

**PENERAPAN ALGORITMA *K-MEANS* UNTUK *CLUSTERING*
DATA MAHASISWA**

(Studi Kasus: Universitas Katolik De La Salle Manado)

TUGAS AKHIR

Disusun Oleh:

Jonathan Adrian Tontong

17013001



PROGRAM STUDI TEKNIK INFORMATIKA

FAKULTAS TEKNIK

UNIVERSITAS KATOLIK DE LA SALLE

MANADO

2021

PENERAPAN ALGORITMA *K-MEANS* UNTUK *CLUSTERING* DATA MAHASISWA

(Studi Kasus: Universitas Katolik De La Salle Manado)

LAPORAN TUGAS AKHIR

Ditulis untuk memenuhi sebagian persyaratan untuk mendapat gelar Sarjana
Teknik pada Program Studi Teknik Informatika

Disusun Oleh:

Jonathan Adrian Tontong

17013001



PROGRAM STUDI TEKNIK INFORMATIKA

FAKULTAS TEKNIK

UNIVERSITAS KATOLIK DE LA SALLE

MANADO

2021

LEMBAR PERNYATAAN

Yang bertanda tangan di bawah ini:

Nama : Jonathan Adrian Tontong
NIM : 17013001
Tempat/Tanggal Lahir : Naha/4 Januari 2000
Fakultas/Program Studi : Teknik/Teknik Informatika

Menyatakan bahwa Laporan Tugas Akhir dan Aplikasi yang berjudul "Penerapan Algoritma *K-Means* Untuk *Clustering* Data Mahasiswa" yang saya buat adalah benar hasil karya saya sendiri dan bukan karya tulis orang lain, baik sebagian atau seluruhnya kecuali dalam bentuk kutipan yang telah disebutkan sumbernya di dalam laporan ini.

Demikianlah surat ini saya buat dengan sebenar-benarnya dan apabila pernyataan ini tidak benar, maka saya bersedia menerima sanksi akademis sesuai dengan yang ditetapkan oleh Fakultas Teknik, berupa pembatalan nilai Tugas Akhir.

Manado, 1 Agustus 2021
Yang Menyatakan,



Jonathan Adrian Tontong

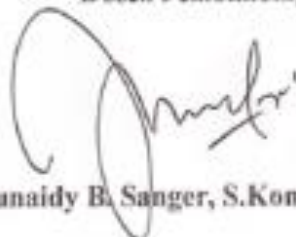
Menyetujui,

Dosen Pembimbing I



Angelia Melani Adrian, Ph.D.

Dosen Pembimbing II



Junaidy B. Sanger, S.Kom., M.Kom.

Mengetahui,

Ketua Program Studi



Vivie D. Kumenap, S.T., M.Cs.

Dekan Fakultas Teknik



Ronald A. Rachmadi, S.T., M.T.



UNIVERSITAS KATOLIK DE LA SALLE
MANADO - INDONESIA

Nama : Jonathan Adrian Tontong
NIM : 17013001
Fakultas : Teknik
Program Studi : Teknik Informatika
Judul Tugas Akhir : Penerapan Algoritma *K-Means* Untuk *Clustering* Data Mahasiswa
Pembimbing I : Angelia Melani Adrian, Ph.D.
Pembimbing II : Junaidy Budi Sanger, S.Kom., M.Kom.

Menyetujui,

Manado, 1 Agustus 2021

Dosen Pembimbing I

Dosen Pembimbing II

Angelia Melani Adrian, Ph.D.

Junaidy B. Sanger, S.Kom., M.Kom.

Mengetahui,

Ketua Program Studi

Dekan Fakultas Teknik

Vivie D. Komenap, S.T., M.Cs.

Ronald A. Rachmadi, S.T., M.T.

