



***CLUSTERING DATA REKOMENDASI PENJURUSAN KE
PERGURUAN TINGGI MENGGUNAKAN METODE K-MEANS
STUDI KASUS SMA NEGERI 2 PALEMBANG***

R. Ahmad Dicky Syarief Purboyo

171420124

Skripsi

Sebagai salah satu syarat memperoleh gelar Sarjana

PROGRAM STUDI TEKNIK INFORMATIKA

FAKULTAS ILMU KOMPUTER

UNIVERSITAS BINA DARMA

PALEMBANG

2022

HALAMAN PENGESAHAN

**Clustering Data Rekomendasi Penjurusan Ke Perguruan
Tinggi Menggunakan Metode K-Means Studi Kasus SMA N 2
Palembang**

R. AHMAD DICKY SYARIEF PURBOYO
171420124

Telah diterima sebagai salah satu syarat untuk memperoleh gelar
Sarjana Komputer pada Program Studi Teknik Informatika

Pembimbing



Dr. H. Jemakmun, M.Si.

Palembang, 23 Maret 2022
Fakultas Ilmu Komputer
Universitas Bina Darma
Dekan,



Universitas **Bina Darma**
Fakultas Ilmu Komputer

Dedy Syamsuar, S.Kom., M.I.T., Ph.D.

HALAMAN PERSETUJUAN

Skripsi Berjudul "Clustering Data Rekomendasi Penjurusan Ke Perguruan Tinggi Menggunakan Metode K-Means Studi Kasus SMA N 2 Palembang" Oleh "R. Ahmad Dicky Syarif Purboyo", telah dipertahankan di depan komisi penguji pada hari Rabu tanggal 23 Maret 2022.

Komisi Penguji

1. Ketua : Dr. H. Jemaknun, M.Si.
2. Anggota : Hadi Syaputra, M.Kom
3. Anggota : Siti Sauda, M.Kom



(.....)





(.....)



(.....)

Mengetahui,
Program Studi Teknik Informatika
Fakultas Ilmu Komputer
Universitas Bina Darma
Ketua,

Universitas Bina Darma
Fakultas Ilmu Komputer



Alek Wijaya, S.Kom., M.I.T.

SURAT PERNYATAAN

Saya yang bertanda tangan dibawah ini :

Nama : R. AHMAD DICKY SYARIEF PURBOYO

Nim : 171420124

Dengan ini menyatakan Bahwa :

1. Karya tulis saya (Skripsi) ini adalah asli dan belum pernah diajukan untuk mendapatkan gelar akademik (sarjana) di Universitas Bina Darma atau perguruan tinggi lainnya;
2. Karya tulis ini murni gagasan, rumusan dan penelitian saya sendiri dengan arahan pembimbing;
3. Didalam karya tulis ini tidak terdapat karya atau pendapat yang telah ditulis di publikasikan orang lain, kecuali secara tertulis dengan jelas dikutip dengan mencantumkan nama pengarang dan memasukan ke daftar pustaka;
4. Saya bersedia tugas Skripsi, di cek keasliannya menggunakan plagiarism checker serta di unggah ke internet, sehingga dapat diakses secara daring;
5. Surat pernyataan ini saya tulis dengan sungguh-sungguh dan apabila terbukti melakukan penyimpangan atau ketidakbenaran dalam pernyataan ini maka saya bersedia menerima sanksi dengan peraturan dan perundang-undangan yang berlaku;

Demikian surat pernyataan ini saya buat agar dapat dipergunakan sebagaimana mestinya.

Palembang, 26 Maret 2022

Yang Membuat Pernyataan,



d Dicky Syarief Purboyo

MOTTO DAN PERSEMBAHAN

MOTTO

"Jangan pernah menyerah hanya karena satu kali kesalahan, karena seharusnya kita belajar dari kesalahan untuk tahu kebenaran."

