

**Perbandingan Seleksi Fitur Teknik *Wrapper* Menggunakan  
*Forward Selection* dengan Teknik *Filter* Menggunakan *Singular  
Value Decomposition* terhadap Hasil Pengklasteran Dokumen**

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



Oleh :

Divinda Morikawa  
09021181520121

**JURUSAN TEKNIK INFORMATIKA  
FAKULTAS ILMU KOMPUTER UNIVERSITAS SIRIWJAYA  
2020**

## LEMBAR PENGESAHAN TUGAS AKHIR

Perbandingan Seleksi Fitur Teknik *Wrapper* Menggunakan *Forward Selection* dengan Teknik *Filter* Menggunakan *Singular Value Decomposition* terhadap Hasil Pengklasteran Dokumen

Oleh :

DIVINDA MORIKAWA

NIM : 09021181520121

Mengetahui,

Indralaya, 8 Agustus 2020

Pembimbing I,

Pembimbing II,



Ir. M. Insan Jambak, M.Sc., MM.

Desty Rodiah, M.T.

NIP. 196804052013081201

NIP. 1671016112890005

Mengetahui,  
Ketua Jurusan Teknik Informatika



Rifkie Primartha, M.T  
NIP. 197706012009121004

## TANDA LULUS SIDANG TUGAS AKHIR

Pada hari Kamis, 22 Juli 2020 telah dilaksanakan ujian sidang tugas akhir oleh Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

Nama : Divinda Morikawa

NIM : 09021181520121

Judul : Perbandingan Seleksi Fitur Teknik Wrapper Menggunakan Forward Selection dengan Teknik Filter Menggunakan Singular Value Decomposition terhadap Hasil Pengklasteran Dokumen

1. Pembimbing I

Ir. M. Ihsan Jambak, M.Sc.  
NIP. 196804052013081201



2. Pembimbing II

Desty Rodiah, M.T  
NIP. 1671016112890005



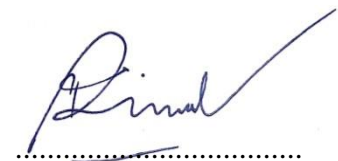
3. Penguji I

Novi Yusliani, M.T.  
NIP.198211082012122001



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 : Divinda Morikawa

NIM : 09021181520121

Program Studi : Teknik Informatika

Judul Skripsi : Perbandingan Seleksi Fitur Teknik *Wrapper* Menggunakan *Forward Selection* dengan Teknik *Filter* Menggunakan *Singular Value Decomposition* terhadap Hasil Pengklasteran Dokumen

Hasil Pengecekan Software *iThenticate/Turnitin* : 10 %

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.



Indralaya, 8 Agustus 2020



Divinda Morikawa

NIM. 09021181520121

## 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 dan seluruh keluarga besarku 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.
4. Bapak Ir. M. Ihsan Jambak, M.Sc. selaku dosen pembimbing I yang telah membimbing, mengarahkan dan memberikan motivasi dalam proses perkuliahan dan pengerjaan Tugas Akhir.
5. Ibu Desty Rodiah.M.T selaku dosen pembimbing II yang telah memberikan arahan serta dukungan dalam proses perkuliahan dan pengerjaan Tugas Akhir.
6. Ibu Novi Yusliani, M.T dan Mastura Diana Marieska, M.T selaku dosen penguji yang telah memberikan masukan dan dorongan dalam proses pengerjaan Tugas Akhir.

7. Seluruh dosen Program Studi Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

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, 8 Agustus 2020

A handwritten signature in black ink, appearing to be 'Divinda Morikawa', enclosed within a circular scribble.

Divinda Morikawa

*Motto:*

- Merantaulah. Karena ia mengajarkan arti ,mengapa kamu harus pulang dan siapa yang kamu rindukan.
- Kita sering melihat namun jarang memperhatikan. Dan menyendiri salah satu jalan untuk paham yang terlewatkan.
- Jikapun satu buku terlalu berat untuk kau baca, bacalah judulnya saja jika hanya itu satu-satunya cara agar kau membaca.
- Konsep membaca adalah, semakin kamu baca semakin kamu tahu bahwa kamu tidak tahu. Dan seharusnya itu yang akan membuat rasa ingin tahumu menggebu.
- Kita terlalu sering berambisi mengontrol perspektif dan emosi orang lain. Namun, kita lupa bahwa tidak ada emosi apapun yang bisa kita kontrol selain emosi diri sendiri. Sebab berusaha membahagiakan semua orang adalah hal yang tidak membahagiakan sama sekali.

