

PENGARUH REDUKSI FITUR MENGGUNAKAN SVD PADA PENGKLASIFIKASIAN *K-NEAREST NEIGHBOR*

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



Oleh :

RIZKY NOVIANTO
NIM : 09021281520006

**JURUSAN TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER UNIVERSITAS SRIWIJAYA
2021**

LEMBAR PENGESAHAN TUGAS AKHIR

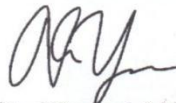
**PENGARUH REDUKSI FITUR MENGGUNAKAN SVD PADA
PENGKLASIFIKASIAN *K-NEAREST NEIGHBOR***

Oleh:

RIZKY NOVIANTO
NIM : 09021181520006

Indralaya, Januari 2021

Pembimbing I,



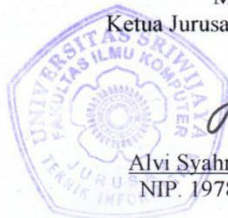
Novi Yushiani M.T
NIP. 198211082012122001

Pembimbing II,



Danny Matthew Saputra, M.Sc
NIP. 198505102015041002

Mengetahui,
Ketua Jurusan Teknik Informatika,



Alvi Syahrini Utami, M.Kom.
NIP. 197812222006042003

TANDA LULUS UJIAN SIDANG TUGAS AKHIR

Pada hari Jumat tanggal 31 Desember 2020 telah dilaksanakan ujian sidang tugas akhir oleh Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

Nama : Rizky Noviano
NIM : 09021181520006
Judul : Pengaruh Reduksi Fitur Menggunakan SVD pada Pengklasifikasian K-Nearest Neighbor

1. Pembimbing I

Novi Yusliani, M.T
NIP. 198211082012122001



2. Pembimbing II

Danny Matthew Saputra, M.Sc
NIP. 198505102015041002



3. Penguji I

Alvi Syahrini Utami, M.Kom
NIP. 197812222006042003

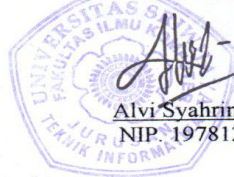


4. Penguji II

Desty Rodiah M.T
NIP. 1671016112890005



Mengetahui,
Ketua Jurusan Teknik Informatika



Alvi Syahrini Utami, M.Kom
NIP. 197812222006042003

HALAMAN PERNYATAAN

Yang bertanda tangan di bawah ini:

Nama : Rizky Novianto
NIM : 09021181520006
Program Studi : Teknik Informatika
Judul Skripsi : Pengaruh Reduksi Fitur Menggunakan SVD
Terhadap : Pada Pengklasifikasian *K-Nearest Neighbor*

Hasil Pengecekan Software *iThenticate/Turnitin* : 10%


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

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



Palembang, Januari 2021




Rizky Novianto
NIM. 09021181520006

MOTTO DAN PERSEMBAHAN

MOTTO: **Jangan Takut Untuk Berbuat Baik**

-Nasihat Ibu-

Kupersembahkan karya tulis ini kepada :

- *Kedua orang tuaku*
- *Keluarga besarku*
- *Sahabat dan teman seperjuanganku*
- *Fakultas Ilmu Komputer*
- *Universitas Sriwijaya*

EFFECT OF FITUR REDUCTION USING SVD ON THE K-NEAREST NEIGHBOR CLASSIFICATION

By:

Rizky Novianto
09021181520006

ABSTRACT

High-dimensional data is a data that has many attributes, one of them is internet traffic data. This research used heart disease data with 76 attributes. If the heart disease 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 for KNN 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 SVD and KNN is proven to be faster than the KNN itself.

Keywords: high dimensional data, heart disease, dimension reduction, singular value decomposition, KNN classification.