PERSEMBAHAN

- ALLAH SWT yang selalu memberikan jalan dan kemudahan dalam hal apapun.
- Kepada kedua Orangtua saya yang selalu memberikan doa dan dukungan penuh.
- Kepada Dosen Pembimbing Bapak Dr. H. Jemakmun, M.Si. yang telah memberikan bimbingan dan arahan dalam menyelesaikan Karya tulis penelitian Skripsi ini
- Kepada teman seperjuangan kuliah terkhusus teman-teman kelas D Teknik Informatika Universitas Bina Darma Angkatan 2017 yang sedikit membantu dalam proses penyelesaian skripsi ini.

ABSTRAK

SMA Negeri 2 Palembang adalah salah satu sekolah menengah atas yang terletak di kota Palembang siswa-siswinya yang terbagi menjadi 2 jurusan yaitu IPA dan IPS. Akibat salah memilih jurusan setelah masuk kuliah siswa/siswi tersebut mengalami kesulitan dan merasa salah jurusan sehingga penulis memberikan solusi untuk penentuan jurusan ke perguruan tinggi menggunakan metode *k-means clustering* dengan metode ini nantinya siswa-siswi dapat mengetahui jurusan apa yang cocok bagi mereka setelah mereka lulus sekolah dan ingin melanjutkan pendidikan ke perguruan tinggi. Dalam penelitian ini siswa-siswi dikelompokkan dengan menerapkan metode data mining. Kelompok didasarkan pada atribut jurusan, minat, sifat, hobi, bakat, dan nilai rata-rata mata pelajaran jurusan ipa dan jurusan ips. untuk melakukan clustering data digunakan metode K-Means dan metode pengukuran jarak Euclidean dan dianalisis menggunakan perhitungan manual menggunakan *Microsoft excel* dan tools *RapidMiner*. Dari hasil penelitian perhitungan manual dengan menggunakan metode K-Means terbentuk menjadi 5 cluster dengan 4 iterasi. Cluster 1 merupakan cluster yang di rekomendasikan mengambil jurusan Bahasa. Cluster 2 merupakan cluster yang direkomendasikan mengambil jurusan Teknik. Cluster 3 merupakan cluster yang direkomendasikan mengambil jurusan Kesehatan/Kedokteran. Cluster 4 merupakan cluster yang di rekomendasikan mengambil jurusan Ekonomi. Cluster 5 merupakan cluster yang direkomendasikan mengambil jurusan Bahasa. Sedangkan hasil dari Penelitian perhitungan menggunakan *RapidMiner Cluster_0* direkomendasikan mengambil jurusan Teknik, *Cluster_1* direkomendasikan mengambil jurusan Ekonomi, *Cluster_2* Direkomendasikan mengambil jurusan Bahasa, *Cluster_3* direkomendasikan mengambil jurusan Pendidikan dan, *Cluster_4* direkomendasikan mengambil jurusan Kedokteran/Kesehatan Sehingga dengan adanya penelitian ini dapat membantu siswa-siswi SMA N 2 Palembang untuk memilih jurusan dengan tepat.

Kata Kunci : *K-Means*, Jurusan , Clustering,

ABSTRACT

SMA Negeri 2 Palembang is one of the senior high schools located in the city of Palembang. The students are divided into 2 classes, namely Science and Social Sciences. As a result of choosing the wrong department after entering college, these students experienced difficulties and felt that they were in the wrong department, so the author provides a solution for determining department for college using the k-means clustering method with this method later students can find out what department are suitable for them after they graduate from school and want to continue their education to college. In this Research the students were grouped by applying the data mining method. The group is based on the attributes of classes, interests, natures, hobbies, talents, and the average value of the subjects of the science department and the social sciences department. To perform data clustering used the K-Means method and the Euclidean distance measurement method and analyzed using manual calculations using Microsoft Excel and RapidMiner tools. From the results of research using the K-Means method formed into 5 clusters. Cluster 1 is a cluster that is recommended to take a language department. Cluster 2 is a cluster that is recommended to join in Engineering Department. Cluster 3 is a cluster that is recommended to join in Medical Department. Cluster 4 is a cluster that is recommended to enter in Economics Department. Cluster 5 is the recommended cluster for entering in Language Department. While the results of the calculation research using RapidMiner, Cluster_0 is recommended to join in Engineering department, Cluster_1 is recommended to join in Economics department, Cluster_2 is recommended to join in Languages department, Cluster_3 is recommended to join in Education department and, Cluster_4 is recommended to join in Medicine/Health department. So that this research can help high school students N 2 Palembang to choose the right department.