*Kupersembahkan karya tulis ini kepada:*

- *Orang tuaku tercinta*
- *Keluarga besarku*
- *Sahabat dan teman seperjuanganku*
- *Fakultas Ilmu Komputer*
- *Universitas Sriwijaya*

## DAFTAR ISI

	Halaman
<u>LEMBAR PENGESAHAN TUGAS AKHIR</u> .....	ii
<u>ABSTRACT</u> .....	xixi
<u>ABSTRAK</u> .....	vii
<u>KATA PENGANTAR</u> .....	viii
<u>DAFTAR ISI</u> .....	x
<u>DAFTAR TABEL</u> .....	xii
<u>DAFTAR GAMBAR</u> .....	xv

### BAB I PENDAHULUAN

1.1 Pendahuluan .....	I-1
1.2 Latar Belakang Masalah .....	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-5
1.7 Kesimpulan .....	I-5

### BAB II KAJIAN LITERATUR

2.1 Pendahuluan .....	II-1
2.2 Landasan Teori .....	II-1
2.2.1 Reduksi Dimensi .....	II-1
2.2.2 <i>Singular Value Decomposition</i> .....	II-2
2.2.3 <i>Forward Selection</i> .....	II-5
2.2.4 K-Means .....	II-8
2.2.5 <i>Davies Boundin Index (DBI)</i> .....	II-9
2.3 Penelitian Lain yang Relevan .....	II-11
2.4 Kesimpulan .....	II-12

### BAB III METEDOLOGI PENELITIAN

3.1 Pendahuluan .....	III-1
3.2 Unit Penelitian .....	III-1
3.3 Pengumpulan Data .....	III-1
3.3.1 Jenis dan Sumber Data .....	III-1
3.3.2 Metode Pengumpulan Data .....	III-2
3.4 Tahap Penelitian .....	III-2
3.4.1 Menetapkan Kerangka Kerja .....	III-3
3.4.2 Menetapkan Kriteria Pengujian .....	III-7



3.4.3 Menetapkan Format Data Pengujian .....	III-7
3.4.4 Menentukan Alat yang Digunakan dalam Pelaksanaan Penelitian .....	III-9
3.4.5 Melakukan Pengujian Penelitian .....	III-10
3.4.6 Melakukan Analisa Hasil Pengujian dan Membuat Kesimpulan Penelitian .....	III-11
3.5 Metode Pengembangan Perangkat Lunak .....	III-12
3.5.1 <i>Rational Unified Process</i> (RUP) .....	III-12
3.5.1.1 Fase Insepsi .....	III-13
3.5.1.2 Fase Elaborasi .....	III-13
3.5.1.3 Fase Konstruksi .....	III-14
3.5.1.4 Fase Transisi .....	III-14
3.6 Manajemen Proyek Penelitian .....	III-15
3.7 Kesimpulan.....	III-27
<b>BAB IV PENGEMBANGAN PERANGKAT LUNAK</b>	
4.1 Pendahuluan.. .....	IV-1
4.2 Fase Insepsi.. .....	IV-1
4.2.1 Pemodelan Bisnis. ....	IV-1
4.2.2 Kebutuhan Sistem .. ..	IV-4
4.2.3 Analisis dan Desain.....	IV-15
4.3 Fase Elaborasi.. .....	IV-17
4.3.1 Pemodelan Bisnis.....	IV-17
4.3.2 Kebutuhan Sistem .. ..	IV-22
4.3.3 Analisis dan Desain.....	IV-26
4.4 Fase Konstruksi.....	IV-32
4.4.1 Pemodelan Bisnis.....	IV-32
4.4.2 Kebutuhan Sistem.....	IV-34
4.4.3 Implementasi.....	IV-34
4.4.3.1 Implementasi Kelas.....	IV-34
4.4.4 Rencana Pengujian.....	IV-36
4.4.4.1 Rencana Pengujian Use Case Melakukan Konversi Format Data Teks Menjadi Numerik.....	IV-36
4.4.4.2 Rencana Pengujian Use Case Mengkluster Data menggunakan K – Means.....	IV-37
4.4.4.3 Rencana Pengujian Use Case Mereduksi Dimensi Data dengan SVD dan Pengklusteran dengan K-Means.....	IV-38
4.4.4.4 Rencana Pengujian Use Case Mereduksi dengan <i>Forward Selection</i> dan Pengklusteran dengan K–Means.....	IV-40
4.5 Fase Transisi.....	IV-41
4.5.1 Pemodelan Bisnis.....	IV-41
4.5.2 Kebutuhan Sistem.....	IV-41
4.5.3 Pengujian.....	IV-42