PENGARUH REDUKSI FITUR MENGGUNAKAN SVD PADA PENGKLASIFIKASIAN K-NEAREST NEIGHBOR

Oleh:

Rizky Novianto
09021181520006

ABSTRAK

Data berdimensi tinggi adalah data yang memiliki banyak atribut. Pada penelitian ini, data penyakit jantung yang digunakan memiliki 76 atribut, maka dari itu data tersebut tergolong data berdimensi tinggi. Apabila ingin dilakukan pengelompokan terhadap data penyakit jantung 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 KNN. Hasil yang didapatkan melalui pengujian ANOVA menunjukkan tidak ada perbedaan signifikan pada nilai akurasi, precision, dan recall. Namun, jika dilihat dari waktu komputasi, klasifikasi KNN yang didahului oleh SVD terbukti lebih cepat 10 detik.

Keywords: Data Berdimensi Tinggi, Data Penyakit Jantung, Reduksi Dimensi, *Singular Value Decomposition*, Klasifikasi KNN

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 tercinta, Ayah Rudianto dan Ibu Lisinta Adikku tersayang Dinda Aprilia dan seluruh keluarga besarku yang selalu mendoakan serta memberikan dukungan baik moril maupun materil.
2. Bapak Jaidan Jauhari, M.T selaku Dekan Fakultas Ilmu Komputer Universitas Sriwijaya, Ibu Alvi Syahrini Utami, M.Kom selaku Ketua Jurusan Teknik Informatika, dan Ibu Hardini Novianti, M.T selaku Sekretaris Jurusan Teknik Informatika.
3. Ibu Novi Yusliani, M.T. selaku dosen pembimbing I dan Danny Matthew Saputra, MSc selaku pembimbing II yang telah membimbing, mengarahkan, dan memberikan motivasi serta kasih sayang kepada penulis selama proses perkuliahan dan pengerjaan Tugas Akhir.
4. Bapak Drs.Megah Mulya, M.T, selaku dosen pembimbing akademik, yang telah membimbing dan selalu memberikan motivasi kepada penulis dalam proses perkuliahan dan pengerjaan Tugas Akhir.
5. Ibu Alvi Syahrini Utami, M.Kom., Ph.D. selaku dosen penguji I, Seminar Proposal, dan Ibu Desty Rodiah, M.T selaku penguji II Komprehensif yang telah memberikan masukan dan dorongan dalam proses pengerjaan Tugas Akhir.
6. Seluruh dosen Program Studi Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

7. Seluruh staf tata usaha yang telah membantu dalam kelancaran proses administrasi dan akademik selama masa perkuliahan.
8. Seluruh teman jurusan Teknik Informatika yang telah berbagi keluh kesah, motivasi, semangat, dan canda tawa selama masa perkuliahan.

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.

Indralaya, Januari 2021

Rizky Novianto

DAFTAR ISI

	Halaman
HALAMAN JUDUL	i
HALAMAN LEMBAR 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	xvii
DAFTAR LAMPIRAN	xx

BAB I PENDAHULUAN

1.1.... Pendahuluan	I-1
1.2.... Latar Belakang	I-1
1.3.... Rumusan Masalah	I-3
1.4.... Tujuan Penelitian	I-4
1.5.... Manfaat Penelitian	I-4
1.6.... Batasan Masalah	I-4
1.7.... Sistematika Penulisan	I-5
1.8.... Kesimpulan	I-6

BAB II KAJIAN LITERATUR

2.1.... Pendahuluan	II-1
2.2.... Landasan Teori	II-1
2.2.1... Reduksi Data.....	II-1
2.2.2... Klasifikasi.....	II-2
2.2.3... <i>K-Nearest Neighbor</i>	II-4
2.2.4... <i>Singular Value Decomposition</i>	II-6
2.2.5... <i>Confussion Matrix</i>	II-7

2.3.... Praproses.....	II-10
2.4.... Klasifikasi <i>K-Nearest Neighbor</i>	II-10
2.5.... Reduksi Dimensi <i>SVD</i>	II-10
2.6.... Evaluasi D.....	II-15
2.7.... <i>Rational Unified Process</i> (RUP)	II-15
2.8.... Penelitian Lain yang Relevan	II-13
2.8.1...Ahmad Nurrudin Safriando.....	II-13
2.8.2...Shameek Biswas.....	II-13
2.9.... Kesimpulan	II-14

