata Kunci: Data Berdimensi Tinggi, Data Trafik Internet, Reduksi Dimensi, *Singular Value Decomposition*, *Principal Component Analysis*, klasifikasi C4.5

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, Ir. Oslan dan Rakhmayanti Raihana S.E., yang selalu mendokan 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, S.T., M.T. selaku Ketua Jurusan Teknik Informatika dan dosen pembimbing I tugas akhir, dan dosen pembimbing akademik yang selalu memberikan pembinaan dalam proses perkuliahan dan pengerjaan Tugas Akhir.
4. Bapak Ir. M. Ihsan Jambak, M.Sc. selaku dosen pembimbing II yang telah membimbing, mengarahkan dan memberikan motivasi dalam proses perkuliahan dan pengerjaan Tugas Akhir.
5. Bapak Rusdi Effendi M.Kom. selaku dosen penguji I yang telah memberikan arahan serta dukungan dalam proses perkuliahan dan pengerjaan Tugas Akhir.
6. Ibu Mastura Diana Marieska, S.T., M.T. selaku dosen penguji II yang telah memberikan masukan dan dorongan dalam proses pengerjaan Tugas Akhir.
7. Seluruh dosen Program Studi Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

8. Seluruh staf tata usaha yang telah membantu dalam kelancaran proses administrasi dan akademik selama masa perkuliahan.
9. Kak Dini dan Inayah, yang telah menjadi tempat belajar dan berbagi selama tugas akhir ini.
10. Teman-teman IF Reguler dan Bilingual, yang selalu berjuang bersama dalam menempuh ilmu, terutama Puri, Winda, dan Joko.
11. Anda dan Citra yang telah menemani kehidupan perantauan dan menjadi tempat berbagi cerita. Kehidupan perantauan semakin mengasyikkan dengan celotehan kalian.
12. Anggota AIESEC UNSRI – MUSI 15/18 terutama Athena, Avengers, dan DnC, yang telah mengisi hari-hari dan selalu menyemangati.

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, Desember 2018

Shofi Salsabila

DAFTAR ISI

	Halaman
HALAMAN JUDUL	i
HALAMAN PENGESAHAN	ii
HALAMAN TANDA LULUS UJIAN SIDANG TUGAS AKHIR	iii
HALAMAN PERNYATAAN	iv
MOTTO DAN PERSEMBAHAN	v
ABSTRACT	vi
ABSTRAK	vii
KATA PENGANTAR	viii
DAFTAR ISI	x
DAFTAR TABEL	xv
DAFTAR GAMBAR	xviii
DAFTAR LAMPIRAN	xix
BAB I PENDAHULUAN	
1.1 Pendahuluan	I-1
1.2 Latar Belakang	I-1
1.3 Perumusan Masalah.....	I-4
1.4 Tujuan Penelitian.....	I-4
1.5 Manfaat Penelitian.....	I-5
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 Reduksi Dimensi	II-1
2.2.1 Singular Value Decomposition	II-2

2.2.1 Principal Component Analysis	II-4
2.3 Klasifikasi C4.5	II-6
2.4 Cross Validation	II-8
2.5 Confusion Matrix	II-9
2.6 Pengujian ANOVA	II-10
2.7 Penelitian Lain yang Relevan	II-11
2.7.11 Li Jun, Zhang Shunyi, Lu Yanqing, Zhang Zailong (2007)	II-11
2.7.2 K. Keerthi Vasan dan B. Surendiran(2016)	II-12
2.8 Kesimpulan	II-13

BAB III METEDOLOGI 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-7
3.4.5 Melakukan Pengujian Penelitian.....	III-8
3.4.6 Melakukan Analisa Hasil Pengujian dan Membuat Kesimpulan	III-9
3.5 Metode Pengembangan Perangkat Lunak	III-9
3.5.1 Rational Unified Process.....	III-9
3.5.2 Fase Insepsi.....	III-11
3.5.3 Fase Elaborasi	III-11
3.5.4 Fase Konstruksi.....	III-12
3.5.5 Fase Transisi	III-12