4.6 Kesimpulan.....	IV-47
<b>BAB V HASIL DAN ANALISIS PENELITIAN</b>	
5.1 Pendahuluan.....	V-1
5.2 Konfigurasi Percobaan.....	V-1
5.3 Hasil Uji Penelitian.....	V-2
5.3.1 Hasil Pengujian Penentuan k Optimum.....	V-2
5.3.2 Hasil Pengujian k-Means.....	V-5
5.3.3 Hasil Pengujian k-Means dengan SVD .....	V-6
5.3.4 Hasil Pengujian k-Means dengan <i>Forward Selection</i> ...	V-7
5.4 Olah Data.....	V-8
5.4.1 Olah Data Hasil Pengujian k Optimum.....	V-8
5.4.2 Olah Data Hasil Perbandingan DBI.....	V-9
5.4.3 Olah Data Hasil Perbandingan Waktu Komputasi Klaster.....	V-15
5.4.4 Olah Data Hasil Perbandingan Jumlah Iterasi Klaster.....	V-19
5.5 Analisa Hasil Pengujian.....	V-24
5.5 Kesimpulan.....	V-26
<b>BAB VI KESIMPULAN DAN SARAN</b>	
6.1 Pendahuluan.....	VI-1
6.2 Kesimpulan.....	VI-1
6.3 Saran.....	VI-2
DAFTAR PUSTAKA .....	xvi
LAMPIRAN.....	xviii

## DAFTAR TABEL

	Halaman
III-1 Rancangan Tabel Hasil Pengujian Penentuan Nilai $k$ Optimum .....	III-7
III-2 Rancangan Tabel Hasil Pengujian Penentuan $k$ -Means.....	III-8
III-3 Rancangan Tabel Hasil Pengklasteran $k$ -Means dengan <i>Forward Selection</i> .....	III-8
III-4 Rancangan Tabel Hasil Pengklasteran $k$ -Means dengan SVD .....	III-9
III-5 Rancangan Tabel Hasil Perbandingan Pengklasteran Tanpa Reduksi Dimensi dan Menggunakan Reduksi Dimensi.....	III-11
III-6 Penjadwalan Penelitian dalam Bentuk <i>Work Breakdown Structure (WBS)</i> .....	III-16
IV-1 Definisi Aktor.....	IV-3
IV-2 Definisi Use Case.. .....	IV-3
IV-3 Kebutuhan Fungsional.....	IV-4
IV-4 Kebutuhan Non-Fungsional.....	IV-5
IV-5 Skenario Muat Data. ....	IV-6
IV-6 Skenario <i>Use Case</i> Mereduksi Dimensi Data Menggunakan SVD dan <i>Clustering k-Means</i> .....	IV-8
IV-7 Skenario <i>Use Case</i> Mereduksi Dimensi data menggunakan <i>Forward Selection</i> dan <i>Clustering k-Means</i> .. .....	IV-11
IV-8 Skenario <i>Use Case</i> Melakukan Pengklasteran menggunakan $k$ -Means..	IV-13
IV-9 Implementasi Kelas.. .....	IV-32
IV-10 Rencana Pengujian <i>Use Case</i> Melakukan Konversi Format Data Text Menjadi Numerik.....	IV-35
IV-11 Rencana Pengujian <i>Use Case</i> Melakukan Clustering dengan K-Means.....	IV-37
IV-12 Rencana Pengujian <i>Use Case</i> Mereduksi dengan SVD dan Clustering dengan K- Means.....	IV-38
IV-13 Rencana Pengujian <i>Use Case</i> Mereduksi dengan <i>Forward Selection</i> dan Pengklasteran dengan $k$ -Means.....	IV-39
IV-14 Pengujian Fitur Konversi Data.....	IV-40
IV-15 Pengujian Fitur Pengklasteran $k$ -Means.....	IV-41
IV-16 Pengujian Fitur Pengklasteran $k$ -Means dengan Reduksi Dimensi ....	IV-42
IV-17 Pengujian Fitur Pengklasteran $k$ -Means dengan Reduksi Dimensi <i>Forward Selection</i> .....	IV-43
V-1 Hasil Pengujian $k$ .....	V-4
V-2 Hasil <i>Clustering</i> dengan $k$ Ideal.....	V-5
V-3 Hasil <i>Clustering</i> dengan SVD.....	V-6
V-4 Hasil Clustering dengan <i>Forward Selection</i> .....	V-7
V-5 Deskripsi Nilai DBI.....	V-9