ABSTRACT

New Student Registration at De La Salle Catholic University Manado (UKDLSM) is opened every year starting from February. Many students apply for admission to this university. Most prospective students comes within the province of North Sulawesi, even many who are come from outside the province such as the provinces of Gorontalo, South Sulawesi, Papua and Maluku. After being accepted as a student at UKDLSM, the data from these students will be stored in the UKDLSM student database. This student data will be useful and produce a lot of information if processed properly. Through this data, it can be seen the distribution of students entering this university from which area and type of high school or equivalent vocational school, and many more. Therefore, this study developed a data mining-based application to cluster student data at UKDLSM using the KMeans Clustering algorithm. The application is built using the Python programming language with the data modeling used is Data Flow Diagrams (DFD). The data processed in this study is 485 student data, consisting of 4 parameters including district/city of origin, study program, school origin and major in school of origin. The clustering process is carried out by selecting the value of $K = 3$ according to the student data which consists of 3 study programs with the largest number of students, and produces the amount of data in each cluster, namely: cluster 1 408 data, cluster 2 19 data and cluster 3 58 data, with the number of iterations: 7 iterations and Silhouette Score: 0.56.

Keywords: Student Data, Data Mining, K-Means Clustering, Clustering.

ABSTRAK

Pendaftaran Mahasiswa Baru di Universitas Katolik De La Salle Manado (UKDLSM) dibuka setiap tahun dimulai dari bulan Februari. Banyak mahasiswa yang mendaftar untuk masuk ke universitas ini. Para calon mahasiswa mayoritas berasal dari dalam daerah di provinsi Sulawesi Utara, bahkan banyak juga yang berasal dari luar provinsi tersebut seperti provinsi Gorontalo, Sulawesi Selatan, Papua dan Maluku. Setelah diterima menjadi mahasiswa di UKDLSM, data-data dari mahasiswa ini akan tersimpan dalam *database* mahasiswa UKDLSM. Data mahasiswa ini akan berguna dan menghasilkan banyak informasi jika diolah dengan baik. Melalui data ini dapat dilihat sebaran mahasiswa yang masuk ke universitas ini berasal dari daerah mana dan jenis sekolah SMA atau SMK sederajat, serta masih banyak lagi. Oleh sebab itu, penelitian kali ini mengembangkan aplikasi berbasis *data mining* untuk melakukan klasterisasi terhadap data mahasiswa di UKDLSM menggunakan algoritma *KMeans Clustering*. Aplikasi di bangun dengan menggunakan bahasa pemrograman Python dengan pemodelan data yang digunakan adalah *Data Flow Diagram* (DFD). Data yang diolah pada penelitian ini berjumlah 485 data mahasiswa, yang terdiri dari 4 parameter antara lain kabupaten/kota asal, program studi, asal sekolah dan jurusan di sekolah asal. Proses klasterisasi dilakukan dengan memilih Nilai $K=3$ sesuai dengan data mahasiswa yang terdiri dari 3 program studi dengan jumlah mahasiswa terbanyak, dan menghasilkan Jumlah Data Pada Tiap klaster yaitu: klaster 1 sebanyak 408 data, 2 sebanyak 19 data dan klaster 3 sebanyak 58 data, dengan jumlah proses iterasi: 7 Iterasi dan *Silhouette Score*: 0.56.

Kata kunci: Data Mahasiswa, *Data Mining*, *K-Means Clustering*, Klasterisasi.

KATA PENGANTAR

Puji dan syukur Penulis panjatkan kepada Tuhan Yang Maha Kuasa karena atas berkat dan kemurahanNya, Penulis dapat menyelesaikan laporan Tugas Akhir yang berjudul Penerapan Algoritma *K-Means* Untuk *Clustering* Data Mahasiswa.