3.6 Manajemen Proyek Penelitian.....	III-13
3.7 Kesimpulan.....	III-16

BAB IV PENGEMBANGAN PERANGKAT LUNAK

4.1 Pendahuluan	IV-1
4.2 Fase Insepsi	IV-1
4.2.1 Permodelan Bisnis	IV-2
4.2.2 Kebutuhan Sistem	IV-3
4.2.2.1 Fitur Prapengolahan	IV-3
4.2.2.2 Fitur Reduksi Dimensi	IV-4
4.2.2.3 Fitur Klasifikasi.....	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-5
4.2.3.2 Analisis Data	IV-7
4.2.3.3 Analisis Prapengolahan	IV-7
4.2.3.4 Analisis Singular Value Decomposition	IV-8
4.2.3.5 Analisis Principal Component Analysis.....	IV-9
4.2.3.6 Analisis Klasifikasi C4.5.....	IV-11
4.2.3.7 Desain Perangkat Lunak.....	IV-12
4.3 Fase Elaborasi.....	IV-23
4.3.1 Permodelan Bisnis	IV-23
4.2.3.1 Perancangan Data	IV-23
4.2.3.2 Perancangan Antarmuka	IV-23
4.3.2 Kebutuhan Sistem	IV-24
4.3.3 <i>Sequence Diagram</i>	IV-25
4.4 Fase Konstruksi	IV-30
4.4.1 Kebutuhan Sistem	IV-30
4.4.2 Diagram Kelas	IV-30
4.4.3 Implementasi.....	IV-32
4.4.3.1 Implementasi Kelas	IV-32

4.4.3.2 Implementasi Antarmuka	IV-33
4.5 Fase Transisi.....	IV-33
4.5.1 Permodelan Bisnis	IV-34
4.5.2 Kebutuhan Sistem	IV-34
4.5.3 Rencana Pengujian.....	IV-34
4.5.3.1 Rencana Use Case melakukan Praproses Data Trafik Internet	IV-35
4.5.3.2 Rencana Use Case melakukan Klasifikasi dengan C4.5	IV-35
4.5.3.3 Rencana Use Case Mereduksi dengan SVD dan Klasifikasi dengan C4.5	IV-35
4.5.3.4 Rencana Use Case Mereduksi dengan PCA dan Klasifikasi dengan C4.5	IV-36
4.5.4 Implementasi.....	IV-36
4.5.4.1 Pengujian Use Case Memasukan Dokumen	IV-37
4.5.4.2 Pengujian Use Case melakukan Klasifikasi dengan C4.5.....	IV-37
4.5.4.3 Pengujian Use Case Mereduksi dengan SVD dan Klasifikasi dengan C4.5	IV-37
4.5.4.4 Pengujian Use Case Mereduksi dengan PCA dan Klasifikasi dengan C4.5	IV-38
4.6 Kesimpulan.....	IV-39

BAB V HASIL DAN ANALISIS PENELITIAN

5.1 Pendahuluan	V-1
5.2 Hasil Percobaan Penelitian.....	V-1
5.3 Hasil Klasifikasi Dengan C4.5	V-2
5.4 Analisa Penelitian.....	V-4
5.4.1 One Way ANOVA – Nilai Akurasi	V-4
5.4.2 One Way ANOVA – Nilai Precision	V-6
5.4.3 One Way ANOVA – Nilai Recall	V-7

5.4.4 One Way ANOVA – Waktu Komputasi	V-8
5.5 Perbandingan Hasil Penelitian.....	V-11
5.6 Kesimpulan.....	V-13
 BAB VI KESIMPULAN DAN SARAN	
6.1 Pendahuluan	VI-1
6.2 Kesimpulan	VI-1
6.3 Saran	VI-2
 DAFTAR PUSTAKA.....	 xx