BAB III METODOLOGI PENELITIAN

3.1.... Pendahuluan	III-1
3.2.... Metode Pengumpulan Data	III-1
3.2.1... Jenis dan Sumber Data.....	III-1
3.2.2... Metode Pengumpulan Data.....	III-1
3.3.... Tahapan Penelitian.....	III-2
3.3.1... Menetapkan Kerangka Kerja	III-2
a)... Praproses.....	III-2
b)... Klasifikasi KNN.....	III-4
c)... Reduksi Dimensi.....	III-3
d)... Evaluasi dan Validasi.....	III-3
3.3.2... Menetapkan Kriteria Pengujian.....	III-4
3.3.3... Menetapkan Format Data Pengujian.....	III-4
3.3.4... Menentukan Alat yang Digunakan dalam Pelaksanaan Penelitian..	III-5
3.3.5... Melakukan Pengajuan Penelitian.....	III-6
3.3.6... Melakukan Analisa Hasil Pengujian dan Membuat Kesimpulan....	III-7
3.4.... Metode Pengembangan Perangkat Lunak.....	III-7
a) Fase Insepsi	III-7
b) Fase Elaborasi.....	III-8
c) Fase Konstruksi	III-8
d) Fase Transisi.....	III-9
3.5.... Manajemen Proyek Perangkat Lunak.....	III-10

3.6.... Kesimpulan.....	III-10
-------------------------	--------

BAB IV PENGEMBANGAN PERANGKAT LUNAK

4.1.... Pendahuluan	IV-1
4.2.... <i>Rational Unified Process</i> (RUP)	IV-1
4.2.1... Fase Insepsi	IV-1
4.2.1.1... Pemodelan Bisnis	IV-1
4.2.1.2... Kebutuhan Sistem	IV-2
4.2.1.3... Analisis dan Desain	IV-5
4.2.1.3.1.. Analisis Kebutuhan Perangkat Lunak	IV-5
4.2.1.3.2.. Analisis Data	IV-6
4.2.1.3.3.. Analisis Praproses Data	IV-6
4.2.1.3.4.. Analisis <i>Singular Value Decomposition</i>	IV-7
4.2.1.3.5.. Analisis Klasisifikasi <i>K-Nearest Neighbor</i>	IV-8
4.2.1.3.6.. Desain Perangkat Lunak	IV-9
4.2.2... Fase Elaborasi	IV-15
4.2.2.1... Pemodelan Bisnis	IV-16
4.2.2.1.1.. Perancangan Data	IV-16
4.2.2.1.2.. Perancangan Antarmuka	IV-16
4.2.2.2... Kebutuhan Sistem	IV-17
4.2.2.3... Diagram	IV-18
4.2.2.3.1.. Diagram Aktivitas	IV-18
4.2.2.3.2.. Diagram <i>Sequence</i>	IV-21
4.2.3... Fase Konstruksi	IV-25
4.2.3.1... Kebutuhan Sistem	IV-25
4.2.3.2... Diagram Kelas	IV-25
4.2.3.3... Implementasi	IV-26
4.2.3.3.1.. Implementasi Kelas	IV-27
4.2.3.3.2.. Implementasi Antar Muka	IV-28
4.2.4... Fase Transisi	IV-29
4.2.4.1... Pemodelan Bisnis	IV-29
4.2.4.2... Kebutuhan Sistem	IV-29

4.2.4.3...Rencana Pengujian	IV-30
4.2.4.3.1 Rencana Pengujian <i>Use Case</i> Melakukan <i>Preprocess</i>	IV-30
4.2.4.3.2 Rencana Pengujian <i>Use Case</i> Melakukan Klasifikasi <i>KNN</i>	IV-30
4.2.4.3.3 Rencana Pengujian <i>Use Case</i> Melakukan Reduksi Dimensi <i>SVD</i>	IV-31
4.2.4.4...Implementasi	IV-32
4.2.4.4.1.. Pengujian <i>Use Case</i> Melakukan <i>Preprocess</i>	IV-34
4.2.4.4.2.. Pengujian <i>Use Case</i> Melakukan Klasifikasi <i>KNN</i>	IV-35
4.2.4.4.3.. Pengujian <i>Use Case</i> Melakukan Reduksi Dimensi <i>SVD</i>	IV-36
4.3.... Kesimpulan	IV-38