Karena dalam proses pembuatan aplikasi dan penyusunan laporan Tugas Akhir, penulis mendapat banyak bantuan, saran, dan arahan dari berbagai pihak, maka dari itu penulis ingin mengucapkan banyak terima kasih kepada:

1. Prof. Dr. Johanis Ohoitumur selaku Rektor Universitas Katolik De La Salle Manado.
2. Bapak Ronald Albert Rachmadi, S.T., M.T selaku Dekan Fakultas Teknik.
3. Bapak Thomas Christian Suwanto, S.Kom, M.Mm selaku Wakil Dekan Fakultas Teknik.
4. Ibu Vivie D. Kumenap, S.T., M.Cs selaku Ketua Program Studi Teknik Informatika.
5. Ibu Angelia Melani Adrian, Ph.D selaku Dosen Pembimbing 1 yang telah banyak membantu memberikan arahan dan bimbingan dalam pembuatan aplikasi dan penyusunan laporan.
6. Bapak Junaidy Budi Sanger, S.Kom., M.Kom selaku Dosen Pembimbing 2 yang telah banyak membantu memberikan arahan dan bimbingan dalam pembuatan aplikasi dan penyusunan laporan.
7. Bapak Apriandy Angdresey, S.T., M.Sc selaku Dosen Pembimbing Akademik yang telah membimbing selama penulis menjalani masa perkuliahan..
8. Seluruh Dosen dan Staf di Universitas Katolik De La Salle Manado, terlebih khusus pada Fakultas Teknik dan Program Studi Teknik Informatika atas semua ilmu dan motivasi yang telah diberikan.
9. Ayah, Ibu, Adik dan seluruh keluarga yang selalu memberikan dukungan, motivasi dan doa.
10. Teman-teman seperjuangan (Aldo, Bragy, Eston, Falentino, Hizkia, Kenny, Ryan dan Krisna) yang selalu membantu mendukung dan saling menopang.
11. Sahabat-sahabat terkasih, Ronaldo Tangkulung, Juolanda Surati, Anggriani Najoan dan Agnesia Yusup.
12. Sahabat-sahabat angkatan 2017 Program Studi Teknik Informatika, UKM *Young Lasallian Corps*, BPA UKDLSM (Ibu Merry, Ibu Fransiska, Ka Aurel, Richard), Ibu Aneth, Ibu Yoan, Ka Angel Fong, Ka Billy Ponto dan semua pihak yang selalu memberikan semangat dan dukungan, dan
13. Sahabat Beliau (Angga Macis, Brenda, Adit, Eka, Rudolfo, Kenny, Agung)

Penulis menyadari bahwa dalam penyusunan laporan tugas akhir ini, tidak terlepas dari banyak kekurangan. Untuk itu kritik dan saran dari pembaca sangat penulis harapkan.

Manado, Agustus 2021

Penulis

DAFTAR ISI

JUDUL	i
LEMBAR PERNYATAAN	ii
LEMBAR PENGESAHAN	iii
ABSTRACT	iv
ABSTRAK	v
KATA PENGANTAR	vi
DAFTAR ISI	vii
DAFTAR TABEL	ix
DAFTAR GAMBAR	x
DAFTAR LAMPIRAN	xi
DAFTAR FORMULA	xii
BAB I PENDAHULUAN	Error! Bookmark not defined.
1.1 Latar Belakang	Error! Bookmark not defined.
1.2 Rumusan Masalah	Error! Bookmark not defined.
1.3 Tujuan Tugas Akhir	Error! Bookmark not defined.
1.4 Manfaat Tugas Akhir	Error! Bookmark not defined.
1.5 Batasan Masalah	Error! Bookmark not defined.
1.6 Metodologi Penelitian	Error! Bookmark not defined.
1.7 Sistematika Penulisan	Error! Bookmark not defined.
BAB II STUDI PUSTAKA	Error! Bookmark not defined.
2.1 Penerimaan Mahasiswa Baru (PMB)	Error! Bookmark not defined.
2.2 <i>Data Mining</i>	Error! Bookmark not defined.
2.2.1 <i>Clustering</i>	Error! Bookmark not defined.
2.2.2 <i>K-Means Algorithm</i>	Error! Bookmark not defined.
2.2.3 <i>Silhouette Coefficient</i>	Error! Bookmark not defined.
2.3 Teknologi Pengembangan Aplikasi	Error! Bookmark not defined.
2.3.1 <i>Hyper Text Markup Language</i> (HTML) ..	Error! Bookmark not defined.
2.3.2 <i>Python</i>	Error! Bookmark not defined.
2.4 Penelitian Terkait	Error! Bookmark not defined.
2.5 Kakas Pemodelan	Error! Bookmark not defined.
2.5.1 <i>Flowchart</i>	Error! Bookmark not defined.
2.5.2 <i>Data Flow Diagram</i> (DFD)	Error! Bookmark not defined.
2.6 Metodologi Pengembangan Perangkat Lunak	Error! Bookmark not defined.
BAB III ANALISIS	Error! Bookmark not defined.
3.1 Analisis Sistem	Error! Bookmark not defined.
3.1.1 Identifikasi Masalah	Error! Bookmark not defined.
3.1.2 Sumber Daya Pendukung	Error! Bookmark not defined.