DAFTAR TABEL

	Halaman
Tabel II-1 Contoh Tabel Confusion Matrix	II-9
Tabel III-1 Rancangan Tabel Confusion Matrix untuk Setiap Hasil Pengujian.....	III-6
Tabel III-2 Rancangan Tabel Hasil Pengujian Untuk Setiap Hasil Klasifikasi	III-6
Tabel III-3 Rancangan Tabel Hasil Rata-rata Perbandingan Klasifikasi Dengan dan Tanpa Reduksi Dimensi.....	III-7
Tabel III-4 Penjadwalan Penelitian dalam Bentuk <i>Work Breakdown Structure (WBS)</i>	III-14
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 Tabel Matrik U (7 x 2)	IV-8
Tabel IV-6 Tabel Matrik S (2 x 2)	IV-8
Tabel IV-7 Tabel Matrik V (2 x 5)	IV-9
Tabel IV-8 Hasil Reduksi Dimensi SVD.....	IV-9
Tabel IV-9 Tabel Hasil Normalisasi (7 x 5)	IV-10
Tabel IV-10 Tabel Hasil Matriks Kovarian (5 x 5)	IV-10
Tabel IV-11 Tabel Matriks U dari Masukan Matriks Kovarian	IV-10
Tabel IV-12 Hasil Principal Component dari PCA	IV-11
Tabel IV-13 Hasil Klasifikasi C4.5	IV-12
Tabel IV-14 Tabel Definisi Aktor Use Case.....	IV-13
Tabel IV-15 Definisi <i>Use Case</i>	IV-13
Tabel IV-16 Skenario <i>Use Case</i> Melakukan Praproses Data Trafik Internet..	IV-14
Tabel IV-17 Skenario <i>Use Case</i> Melakukan Klasifikasi dengan C4.5.....	IV-16
Tabel IV-18 Skenario <i>Use Case</i> Mereduksi dimensi dengan SVD dan Klasifikasi dengan C45.....	IV-17

Tabel IV-19 Skenario <i>Use Case</i> Mereduksi Data Trafik Internet dengan PCA dan Klasifikasi dengan C4.5	IV-18
Tabel IV-20 Implementasi Kelas	IV-32
Tabel IV-21 Rencana Pengujian <i>Use Case</i> Melakukan Praproses Data	IV-35
Tabel IV-22 Rencana Pengujian <i>Use Case</i> Melakukan Klasifikasi dengan C4.5	IV-35
Tabel IV-23 Rencana Pengujian <i>Use Case</i> Mereduksi dengan SVD dan Klasifikasi dengan C4.5	IV-36
Tabel IV-24 Rencana Pengujian <i>Use Case</i> Mereduksi dengan PCA dan Klasifikasi dengan C4.5	IV-36
Tabel IV-25. Pengujian <i>Use Case</i> Melakukan Praproses Data Trafik Internet	IV-37
Tabel IV-26 Rencana Pengujian <i>Use Case</i> Melakukan Klasifikasi dengan C4.5	IV-37
Tabel IV-27 Rencana Pengujian <i>Use Case</i> Mereduksi dengan SVD dan Klasifikasi dengan C4.5	IV-38
Tabel IV-28 Rencana Pengujian <i>Use Case</i> Mereduksi dengan PCA dan Klasifikasi dengan C4.5	IV-38
Tabel V-1 Tabel Hasil Klasifikasi	V-3
Tabel V-2 Tabel Deskripsi Nilai Akurasi	V-5
Tabel V-3 Tabel Hasil Uji ANOVA Nilai Akurasi	V-5
Tabel V-4 Tabel Deskripsi Nilai <i>Precision</i>	V-6
Tabel V-5 Tabel Hasil Uji ANOVA Nilai <i>Precision</i>	V-7
Tabel V-6 Tabel Deskripsi Nilai <i>Recall</i>	V-7
Tabel V-7 Tabel Hasil Uji ANOVA Nilai <i>Recall</i>	V-8
Tabel V-8 Tabel Deskripsi Waktu Komputasi	V-9
Tabel V-9 Tabel Hasil Uji ANOVA Waktu Komputasi	V-9
Tabel V-10 Tabel Hasil Uji Post Hoc Tukey Waktu Komputasi	V-10