BAB V HASIL DAN ANALISIS PENELITIAN

5.1.... Pendahuluan	V-1
5.2.... Data Hasil Penelitian.....	V-1
5.3.... Hasil Klasifikasi <i>KNN</i>	V-2
5.4.... Analisis Hasil Penelitian	V-5
5.5.... Analisis Hasil Akurasi	V-5
5.6.... Analisis Hasil <i>Precision</i>	V-7
5.7.... Analisis Hasil <i>Recall</i>	V-8
5.8.... Analisis Hasil Waktu Komputasi.....	V-10
5.9.... Perbandingan Hasil Penelitian	V-12
5.10.. 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	xix
----------------------	-----

LAMPIRAN.....	xxii
---------------	------

DAFTAR TABEL

	Halaman
II-1 Contoh Tabel Confusion Matrix	II-8
III-1.....Tabel Rancangan <i>Confussion Matrix</i>	III-4
III-2.... Tabel Rancangan Pengujian Hasil Klasifikasi.....	III-5
III-3.... Tabel Rancangan Perbandingan Hasil Klasifikasi.....	III-5
IV-1.... Tabel Kebutuhan Fungsional	IV-4
IV-2.... Tabel Kebutuhan Non-Fungsional	IV-5
IV-3.... Tabel Contoh Data Uji Coba	IV-6
IV-4.... Tabel Contoh Data Hasil Praproses.....	IV-7
IV-5.... Tabel Matriks U(5x2).....	IV-8
IV-6.... Tabel Matriks S(2x2).....	IV-8
IV-7.... Tabel Matriks S(2x4).....	IV-8
IV-8.... Tabel Hasil Reduksi Dimensi SVDl.....	IV-8
IV-9.... Tabel Contoh Hasil Klasifikasi.....	IV-9
IV-10.. Tabel Definisi Aktor.....	IV-10
IV-11.. Tabel Definisi <i>Use Case</i>	IV-11
IV-12.. Tabel Skenario <i>Use Case Preprocess</i>	IV-12
IV-13.. Tabel Skenario <i>Use Case</i> Klasifikasi KNN.....	IV-13
IV-14.. Tabel Skenario <i>Use Case</i> Reduksi Dimensi SVD	IV-14
IV-15.. Tabel Implementasi Kelas	IV-27
IV-16.. Tabel Rencana Pengujian <i>Use Case</i> Melakukan <i>Preprocess</i>	IV-30
IV-17.. Tabel Rencana Pengujian <i>Use Case</i> Melakukan Klasifikasi <i>KNN</i>	IV-31
IV-18.. Tabel Rencana Pengujian <i>Use Case</i> Melakukan Reduski Dimensi <i>SVD</i>	IV-31
IV-19.. Tabel Rencana Pengujian <i>Use Case</i> Melakukan <i>Preprocess</i>	IV-34
IV-20.. Tabel Rencana Pengujian <i>Use Case</i> Melakukan Klasifikasi <i>KNN</i>	IV-35
IV-21.. Tabel Rencana Pengujian <i>Use Case</i> Melakukan Reduksi Dimensi <i>SVD</i>	IV-36
V-1..... Tabel Hasil Klasifikasi.....	V-3
V-2..... Tabel Hasil Deskripsi Nilai Akurasi	V-6

V-3.....	Tabel Hasil Ujian Nilai Akurasi menggunakan ANOVA.....	V-7
V-4.....	Tabel Hasil Deskripsi Nilai <i>Precision</i>	V-7
V-5.....	Tabel Hasil Ujian Nilai <i>Precision</i> menggunakan ANOVA.....	V-8
V-6.....	Tabel Hasil Deskripsi Nilai <i>Recall</i>	V-9
V-7.....	Tabel Hasil Ujian Nilai <i>Recall</i> menggunakan ANOVA.....	V-9
V-8.....	Tabel Hasil Deskripsi Waktu Komputasi.....	V-10
V-9.....	Tabel Hasil Ujian Waktu Komputasi menggunakan ANOVA	V-11
V-10...	Tabel Hasil Uji <i>Post Hoc Tukey</i> Waktu Komputasi.....	V-11
V-11...	Tabel Rangkuman Uji Waktu Komputasi Menggunakan <i>Hoc Tukey</i>	V-12