V-6	Hasil Uji Normalitas Nilai DBI.....	V-10
V-7	Hasil Uji Mean Rank Perbandingan Nilai DBI.....	V-12
V-8	Hasil Uji Kruskal-Wallis Perbandingan Nilai DBI.....	V-12
V-9	Hasil Uji Mean Rank Perbandingan Nilai DBI pada Data Hasil Pengklastran K-Means dan SVD+K-Means .....	V-13
V-10	Hasil Uji Mann Whitney Perbandingan Nilai DBI pada Data Hasil Pengklastran K-Means dan SVD+K-Means.....	V-13
V-11	Hasil Uji Mean Rank Perbandingan Nilai DBI pada Data Hasil Pengklastran K-Means dan FS+K-Means.....	V-13
V-12	Hasil Uji Mann Whitney Perbandingan Nilai DBI pada Data Hasil Pengklastran K-Means dan FS+K-Means.....	V-14
V-13	Hasil Uji Mean Rank Perbandingan Nilai DBI pada Data Hasil Pengklastran FS+K-Means dan SVD+K-Means.....	V-14
V-14	Hasil Uji Mann-Whitney Perbandingan Nilai DBI pada Data Hasil Pengklastran FS+K-Means dan SVD+K-Means.....	V-14
V-15	Deskripsi Waktu Komputasi.....	V-15
V-16	Hasil Uji Normalitas Waktu Komputasi.....	V-16
V-17	Hasil Uji Kruskal-Wallis Perbandingan Waktu Komputasi.....	V-16
V-18	Hasil Uji Kruskal-Wallis Perbandingan Waktu Komputasi.....	V-17
V-19	Hasil Uji Mean Rank Perbandingan Waktu Komputasi pada Data Hasil Pengklastran K-Means dan SVD+K-Means.....	V-17
V-20	Hasil Uji Mann-Whitney Perbandingan Waktu Komputasi pada Data Hasil Pengklastran K-Means dan SVD+K-Means.....	V-17
V-21	Hasil Uji Mean Rank Perbandingan Waktu Komputasi pada Data Hasil Pengklastran K-Means dan FS+K-Means.....	V-18
V-22	Hasil Uji Mann-Whitney Perbandingan Waktu Komputasi pada Data Hasil Pengklastran K-Means dan FS+K-Means.....	V-18
V-23	Hasil Uji Mean Rank Perbandingan Waktu Komputasi pada Data Hasil Pengklastran SVD+K-Means dan FS+K-Means.....	V-19
V-24	Hasil Uji Mann-Whitney Perbandingan Waktu Komputasi pada Data Hasil Pengklastran SVD+K-Means dan FS+K-Means.....	V-19
V-25	Deskripsi Jumlah Iterasi.....	V-19
V-26	Hasil Uji Normalitas Jumlah Iterasi.....	V-20
V-27	Hasil Uji Mean Rank Perbandingan Jumlah Iterasi.....	V-21
V-28	Hasil Uji Kruskal-Wallis Perbandingan Jumlah Iterasi.....	V-21
V-29	Hasil Uji Mean Rank Perbandingan Jumlah Iterasi pada Data Hasil Pengklastran K-Means dan SVD+K-Means.....	V-22
V-30	Hasil Uji Mean Whitney Perbandingan Jumlah Iterasi pada Data Hasil Pengklastran K-Means dan SVD+K-Means.....	V-22
V-31	Hasil Uji Mean Rank Perbandingan Jumlah Iterasi pada Data Hasil	

Pengklastran K-Means dan FS+K-Means.....	V-22
V-32 Hasil Uji Mann-Whitney Perbandingan Jumlah Iterasi pada Data Hasil Pengklastran K-Means dan FS+K-Means.....	V-23
V-33 Hasil Uji Mean Rank Perbandingan Jumlah Iterasi pada Data Hasil Pengklastran SVD+K-Means dan FS+K-Means.....	V-23
V-34 Hasil Uji Mann-Whitney Perbandingan Jumlah Iterasi pada Data Hasil Pengklastran SVD+K-Means dan FS+K-Means.....	V-24