Tabel V-11 Tabel Rangkuman Uji Post Hoc Tukey Waktu Komputasi..... V-11

DAFTAR GAMBAR

	Halaman
Gambar II-1 Representasi dari <i>Singular Value Decomposition</i>	II-2
Gambar III-1 Tahapan Pengujian Penelitian.....	III-8
Gambar III-2 Arsitektur RUP	III-10
Gambar IV-1 Diagram Use Case Current Existing.....	IV-2
Gambar IV-2 Diagram <i>Use Case</i>	IV-12
Gambar IV-3 Diagram Aktivitas Use Case Melakukan Praproses Data Trafik Internet.....	IV-19
Gambar IV-4 Diagram Aktivitas Use Case Melakukan Klasifikasi dengan C4.5.....	IV-20
Gambar IV-5 Diagram Aktivitas Use Case Mereduksi dengan SVD dan Klasifikasi dengan C4.5.....	IV-21
Gambar IV-6 Diagram Aktivitas Use Case Mereduksi dengan PCA dan Klasifikasi dengan C4.5.....	IV-22
Gambar IV-7 Rancangan Antarmuka Menu Utama	IV-24
Gambar IV-8 <i>Sequence Diagram</i> Melakukan Praproses Data Trafik Internet	IV-26
Gambar IV-9 <i>Sequence Diagram</i> Melakukan Klasifikasi dengan C4.5	IV-27
Gambar IV-10 <i>Sequence Diagram</i> Mereduksi dengan SVD dan Klasifikasi dengan C4.5	IV-28
Gambar IV-11 <i>Sequence Diagram</i> Mereduksi dengan PCA dan Klasifikasi dengan C4.5	IV-29
Gambar IV-12 Kelas Diagram Perangkat Lunak.....	IV-31
Gambar IV-13 Antarmuka Halaman Utama Perangkat Lunak.....	IV-33
Gambar V-1 Perbandingan Hasil Penelitian	V-12

DAFTAR LAMPIRAN

	Halaman
LAMPIRAN 1 Nama Atribut.....	L-1
LAMPIRAN 2 Perhitungan Manual SVD	L-9
LAMPIRAN 3 Perhitungan Manual PCA	L-11
LAMPIRAN 4 Perhitungan Manual C4.5	L-13
LAMPIRAN 5 Kodingan.....	L-16

Bab 1

Pendahuluan

1.1 Pendahuluan

Bab ini membahas mengenai latar belakang penelitian perbandingan algoritma *Principal Component Analysis* (PCA) dan *Singular Value Decomposition* (SVD) dalam reduksi fitur untuk klasifikasi data trafik internet yang kemudian dirumuskan menjadi rumusan masalah yang memiliki tujuan dan manfaat.

1.2 Latar Belakang Masalah

Kita hidup di dunia di mana sejumlah besar data dikumpulkan setiap hari. Menganalisis data tersebut merupakan kebutuhan yang penting, supaya data tersebut kaya akan informasi dan dapat membantu untuk mengambil keputusan (Gorunescu, 2011). Serangkaian proses analisis untuk mendapatkan informasi yang berguna dari basis data yang besar untuk membantu dalam pengambilan keputusan disebut dengan *data mining* (Stanton, 2013). Trafik internet menghasilkan sejumlah besar data yang membantu pengelola jaringan untuk mengelola dan mempelajari berbagai karakteristik trafik internet (Andersen & Feamster, 2006). Ketersediaan data yang ada banyak mengandung dimensi atau fitur. Memungkinkan menjadi data yang kompleks dengan banyak *noise*, *outliner*, dan elemen yang hilang (*missing value*) serta entitas data yang tidak berhubungan (Chen et al. 2009; Tajunisha & Saravanan, 2010) sehingga data tersebut

dikelompokkan sebagai data berdimensi tinggi. Menggunakan seluruh fitur untuk sebuah algoritma *machine learning* menjadi pilihan yang kurang tepat, karena tidak semua fitur data relevan dengan *machine learning* tersebut (Adi Suryaputra P., Febriliyan Samopa, 2014), maka dari itu untuk meningkatkan akurasi dibutuhkan reduksi dimensi saat pengolahan data awal (Addison et al. 2003; Gorunescu, 2011).

Teknik reduksi dimensi tergolong menjadi dua yaitu seleksi fitur dan ekstraksi fitur. Seleksi fitur bertujuan menemukan fitur terbaik pada dimensi untuk mendapatkan informasi dan membuang fitur yang tidak signifikan pada dimensi lain (v. Kumar, 2014). Ekstraksi fitur bertujuan untuk mencari satu set fitur yang baru di dimensi yang merupakan kombinasi ukuran asli dengan cara mengubah dimensi tinggi ke dimensi yang lebih rendah. Salah satu seleksi fitur yang umum digunakan adalah algoritma *Singular Value Decomposition* (Wall, 2003), sedangkan ekstraksi fitur menerapkan *Principal Component Analysis* (Pechenizkiy, 2003).