3.1.3	Identifikasi Spesifikasi Fungsional Sistem	Error! Bookmark not defined.
3.2	Identifikasi Pengguna	Error! Bookmark not defined.
3.3	Analisis Deskriptif	Error! Bookmark not defined.
3.3.1	<i>Data Selection</i>	Error! Bookmark not defined.
3.3.2	<i>Data Processing</i>	Error! Bookmark not defined.
3.3.3	<i>Data Transformation</i>	Error! Bookmark not defined.
3.3.4	<i>Clustering Algoritma K-Means</i>	Error! Bookmark not defined.
BAB IV	PERANCANGAN	Error! Bookmark not defined.
4.1	Pemodelan Sistem	Error! Bookmark not defined.
4.1.1	<i>Flowchart</i>	Error! Bookmark not defined.
4.1.2	<i>Data Flow Diagram</i>	Error! Bookmark not defined.
4.2	<i>Storyboard</i>	Error! Bookmark not defined.
BAB V	IMPLEMENTASI	Error! Bookmark not defined.
5.1	Lingkungan Implementasi	Error! Bookmark not defined.
5.1.1	Lingkungan Perangkat Keras	Error! Bookmark not defined.
5.1.2	Lingkungan Perangkat Lunak	Error! Bookmark not defined.
5.2	Implementasi Antarmuka	Error! Bookmark not defined.
5.3	Implementasi Pengodean	Error! Bookmark not defined.
BAB VI	PENGUJIAN	Error! Bookmark not defined.
6.1	Lingkungan Pengujian	Error! Bookmark not defined.
6.2	Tujuan Pengujian	Error! Bookmark not defined.
6.3	Kriteria Pengujian	Error! Bookmark not defined.
6.4	Kasus Pengujian	Error! Bookmark not defined.
6.5	Pelaksanaan Pengujian	Error! Bookmark not defined.
6.5.1	Pengujian Fitur	Error! Bookmark not defined.
6.5.2	Pengujian Nilai K	Error! Bookmark not defined.
6.5.3	Pengujian <i>User Acceptance Test</i> (UAT)..	Error! Bookmark not defined.
6.5.4	Pengujian <i>Web Browser</i>	Error! Bookmark not defined.
6.6	Analisis Hasil Pengujian	Error! Bookmark not defined.
BAB VII	KESIMPULAN DAN SARAN	Error! Bookmark not defined.
7.1	Kesimpulan	Error! Bookmark not defined.
7.2	Saran	Error! Bookmark not defined.
DAFTAR PUSTAKA	Error! Bookmark not defined.
LAMPIRAN		