## DAFTAR GAMBAR

	Halaman
II-1 Dekomposisi <i>Singular Value Decomposition</i> .....	II-3
III-1 Diagram Tahapan Penelitian.....	III-2
III-2 Tahap Pengujian Penelitian .....	III-10
III-3 Arsitektur RUP .....	III-12
III-4 Penjadwalan Penelitian Tahap Menentukan Ruang Lingkup dan Unit Penelitian.....	III-22
III-5 Penjadwalan Penelitian Tahap Menentukan Dasar Teori yang Berkaitan dengan Penelitian .....	III-23
III-6 Panjadwalan Penelitian Tahap Penentuan Kriteria Pengujian.....	III-23
III-7 Penjadwalan Penelitian Tahap Menentukan Alat yang Digunakan untuk Pelaksanaan Penelitian pada Fase Insepsi .....	III-24
III-8 Penjadwalan Penelitian Tahap Menentukan Alat yang Digunakan untuk Pelaksanaan Penelitian pada Fase Elaborasi .....	III-24
III-9 Penjadwalan Penelitian Tahap Menentukan Alat yang Digunakan untuk Pelaksanaan Penelitian pada Fase Konstruksi.....	III-25
III-10 Penjadwalan Penelitian Tahap Menentukan Alat yang Digunakan untuk Pelaksanaan Penelitian pada Fase Transisi.....	III-25
III-11 Penjadwalan Penelitian Tahap Melakukan Pengujian Penelitian .....	III-26
III-12 Penjadwalan Penelitian Tahap Melakukan Analisa Hasil Pengujian dan Membuat Kesimpulan .....	III-26
IV-1 Diagram Use Case. ....	IV-12
IV-2 Diagram <i>sequence</i> melakukan konversi format data teks.. .....	IV-18
IV-3 Diagram <i>sequence</i> pengklasteran data dengan menggunakan <i>k-Means</i> dan evaluasi DBI .....	IV-19
IV-4 Diagram <i>sequence</i> mereduksi data menggunakan SVD.....	IV-20
IV-5 Diagram <i>sequence</i> mereduksi data menggunakan <i>Forward Selection</i> ..	IV-21
IV-6 Rancangan Antar Muka Perangkat Lunak.....	IV-22
IV-7 Rancangan Fitur Input Dokumen.. .....	IV-23
IV-8 Rancangan Fitur Input Jumlah Cluster.. .....	IV-23
IV-9 Rancangan Fitur Metode <i>Clustering</i> .. .....	IV-24
IV- 10 Rancangan Fitur Tombol Mulai Proses.. .....	IV-25
IV-11 Rancangan Fitur Menampilkan Waktu Komputasi, nilai DBI dan iterasi .....	IV-25
IV-12 Rancangan Fitur Menampilkan Hasil <i>Clustering</i> Dokumen.. .....	IV-26
IV-13 Diagram Aktivitas <i>Use Case</i> Melakukan Load Data.....	IV-27
IV-14 Diagram Aktivitas <i>Use Case</i> Melakukan <i>Clustering</i> dengan <i>K-Means</i> .....	IV-28
IV-15 Diagram Aktivitas <i>Use Case</i> Mereduksi dengan SVD dan <i>Clustering</i> dengan <i>K-Means</i> .. .....	IV-28
IV-16 Diagram Aktivitas <i>Use Case</i> Mereduksi dengan FS dan <i>Clustering</i> dengan <i>K-Means</i> .. .....	IV-29

IV-17 Diagram Kelas.....	IV-31
V-1 Grafik Hasil Pengujian k.....	V-9

## **DAFTAR LAMPIRAN**

Lampiran I Kode Program

Lampiran II Contoh Perhitungan Singular Value Decomposition (SVD)

Lampiran III Contoh perhitungan k-Means

Lampiran IV Contoh perhitungan forward selection

Lampiran V Contoh perhitungan BDI



COMPARISON OF FITUR WRAPPER TECHNIQUE SELECTION USING  
FORWARD SELECTION WITH FILTER TECHNIQUE USING SINGULAR  
VALUE DECOMPOSITION TO DOCUMENT CLUSTERING  
PERFORMANCE

By:

Divinda Morikawa  
09021181520121

**ABSTRACT**

k-Means is the simple clustering algorithm that has fast time of processing data. However, k-Means can cause curse of dimensionality on high dimensional data. From previous study, feature selection using filter technique in Singular Value Decomposition can reduce dimensional data and improve kmeans clustering results. Besides, there is also wrapper technique in feature selection. Forward Selection is wrapper approaches used in feature selection oftenly. It can remove irrelevant feature, develop and improve the quality of data, performance and accurateness of the model. The results obtained by the combination of forward selection and K-Means improve the performance of text clustering by 29,3%, while combination of SVD and k-Means just improve the performance of text clustering by 17,9%. On the other hand, SVD has a faster time compared to forward selection and SVD can reduce the number of repetitions of processes.