DAFTAR GAMBAR

	Halaman
II-1..... Gambar Arsitektur Model Iteratif RUP (Kruchten, 2003).....	II-12
III-1.... Gambar Tahapan Pengujian Penelitian.....	III-6
IV-1.... Gambar Diagram <i>Use Case</i>	IV-1
IV-2.... Gambar Rancangan Antarmuka.....	IV-17
IV-3.... Gambar Diagram Aktivitas <i>Preprocces</i>	IV-19
IV-4.... Gambar Diagram Aktivitas Klasifikasi KNN.....	IV-20
IV-5.... Gambar Diagram Aktivitas Reduksi Dimensi SVD.....	IV-21
IV-6.... Gambar Diagram <i>Sequence Preprocess</i>	IV-22
IV-7.... Gambar Diagram <i>Sequence</i> Klasifikasi KNN	IV-23
IV-8.... Gambar Diagram <i>Sequence</i> Reduksi Dimensi SVD.....	IV-24
IV-9.... Gambar Diagram Kelas	IV-26
IV-10.. Gambar Antarmuka Perangkat Lunak	IV-28
V-1..... Gambar Perbandingan Hasil Penelitian.....	V-13

DAFTAR LAMPIRAN

	Halaman
L-1.....Nama Attribut.....	L-1
L-2.....Perhitungan Manual SVD.....	L-14
L-3.....Perhitungan Manual KNN.....	L-20
L-4..... <i>Source Code Program</i>	L-39

BAB I

PENDAHULUAN

1.1 Pendahuluan

Pada bab ini ini dilakukan penjelasan mengenai pokok pikiran yang melandasi skripsi ini. Pokok pikiran yang dimaksud ialah latar belakang masalah penelitian, perumusan masalah/permasalahan penelitian, tujuan dan manfaat penelitian, serta batasan masalah..

1.2 Latar Belakang Masalah

Data Mining merupakan suatu proses untuk mendapatkan *knowledge* atau pola dari berbagai kumpulan data (Ian&Witten,2011). *Data mining* melakukan pemecahan pada suatu pola masalah dengan cara menganalisa data yang sudah ada pada database. Data mining sering disebut dengan istilah *knowledge discovery in database(KDD)* yaitu kegiatan yang meliputi pengumpulan data, pemakaian riwayat data untuk menemukan pola data yang teratur pada data yang berukuran besar (Santoso,2007).

Data mining dibagi menjadi beberapa kelompok tugas yang dapat dilakukan, salah satunya ialah klasifikasi. Metode klasifikasi sering digunakan sebagai penelitian pada bidang kesehatan (*medical science*) seperti melakukan diagnosa suatu penyakit yang diidap oleh pasien. Salah satu metode klasifikasi yang paling umum digunakan ialah metode *k-nearest neighbor*. Metode ini

bertujuan untuk mengklasifikasikan obyek data berdasarkan *attribute* dan *training sample* . Prinsip kerja dari *k-nearest neighbor* yaitu mencari jarak terdekat diantara data yang akan dievaluasi dengan k tetangga(*neighbor*) yang terdekat.

Berdasarkan penelitian yang telah dilakukan oleh Abdul Rohman (2016) mengenai prediksi penyakit jantung menggunakan metode *k-nearest neighbor* yang diaplikasikan ke seluruh data pasien yang ada , baik yang sehat maupun sakit mendapatkan nilai akurasi sebesar 77,58%.

Alasan metode *k-nearest neighbor* ini digunakan ialah karena metode ini sangat mudah dipahami dan diimplementasikan.. Namun, algoritma *k-nearest neighbor* memiliki suatu kelemahan yaitu rentan terhadap data yang memiliki dimensionalitas yang tinggi(S.Mutrofin,2015). Karena semakin banyaknya dimensi, maka ruang yang ditempati oleh suatu instance semakin besar, sehingga semakin besar pula probabilitas dari *k nearest neighbour* pada suatu instance tidak “near”. Selain itu kelemahan dari *k-nearest neighbor tidak bisa* menangani *missing value* secara jelas yang berakibat waktu komputasiyang dibutuhkan antara *data training* dan *data sample* memerlukan waktu yang cukup tinggi. Untuk itulah kekurangan dari klasifikasi *k-nearest neighbor* ini diatasi dengan cara reduksi dimensi.