DAFTAR TABEL

Tabel 2.1 Daftar Fakultas dan Program Studi di UKDLSM	Error! Bookmark not defined.
Tabel 2.2 Notasi <i>Flowchart</i>	Error! Bookmark not defined.
Tabel 2.3 Notasi <i>Data Flow Diagram</i> (DFD)	Error! Bookmark not defined.
Tabel 3.1 Daftar Sumber Daya Pendukung	Error! Bookmark not defined.
Tabel 3.2 Data Mahasiswa Tahun 2019-2020	Error! Bookmark not defined.
Tabel 3.3 Keterangan Untuk Mengisi Setiap Parameter	Error! Bookmark not defined.
Tabel 3.4 Tabel Inisialisasi Data	Error! Bookmark not defined.
Tabel 3.5 Daftar <i>Centroid</i> Awal	Error! Bookmark not defined.
Tabel 3.6 Hasil Perhitungan Iterasi 1	Error! Bookmark not defined.
Tabel 3.7 Pembaruan <i>Centroid</i> dan Nilai Rata-Rata Tiap <i>Cluster</i>	Error! Bookmark not defined.
Tabel 3.8 Hasil Perhitungan Iterasi 2	Error! Bookmark not defined.
Tabel 3.9 Hasil Perhitungan Iterasi 7	Error! Bookmark not defined.
Tabel 4.1 Proses DFD Level 1	Error! Bookmark not defined.
Tabel 4.2 Proses DFD Level 2	Error! Bookmark not defined.
Tabel 5.1 Lingkungan Perangkat Keras	Error! Bookmark not defined.
Tabel 5.2 Lingkungan Perangkat Lunak	Error! Bookmark not defined.
Tabel 6.1 Pelaksanaan Pengujian	Error! Bookmark not defined.
Tabel 6.2 Hasil Pengujian Nilai K dan <i>Silhouette Score</i>	Error! Bookmark not defined.
Tabel 6.3 Hasil Pengujian <i>Web Browser</i>	Error! Bookmark not defined.
Tabel 6.4 Data Hasil Proses Klasterisasi	Error! Bookmark not defined.
Tabel 6.5 Hasil Data pada tiap Klaster	Error! Bookmark not defined.

DAFTAR GAMBAR

Gambar 2.1 Ilustrasi Model <i>Waterfall</i>	Error! Bookmark not defined.
Gambar 4.1 <i>Flowchart K-Means Clustering</i>	Error! Bookmark not defined.
Gambar 4.2 <i>Flowchart</i> Aplikasi Klasterisasi Data Mahasiswa ..	Error! Bookmark not defined.
Gambar 4.3 Diagram Konteks.....	Error! Bookmark not defined.
Gambar 4.4 <i>Data Flow Diagram Level 1</i>	Error! Bookmark not defined.
Gambar 4.5 <i>Data Flow Diagram Level 2</i>	Error! Bookmark not defined.
Gambar 4.6 <i>Storyboard</i> Aplikasi Klasterisasi Data Mahasiswa .	Error! Bookmark not defined.
Gambar 4.7 <i>Storyboard</i> Halaman Visualisasi Data	Error! Bookmark not defined.
Gambar 5.1 Tampilan Antarmuka Aplikasi	Error! Bookmark not defined.
Gambar 5.2 Tampilan Halaman Visualisasi Data ..	Error! Bookmark not defined.
Gambar 6.1 Pengujian Nilai K dan <i>Silhouette Score</i>	Error! Bookmark not defined.

DAFTAR LAMPIRAN

LAMPIRAN A *USER ACCEPTANCE TEST ...* A-Error! Bookmark not defined.
LAMPIRAN B DATA MAHASISWA TAHUN 2019-2020 .B-Error! Bookmark not defined.

DAFTAR FORMULA

(2.1) Rumus <i>Euclidean</i>	10
(2.2) Rumus BCV	11
(2.3) Rumus WCV	11
(2.4) Rumus Rasio	11
(2.5) Rumus <i>Silhouette Coefficient</i> (1)	12
(2.6) Rumus <i>Silhouette Coefficient</i> (2)	12
(2.7) Rumus <i>Silhouette Coefficient</i> (3)	12