Penelitian oleh Wibawa dan Novianti (2017) menerapkan algoritma *k-nearest neighbor* dan beberapa teknik reduksi ekstraksi fitur, salah satunya *Principal Component Analysis* (PCA), pada data citra *Fine Needle Aspiration* (FNA) untuk optimalisasi klasifikasi tumor payudara. Penelitian ini menggunakan tiga teknik reduksi ekstraksi fitur yang dimana *Principal Component Analysis* (PCA) memiliki tingkat akurasi yang paling tinggi di antara ketiganya, yaitu 97,36%. Hasil penelitian ini menunjukkan reduksi dimensi dapat meningkatkan klasifikasi data citra yang kemudian digunakan untuk mendeteksi tumor payudara.

Algoritma PCA mampu menghilangkan korelasi 100% hingga mampu menghasilkan data ekstrak yang signifikan (Paukkeri et al. 2011)

Selanjutnya penelitian yang dilakukan oleh Ruangpainsarn (2015) menerapkan metode klasifikasi dan teknik reduksi dimensi *Singular Value Decomposition* (SVD) yang membantu proses klasifikasi dokumen berdimensi tinggi dengan metode *Sequential Minimal Optimization* (SMO). Algoritma SVD pada penelitian ini menghasilkan 120% pengurangan data dan mengurangi kompleksitas waktu. Untuk evaluasi kinerja, penelitian ini menggunakan 5 metode klasifikasi, yaitu *decision tree*, *K-nearest neighbor*, *naive bayes*, *RBF neural network*, dan SMO. Hasil percobaan menunjukkan semua klasifikasi menghasilkan akurasi lebih dari 90%. Algoritma SVD memiliki kelebihan pada efisiensi waktu proses pada dataset yang berskala besar (Navas & Ordonez, 2015) namun memiliki kelemahan tidak dapat melakukan perhitungan jika dataset memiliki elemen yang hilang (*missing value*) (n. Kumar, 2011) dan juga hasil data ekstrak kurang maksimal disebabkan korelasi antar fitur belum bersih 100%. Algoritma PCA tentu menghasilkan data lebih signifikan dibanding algoritma SVD, karena algoritma tersebut mampu menghilangkan korelasi 100%. Namun kelemahan dari algoritma PCA adalah lebih lambat memproses dataset dibanding dengan algoritma SVD dan memerlukan memori yang besar.

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 menguji perbandingan hasil reduksi yang dihasilkan oleh reduksi PCA dan reduksi SVD agar didapatkan hasil reduksi yang efektif untuk proses klasifikasi data trafik internet sehingga dihasilkan klasifikasi data dimensi tinggi dengan akurasi yang baik.

1.3 Rumusan Masalah

Rumusan masalah pada penelitian ini adalah bagaimana pengaruh teknik reduksi dimensi *Principal Component Analysis* (PCA) dan *Singular Value Decomposition* (SVD) terhadap klasifikasi data trafik internet. Untuk menjawab rumusan masalah tersebut, diuraikan beberapa *research question* sebagai berikut:

1. Bagaimana mekanisme *Principal Component Analysis* (PCA) dan *Singular Value Decomposition* (SVD) untuk mereduksi data?
2. Bagaimana jika metode *Principal Component Analysis* (PCA) dan *Singular Value Decomposition* (SVD) masing-masing dikombinasikan dengan metode klasifikasi data berdimensi tinggi.

1.4 Tujuan Penelitian

Tujuan penelitian ini adalah sebagai berikut:

1. Mengetahui mekanisme *Principal Component Analysis* (PCA) dan *Singular Value Decomposition* (SVD) pada klasifikasi data berdimensi tinggi.

2. Mengetahui pengaruh *Principal Component Analysis* (PCA) dan *Singular Value Decomposition* (SVD) pada waktu eksekusi dan akurasi klasifikasi data berdimensi tinggi.

1.5 Manfaat Penelitian

Manfaat dari penelitian ini adalah sebagai berikut:

1. Memahami *Principal Component Analysis* (PCA) dan *Singular Value Decomposition* (SVD) sebagai metode reduksi dimensi data berdimensi tinggi;
2. Mengetahui perbandingan pengaruh *Principal Component Analysis* (PCA) dan *Singular Value Decomposition* (SVD) pada klasifikasi data berdimensi tinggi;
3. Mampu menerapkan teknik reduksi dimensi *Principal Component Analysis* (PCA) dan *Singular Value Decomposition* (SVD) pada metode klasifikasi data berdimensi tinggi.