**Keywords** *K-Means, fitur selection, forward selection, singular value decomposition.*

**Perbandingan Seleksi Fitur Teknik *Wrapper* Menggunakan *Forward Selection* dengan Teknik *Filter* Menggunakan *Singular Value Decomposition* terhadap Hasil Pengklasteran Dokumen**

Oleh:

Divinda Morikawa

09021181520121

**ABSTRAK**

Salah satu algoritma *clustering* yang sederhana dan memiliki waktu pemrosesan yang cepat adalah *K-Means*. Namun, sering kali *K-Means* mengalami permasalahan seperti *curse of dimensionality* pada data berdimensi tinggi, telah dibahas pada penelitian sebelumnya, dengan menggunakan seleksi fitur teknik filter yaitu *Singular Value Decomposition*, mampu mengurangi dimensi data sehingga meningkatkan hasil kinerja *clustering K-Means*. Disisi lain, ada juga teknik *wrapper* untuk seleksi fitur. *Forward selection* adalah pendekatan *wrapper* yang sering digunakan dalam seleksi fitur, *forward selection* mampu menghapus fitur yang tidak relevan, mengembangkan dan menambah kualitas data, serta meningkatkan performa dan akurasi model. Hasil yang didapatkan kombinasi *forward selection* dan *k-Means* meningkatkan kinerja pengelompokan dokumen sebesar 29,3% dibandingkan kombinasi *SVD* dan *k-Means* hanya meningkatkan sebesar 17,9%. Di sisi lain, *SVD* memiliki waktu komputasi yang lebih cepat dari *forward selection* dan mampu menurunkan jumlah proses pengulangan.

**Kata Kunci :** *k-Means*, seleksi fitur, *forward selection*, *singular value decomposition*.

# **BAB I**

## **PENDAHULUAN**

### **1.1 Pendahuluan**

Pada bab ini membahas latar belakang masalah, rumusan masalah, tujuan, manfaat dan batasan masalah pada penelitian ini. Bab ini akan menjelaskan deskripsi umum tentang keseluruhan penelitian. Pendahuluan dimulai dengan penjelasan mengenai tantangan pengelompokan dokumen. Serta penelitian yang berkaitan dengan teknik reduksi dimensi yaitu seleksi fitur dan proses pengklasteran data berdimensi tinggi.

### **1.1 Latar Belakang**

Pengklasteran merupakan salah satu metode data mining yang bersifat tanpa arahan (*unsupervised*), maksudnya metode ini diterapkan tanpa adanya latihan (*training*) dan tanpa ada guru (*teacher*) serta tidak memerlukan target *output*. Salah satu metode pengklasteran ialah *k-Means*. Metode *k-Means* akan mengelompokkan data yang memiliki karakteristik yang sama kedalam satu kluster dan data yang memiliki karakteristik yang berbeda dikelompokkan dengan kluster yang lain sehingga data yang berada dalam satu kluster memiliki tingkat variasi yang kecil (Agusta,2007; Han et al, 2011; Jumadi, 2014).

*K-Means* merupakan algoritma yang sederhana sehingga mudah dipahami dan sering digunakan pada pengklasteran dokumen . Algoritma *k-Means* memiliki waktu pemrosesan yang cepat (Badriyah, 2014). Namun *k-Means* sebagai algoritma

pengklasteran konvensional hanya mampu mengatasi data berdimensi rendah (Han, Pei, & Kamber, 2011). Sedangkan pengelompokan dokumen seringkali merupakan dokumen berdimensi tinggi. Hal itu disebabkan karena semakin besar dokumen maka fitur yang dihasilkan akan semakin banyak sehingga berdampak pada tingginya dimensi. Sehingga teknik reduksi dimensi dapat dilakukan untuk mengatasi permasalahan tersebut (Supriyanto dan Affandy, 2011; Hidayati et al. 2017).

Penelitian oleh Hasanah et al. (2018) mengatakan bahwa teknik reduksi dimensi dengan seleksi fitur SVD lebih baik dalam hasil pengklasteran dan waktu pemrosesan juga lebih cepat dibandingkan reduksi dimensi dengan PCA.

Teknik reduksi dimensi *Singular Value Decomposition* dikombinasikan dengan metode clustering *k-Means*, *k-Medoids*, dan *fuzzy c-Means* untuk mengelompokkan dokumen. Berdasarkan hasil pengujian, ditemukan bahwa SVD berpengaruh terhadap *clustering k-Means* dan *k-Medoids* (Hidayati et al, 2017; Jambak et al, 2018).