Keuntungan dari reduksi dimensi ialah algoritma *k-nearest neighbor* dapat bekerja secara optimal jika memiliki dimensi yang lebih rendah. Karena reduksi dimensi dapat mengatasi *irrelevant feature*,*missing value*,mengurangi noise, dan

bisa mengurangi kelemahan dari *k-nearest neighbor ini* yaitu dimensionalitas yang tinggi atau sering dikenal *curse of dimensionality* (Shanwen Zhang,2011). Reduksi dimensi mampu membuat model agar lebih mudah dipahami karena menggunakan fitur yang relatif lebih sedikit dan hal ini juga akan berdampak pada waktu dan memori yang diperlukan dalam proses komputasi..

Metode reduksi dimensi data yang paling umum digunakan ialah *Principal Component Analysis (PCA)* dan *Singular Value Decomposition(SVD)*. Metode *Principal Component Ananlysis* dinilai kurang efektif karenaa kurang optimal pada pemisahan data antar kelas. Maka dari itu metode *Singular Value Decomposition* saya pilih karena dinilai cocok untuk mengatasi kelemahan yang ada pada klasifikasi *k-nearest neighbor*.

1.3 Rumusan Masalah

Rumusan masalah dalam penelitian ini ialah bagaimana pengaruh *Singular Value Decompsition (SVD)* dalam pengklasifikasian menggunakan *k-nearest neighbor (k-NN)* pada kasus penyakit jantung ?. untuk menjawab hal tersebut, maka diuraikan pertanyaan sebagai berikut :

1. Bagaimana implementasi metode k-NN dalam mengklasifikasikan dta berdimensi tinggi pada penyakit jantung ?
2. Bagaimana impleementasi metode SVD dalam mereduksi data berdimensi tinggi pada penyakit jnatung ?

3. Bagaimana tingkat akurasi pengklasifikasian data sebelum dan sesudah direduksi menggunakan metode SVD ?

1.4 Tujuan Penelitian

Berdasarkan rumusan masalah diatas maka tujuan yang ingin dicapai dalam penelitian adalah sebagai berikut :

1. Mengimplementasikan metode k - NN dalam mengklasifikasikan data berdimensi tinggi.
2. Mengimplementasikan metode SVD dalam mereduksi data berdimensi tinggi.
3. Mengetahui tingkat akurasi hasil dari pengklasifikasian.

1.5 Manfaat Penelitian

Manfaat yang diperoleh dari penelitian ialah sebagai berikut:

1. Memahami implementasi k - NN dalam pengklasifikasian data berdimensi tinggi.
2. Memahami SVD sebagai metode reduksi data berdimensi tinggi.

1.6 Batasan Masalah

Batasan masalah pada penelitian ini adalah sebagai berikut:

1. *Dataset* yang digunakan merupakan dataset yang diperoleh dari public repository UCI yaitu dataset penyakit jantung.
2. Metode klasifikasi yang digunakan ialah k - NN .
3. Evaluasi klasifikasi dilakukan dengan cara *Confusion Matrix*.

4. Validasi data penyakit jantung menggunakan *10-fold cross validation*.

1.7 Sistematika Penulisan

Sistematika penulisan ini adalah sebagai berikut:

BAB I. PENDAHULUAN

Bab I membahas latar belakang masalah, rumusan masalah, tujuan penelitian dan manfaat penelitian, batasan masalah, sistematika penulisan penelitian dan kesimpulan.

BAB II. KAJIAN LITERATUR

Bab II membahas dasar-dasar teori yang akan digunakan dalam penelitian, seperti pengetahuan dasar mengenai reduksi dimensi dan metode yang digunakan dalam proses klasifikasi *k-NN*.

BAB III. METODOLOGI PENELITIAN

Bab III membahas unit penelitian, tahapan penelitian, metode pengembangan perangkat lunak, teknik pengujian dan manajemen proyek penelitian.