1.6 Batasan Masalah

Batasan masalah pada penelitian ini adalah sebagai berikut:

1. Data yang digunakan berupa data trafik internet yang didownload dari situs *Computer Laboratory University of Cambridge* (<http://www.cl.cam.ac.uk/>) dalam bentuk .xls.
2. Metode klasifikasi yang digunakan adalah C4.5.
3. Evaluasi kualitas pengklasteran dilakukan dengan *Confusion Matrix*.

4. Validasi data trafik internet 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 reduksi dimensi 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 analisa 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 dalam pengaruh reduksi dimensi pada teknik pengklasteran data berdimensi tinggi.

1.8 Kesimpulan

Penelitian mengenai perbandingan pengaruh reduksi dimensi klasifikasi data trafik internet akan dilakukan dengan *Principal Component Analysis* (PCA) sebagai ekstraksi fitur dan *Singular Value Decomposition* (SVD) sebagai seleksi fitur.

DAFTAR PUSTAKA

- Abbas, M. I. and A. I. S. Azis (2014). "Integrasi Algoritma Singular Value Decomposition (SVD) Dan Principal Component Analysis (PCA) Untuk Pengurangan Dimensi Pada Data Rekam Medis." *ILKOM UMI* 6: 96-111.
- 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.
- Adi Suryaputra P.1, Febriliyan Samopa2, B. C. H. (2014). Klasterisasi Dan Analisis Trafik Internet Menggunakan Fuzzy C Mean Dengan Ekstraksi Fitur Data. *Jurnal Informatika*, 12(1), 33–39.
- 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*.
- Burges, C. J. (2009). *Dimension Reduction: A Guided Tour* (Vol. 2).
- Chen, T. C., Sanga, S., Chou, T. Y., Cristini, V., & Edgerton, M. E. (2009). Neural Network with K-Means Clustering via PCA for Gene Expression Profile Analysis. *World Congress on Computer Science and Information Engineering*, 670–673.
- 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)."
- Kohavi, R., and Provost, F. (1998). "On Applied Research in Machine Learning". Columbia University.
- Kohavi, R., and Sahami, M. (1996). "Error-based and Entropy-based Discretization of Continuous Features ". Second International Conference on Knowledge Discovery and Data Mining: 114-119.

- Kumar, N. (2011). A New Singular Value Decomposition Based Robust Graphical Clustering Technique and Its Application in Climatic Data. *Journal of Geography and Geology*, 3(1), 227–238.
- Kumar, V. (2014). Feature Selection: A literature Review. *The Smart Computing Review*, 4(3).
- Kursrini, dan Luthfi, T. E. (2019). *Algoritma Data Mining*. Yogyakarta: Andi Publisher.
- Leskovec, J., A. Rajaraman and J. Ullman (2014). "Mining of Massive Datasets (2th Edition)."
- Navas, M., Ordonez, C., Navas, M., & Ordonez, C. (2015). Efficient computation of PCA with SVD in SQL Efficient computation of PCA with SVD in SQL, (January 2009).
- Paukkeri, M., & Kivim, I. (n.d.). Effect of Dimensionality Reduction on Different Distance Measures in Document Clustering. *International Conference on Neural Information Processing (ICONIP)*, 167–176.
- Pechenizkiy, M., Puuronen, S., & Tsymbal, A. (2003). Feature Extraction for Classification in Knowledge Discovery Systems Eigenvector-Based Feature Extraction, 526–532.
- Ruangpaisarn, Y. (2015). sEMG Signal Classification Using SMO Algorithm and Singular Value Decomposition. *International Conference on Information Technology and Electrical Engineering (ICITEE)*, 46–50.
- Stanton, J. (2013). *Introduction to Data Science*, 196.
- SUDARYONO, SAEFULLAH, A. & RAHARDJA, U. 2012. *Statistik Deskriptif for IT*.
- 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.
- Wall, M., Rechtsteiner, A., & Rocha, L. (2003). Singular value decomposition and principal component analysis. *A Practical Approach to Microarray Data Analysis*, 91–109.
- Wibawa, M. S., & Novianti, K. D. P. (2017). Reduksi Fitur Untuk Optimalisasi Klasifikasi Tumor Payudara Berdasarkan Data Citra FNA. *Konferensi Nasional Sistem & Informatika 2017*, 73–78.