Keywords : K-Means, Department, Clustering,

KATA PENGANTAR

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

Puji syukur kehadirat Allah SWT karena berkat rahmat dan karunia nya – nya jugalah penelitian skripsi ini dapat diselesaikan guna memenuhi salah satu syarat untuk diteruskan sebagai proses akhir dalam menyelesaikan pendidikan di bangku kuliah

Dalam penelitian skripsi ini, tentunya masih jauh dari sempurna. Hal ini dikarenakan keterbatasan pengetahuan yang dimiliki. Oleh karena itu dalam rangka melengkapi kesempurnaan dari penulisan skripsi ini diharapkan adanya saran dan kritik yang diberikan bersifat membangun. Pada kesempatan yang baik ini, tak lupa penulis mengucapkan terimakasih kepada semua pihak yang telah memberikan bimbingan, pengalaman, nasihat, dan pemikiran dalam penulisan skripsi ini, terutama kepada :

1. Dr. Sunda Ariana, M.Pd., MM selaku Rektor Universitas Bina Darma Palembang.
2. Dedy Syamsuar, Ph.D. selaku Dekan Fakultas Ilmu Komputer Universitas Bina Darma Palembang
3. Alex Wijaya, S.Kom., M.I.T. Selaku Ketua Program Studi Teknik Informatika Universitas Bina Darma Palembang

4. Dr. H. Jemakmun, M.Si selaku pembimbing yang telah memberikan bimbingan dan arahan dalam penulisan Skripsi ini.
5. Kedua orangtua yang telah memberikan doa dan dukungan penuh dalam penyusunan Skripsi ini.
6. Kepada seluruh dosen dan mahasiswa Universitas Bina Darma yang telah membantu atas terlaksananya Skripsi tersebut.
7. Kepada teman dekat dan teman teman seperjuangan Program Studi Teknik Informatika Angkatan 2017. Demikianlah yang dapat saya sampaikan, dan harapan saya semoga ini dapat memberikan manfaat bagi kita semua.
8. Kepada Ibu Ina Wahyuli, S.Ag & Ibu Wiwiek Endah Noviyanti, S.Si yang telah mempermudah saat penelitian di SMA N 2 Palembang serta guru – guru yang tidak bisa saya sebutkan semuanya.