Dua pendekatan umum untuk seleksi fitur otomatis adalah teknik *filter* dan teknik *wrapper*. Teknik *filter* biasanya lebih cepat dalam proses seleksi namun teknik *wrapper* memberikan kinerja yang lebih tinggi (Chandrashekar and Sahin, 2014). Seleksi fitur teknik *filter* dilakukan sebelum algoritma *Data Mining* (pada saat ini klasterisasi) dijalankan. Sedangkan teknik *wrapper* merupakan teknik yang menggunakan algoritma dari *Data Mining* target (pada saat ini klasterisasi) sebagai sebuah *black box* untuk menemukan subset atribut terbaik. Sehingga SVD

merupakan seleksi fitur teknik *filter* karena cara kerja SVD dilakukan sebelum algoritma *Data Mining* dijalankan (Agustine et al., 2008).

Salah satu pendekatan dalam seleksi fitur teknik *wrapper* yaitu *Forward Selection*. *Forward Selection* dilakukan dengan cara memasukkan variabel independen yang signifikan masuk ke dalam model satu persatu hingga tidak ada lagi variabel independen yang signifikan untuk dimasukkan ke dalam model (Ladha and Deepa, 2011).

Purnanditya dan Fanani (2014) melakukan penelitian dengan *dataset* status kelulusan mahasiswa Fasilkom UNAKI dari tahun 2008-2010 untuk pengklasifikasian menggunakan *Naive Bayes*. Hasil penelitiannya menunjukkan bahwa *forward selection* memiliki pengaruh yang baik dalam mengklasifikasikan status kelulusan mahasiswa menggunakan algoritma *Naive Bayes*.

Berdasarkan hal tersebut, maka perlu diuji apakah seleksi fitur tipe *wrapper* menggunakan *Forward Selection* lebih baik dibandingkan dengan seleksi fitur tipe *filter* menggunakan SVD dalam hasil pengklasteran.

## 1.2 Rumusan Masalah

Rumusan masalah dalam penelitian ini ialah bagaimana perbedaan seleksi fitur teknik *wrapper* dengan *Forward Selection* (FS) dibandingkan teknik *filter Singular Value Decomposition* (SVD) terhadap hasil pengklasteran dokumen berdimensi tinggi? Untuk menjawab rumusan masalah tersebut, maka diurai beberapa pertanyaan penelitian sebagai berikut:

1. Bagaimana cara kerja seleksi fitur teknik *wrapper* dengan *Forward Selection* (FS) dan teknik *filter* dengan *Singular Value Decomposition* (SVD) dalam mereduksi data berdimensi tinggi?
2. Bagaimana cara kerja algoritma *k-Means* dalam pengklasteran dokumen?
3. Bagaimana pengaruh seleksi fitur teknik *filter* dengan SVD dan teknik *wrapper* dengan *forward selection* terhadap hasil klaster?

### **1.3 Tujuan Penelitian**

Tujuan dilakukannya penelitian ini ialah sebagai berikut:

1. Memahami perbedaan mekanisme reduksi dimensi dengan metode seleksi fitur dengan *Forward Selection* dan SVD.
2. Memahami mekanisme *k-Means* dalam pengklasteran dokumen.
3. Mengetahui metode seleksi yang lebih baik dalam mereduksi data berdimensi tinggi.

### **1.4 Manfaat Penelitian**

Manfaat yang diperoleh dari penelitian ini ialah sebagai berikut:

1. Dapat mengetahui alternatif lain dalam seleksi fitur selain SVD.
2. Dapat mengetahui teknik reduksi yang sesuai dengan cara kerja *k-Means*.
3. Dapat mereduksi data berdimensi tinggi sehingga mampu melakukan pengklasteran yang lebih baik.

## 1.5 Batasan Masalah

Batasan masalah pada penelitian ini ialah sebagai berikut:

1. Metode pengklasteran hanya menggunakan algoritma *k-Means*.
2. Data yang digunakan merupakan data teks berbahasa Indonesia, sehingga menggunakan kaidah praproses yang diatur dalam tata bahasa Indonesia.
3. Evaluasi hasil pengklasteran akan diukur menggunakan *Davies Bouldin Index* (DBI).

## 1.6 Kesimpulan

Fokus pada penelitian ini ialah mengetahui perbedaan hasil antara seleksi fitur teknik *wrapper* dengan *Forward Selection* dan teknik *filter* dengan SVD terhadap hasil pengklasteran dokumen berdimensi tinggi. Untuk mengetahui pengaruh kedua teknik seleksi fitur tersebut, maka akan dilakukan evaluasi hasil *klaster*. Evaluasi hasil *klaster* diukur dengan perhitungan jarak antar *klaster* dengan *Davis Bouldin Index* (DBI).