BAB IV. PENGEMBANGAN PERANGKAT LUNAK

Bab IV membahas analisa dan perancangan perangkat lunak yang akan digunakan pada penelitian ini.

BAB V. HASIL DAN ANALISIS PENELITIAN

Bab V menguraikan hasil pengujian perangkat lunak berdasarkan langkah-langkah yang telah disusun. Analisis disajikan sebagai basis dari kesimpulan yang diambil dari penelitian ini.

BAB VI. KESIMPULAN DAN SARAN

Bab VI berisi kesimpulan dari semua uraian uraian pada semua bab dan juga berisi saran yang diharapkan dapat membantu pada pengembangan penelitian selanjutnya.

1.8 Kesimpulan

Berdasarkan penjelasan yang telah dijelaskan diatas penulis melakukan penelitian ini agar penulis mengetahui tingkat akurasi dari suatu algoritma *k-nearest neighbor(k-NN)* dan algoritma *k-nearest neighbor(k-NN)* yang digabungkan dengan metode *Singular Value Decomposition(SVD)* pada kasus prediksi penyakit jantung. Klasifier yang digabungkan dengan metode *Singular Value Decomposition ini* bertujuan untuk mengatasi kelemahan pada algoritma *k-nearest neighbor(k-NN)*.

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.
- Anand, S., Padmanabham, P., & Govardhan, A. (2015). Effect of *Distance Measures* on Partitional *Clustering* Algorithms using Transportation Data. *International Journal of Computer Science and Information Technologies*, 6(6), 5308–5312.
- Arechiga, A., Barocio, E., Ayon, J. J., & Garcia-Baleon, H. A. (2017). Comparison of Dimensionality Reduction Techniques for *Clustering* and Visualization of Load Profiles. *2016 IEEE PES Transmission and Distribution Conference and Exposition-Latin America, PES T and D-LA 2016*, 2, 1–6. <https://doi.org/10.1109/TDC-LA.2016.7805661>
- Awodele, O., & Jegede, O. (2009). Neural Networks and its Application in Engineering. *Proceedings of Informing Science & IT Education Conference (InSITE) 2009*, 1–13. <http://proceedings.informingscience.org/InSITE2009/InSITE09p083-095Awodele542.pdf>
- Bora, M. D. J., & Gupta, D. A. K. (2014). *Effect of Different Distance Measures on the Performance of K-Means Algorithm: An Experimental Study in Matlab*. 5(2), 2501–2506. <http://arxiv.org/abs/1405.7471>
- Fausett, L. (1994). *Fundamentals of Neural Networks: Architectures, Algorithms, and Applications*. Prentice-Hall, Inc.
- Galluccio, L., Michel, O., Comon, P., & Hero, A. O. (2012). Graph based K-Means *Clustering*. *Signal Processing*, 92(9), 1970–1984. <https://doi.org/10.1016/j.sigpro.2011.12.009>
- Halkidi, M., Batistakis, Y., & Vazirgiannis, M. (2001). On *Clustering* Validation Techniques. *Journal of Intelligent Information Systems*, 17(3), 107–145. <https://doi.org/10.1023/A>
- Han, J., Kamber, M., & Pei, J. (2012). Data Mining. In *Data Mining* (Third). Morgan Kaufmann. <https://doi.org/10.1016/b978-0-12-381479-1.00001-0>
- Hasanah, S. I. R., Jambak, M. I., & Saputra, D. M. (2018). *Comparison of Dimensional Reduction Using Singular Value Decomposition and Principal*

Component Analysis for Clustering Results of Indonesian Language Text Documents. Sriwijaya University.

Jun, S., Park, S. S., & Jang, D. S. (2014). Document *Clustering* Method using Dimension Reduction and Support Vector *Clustering* to Overcome Sparseness. *Expert Systems with Applications*, 41(7), 3204–3212. <https://doi.org/10.1016/j.eswa.2013.11.018>

Kirdat, T., & Patil, V. V. (2016). *INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH IN TECHNOLOGY* 28 *Application of Chebyshev Distance and Minkowski Distance to CBIR Using Color Histogram.* 2(9), 28–31.

Kohonen, T. (2001). *Self-Organizing Maps* (3rd ed.). Springer International Publishing.