Sehingga hasil klaster *k-Means* menggunakan seleksi fitur teknik *wrapper* dengan *Forward Selection* (FS) akan dibandingkan dengan hasil klaster *k-Means* menggunakan seleksi fitur teknik *filter* dengan *Singular Value Decomposition* (SVD).

## DAFTAR PUSTAKA

- Agusta, Y. (2007). 'K-means - Penerapan, Permasalahan dan Metode Terkait'. *Jurnal Sistem dan Informatika*, 3, pp. 47-60.
- Agustine, R.A., Bijaksana, M.A., & Kiki (2008) 'Analisis dan Implementasi Feature Selection Pada Unsupervised Learning (Klasterisasi dengan Metode Wrapper)'.
- Aryani, Fitri., and Yulianti, Dewi., (2012) 'Aplikasi Metode Singular Value Decomposition (SVD) Pada Sistem Persamaan Linier Kompleks', *Sains*, 10(1).
- Badriyah, T. (2014) 'Hybrid Modeling Kmeans-Genetic Algorithms in the Health Care Data', 2(1).
- Chandrashekar, Girish., Sahin, Ferat (2014) 'A Survey On Feature Selection Methods', Elsevier, pp.16-28.
- Ghani, I. M. M., and Ahmad, S., (2011) 'Comparison Methods of Multiple Linear Regressions in Fish Landing', *Australian Journal of Basic and Applied Sciences*, pp. 25-30, 5(1).
- Han, J., Pei, J., & Kamber, M. (2011) 'Data mining: concepts and techniques: Elsevier'.
- Hasanah, S.I.R., Jambak, Saputra, (2018) 'Comparasion of Dimensional Reduction Using Singular Value Decomposition and Principal Component Analysis for Clustering Result of Indonesian Language Text Documents'.
- Hidayati, N., Jambak, Saputra (2017) 'Pengaruh Singular Value Decomposition Terhadap Metode-Metode Klastering', pp.95-104.
- Jambak, M.I., et al. (2018) 'The Impacts of Singular Value Decomposition Algorithm toward Indonesian Language Text Documents Clustering'.
- Jumadi., Winarko, E. (2014). 'Klasifikasi Dokumen Teks Berita Hasil Proses Pengklasteran Yang Tak terkelompokan.
- Kruchten, P. (2000). *The Rational Unified Process An Introduction, Second Edition*. United States of America: Addison Wesley Longman, Inc.
- Ladha, L., Deepa, T. (2011) 'Feature Selection Methods And Algorithms', *IJCSE*, 0975-3397, 3(2).
- Muhammad, A.F., (2015), 'Klasterisasi Proses Seleksi Pemain Menggunakan Algoritma K-Means'.



- Mulyana, Rian., (2012) ‘Aplikasi Metode Variable Selection Untuk Menentukan Faktor Dominan Yang Mempengaruhi Pendidikan Dan Kesehatan’. <http://repository.ipb.ac.id>.
- Nazief, B., M. Adriani, J. Asian, S. M. M. Tahaghoghi and H. E. Williams (2007). "Stemming Indonesian: A confix-stripping approach." *ACM Transactions on Asian Language Information Processing (TALIP)*, **6**(4).
- Prunanditya, B.A., Fanani, A.Z. (2015) . ‘Penerapan Fitur Seleksi Forward Selection Menggunakan Algoritma Naive Bayes Untuk Menentukan Atribut Yang Berpengaruh Pada Klasifikasi Kelulusan Mahasiswa Universitas AKI Semarang’.
- Revathi, M. and T. Ramesh (2010) ‘Network Intrusion Detection System Using Reduced Dimensionality’ *IJCSE*, pp. 61-67, 2(1).
- Saleh, H. (2017). ‘Prediksi Kebangkrutan Perusahaan Menggunakan Algoritma C4.5 Berbasis Forward Selection’.
- Supriyanto, Catur., Affandy. (2011) ‘Kombinasi Teknik Chi Square dan Singular Value Decomposition Untuk Reduksi Fitur Pada Pengelompokan Dokumen’. *Semantik*, 979-26-0255-0.
- Safriandono, A.N., (2017) ‘ Algoritma K-Nearest Neighbor Berbasis Forward Selection untuk Mendiagnosa Penyakit Jnatung Koroner’, *Komputaki*, 3(1).
- Santosa, B. (2007). ‘Data Mining: Teknik Pemanfaatan Data untuk Keperluan Bisnis’. Yogyakarta: Graha Ilmu.
- Zhang, S. and R. Jing (2011) ‘Dimension Reduction Based on Modified Maximum Margin Criterion for Tumor Classification’ *ICIC*, pp.552-554.