Kouser, K., & Sunita, S. (2013). A Comparative Study of K Means Algorithm by Different *Distance* Measures. *International Journal of Innovative Research in Computer and Communication Engineering*, 1(9), 2443–2447. www.ijirce.com

Kruchten, P. (2003). The Rational Unified Process--An Introduction. *Rational Software*, May.

Kumar, V., Chhabra, J. K., & Kumar, D. (2014a). Impact of *Distance* Measures on the Performance of *Clustering* Algorithms. *Advances in Intelligent Systems and Computing*, 243, 183–190. https://doi.org/10.1007/978-81-322-1665-0_17

Liu, P., El-Darzi, E., Lei, L., Vasilakis, C., Chountas, P., & Huang, W. (2005). An Analysis of Missing Data Treatment Methods and Their Application to Health Care *Dataset.* *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 3584 *LNAI*, 583–590.

Miljkovic, D. (2017). Brief Review of Self-Organizing Maps. *2017 40th International Convention on Information and Communication Technology, Electronics and Microelectronics, MIPRO 2017 - Proceedings, October*, 1061–1066. <https://doi.org/10.23919/MIPRO.2017.7973581>

Mythili, S., & Madhiya, E. (2014). An Analysis on *Clustering* Algorithms in Data Mining. *International Journal of Computer Science and Mobile Computing*, 3(1), 334–340.

Nawrin, S., Rahatur, M., & Akhter, S. (2017). Exploreing K-Means with Internal Validity Indexes for Data *Clustering* in Traffic Management System. *International Journal of Advanced Computer Science and Applications*, 8(3). <https://doi.org/10.14569/ijacsa.2017.080337>

Pachghare, V. K., Patole, V. A., & Kulkarni, D. P. (2010). Self Organizing Maps to Build Intrusion Detection System. *International Journal of Computer Applications*, 1(8), 1–4. <https://doi.org/10.5120/191-328>

- Pande, S. R., Sambare, M. S. S., & Thakre, V. M. (2012). *Data Clustering Using Data Mining Techniques*. 1(8), 494–499.
- Patro, S. G. K., & sahu, K. K. (2015). Normalization: A Preprocessing Stage. *Iarjset*, 20–22. <https://doi.org/10.17148/iarjset.2015.2305>
- Paukkeri, M. S., Kivimäki, I., Tirunagari, S., Oja, E., & Honkela, T. (2011). Effect of dimensionality reduction on different *distance* measures in document *clustering*. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 7064 LNCS(PART 3), 167–176. https://doi.org/10.1007/978-3-642-24965-5_19
- Putra, J. W. G. (2020). *Pengenalan Konsep Pembelajaran Mesin dan Deep Learning* (1.4, Vol. 4, p. 227).
- Rani, S., & Sonika, S. (2014). Effectiveness of Data Preprocessing for Data Mining. *International Journal of Current Engineering and Technology*, 4(5), 3480–3483. <http://www.theseus.fi/handle/10024/62626>
- Sari, T., Rini, D. P., & Yusliani, N. (2018). *PENGARUH PERHITUNGAN RENTANG DATA TERHADAP PERFORMANSI ALGORITMA CLUSTERING SELF ORGANIZING MAP*. Sriwijaya University.
- Singh, A., Yadav, A., & Rana, A. (2013). K-Means with Three Different *Distance* Metrics. *International Journal of Computer Applications*, 67(10), 13–17. <https://doi.org/10.5120/11430-6785>
- Tamilselvi, R., Sivasakthi, B., & Kavitha, R. (2015). *An Efficient Preprocessing and Postprocessing Techniques in Data Mining*.
- Vimal, A., Valluri, S. R., & Karlapalem, K. (2008). *An Experiment with Distance Measures for Clustering . An Experiment with Distance Measures for Clustering. June 2014*.
- Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2016). *Data Mining Practical Machine Learning Tools and Techniques* (Fourth). Elsevier Inc.
- Zaware, P. S. N., Gautam, M. A., Nashte, M. S., & Khanuja, M. P. (2015). *AN EFFECTUAL APPROACH FOR CALCULATING COSINE SIMILARITY*. 13–18.