Palembang, 26 Maret 2022

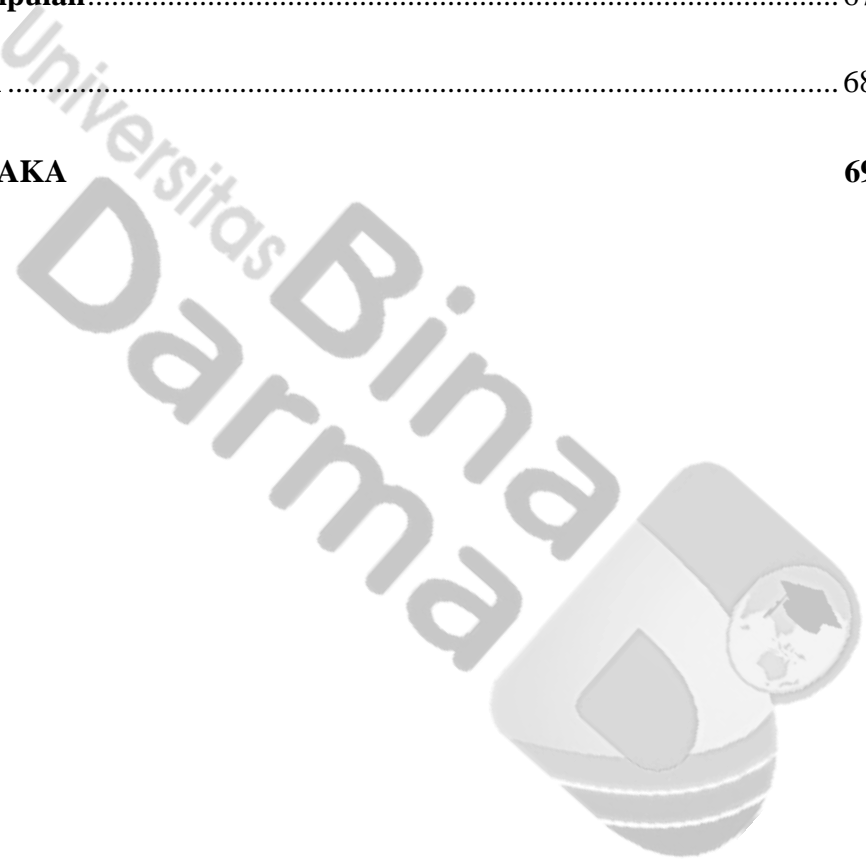
Penulis

DAFTAR ISI

HALAMAN JUDUL.....	i
HALAMAN PENGESAHAN	i
HALAMAN PERSETUJUAN	ii
HALAMAN PERNYATAAN	iii
MOTTO DAN PERSEMBAHAN	iv
ABSTRAK	v
ABSTRACT	vi
KATA PENGANTAR.....	vii
DAFTAR ISI	ix
DAFTAR GAMBAR	xii
BAB I	1
PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	4
1.3 Batasan Masalah	4
1.4 Tujuan Penelitian	5
1.5 Manfaat Penelitian	5
1.6 Metodologi Penelitian	6
1.7 Sistematika Penulisan	13
BAB II	15
TINJAUAN PUSTAKA	15

2.1.	<i>Data Mining</i>	15
2.2.	<i>Clustering (Klasterisasi)</i>	16
2.3.	<i>RapidMiner</i>	17
2.4.	<i>K-Means</i>	18
2.5.	Penelitian Sebelumnya	21
2.6.	Kerangka Berpikir	22
BAB III		24
	ANALISIS DATA MINING	24
3.1.	Metode Analisis Data	24
BAB IV		35
	HASIL DAN PEMBAHASAN	35
4.1.	<i>Representasi Data</i>	35
4.2.	Perhitungan <i>K-Means Clustering</i>	37
4.3.	Hasil Analisis	53
BAB V		55
	HASIL DAN PEMBAHASAN APLIKASI	55
5.1.	Proses <i>Data Mining</i> menggunakan <i>RapidMiner</i>	55
5.2	Penerapan <i>K-Means Clustering</i> Pada <i>RapidMiner</i>	60
5.3	Hasil Pola <i>RapidMiner</i>	64

BAB VI	67
KESIMPULAN DAN SARAN	67
6.1 Kesimpulan	67
6.2 Saran	68
DAFTAR PUSTAKA	69



DAFTAR GAMBAR

Gambar 1.1 Proses Knowledge Discovery in Database (KDD)	9
Gambar 1.2 Penerapan Metode K-Means.	10
Gambar 2.1 Kerangka Berpikir	23
Gambar 3.1 Daftar Siswa Jurusan IPA	26
Gambar 3.2 Daftar Siswa Jurusan IPS	27
Gambar 3.3 Hasil Data Duplikat	28
Gambar 3.4 Hasil Data Cleaning	29
Gambar 5.1 Tampilan New Process	55
Gambar 5.2 Langkah Pertama Data Import Wizard	56
Gambar 5.3 Langkah Kedua Data Import Wizard	57
Gambar 5.4 Langkah Ketiga Data Import Wizard	58
Gambar 5.5 Langkah Keempat Data Import Wizard	59
Gambar 5.6 Pemodelan <i>Clustering K-means Rapidminer</i>	60
Gambar 5.7 Tampilan Hasil <i>Cluster</i> Pada <i>Data view</i>	61
Gambar 5.8 <i>Result History</i>	63
Gambar 5.9 Tampilan <i>Folder View</i>	64
Gambar 5.10 Pola Output <i>Scatter 3D</i>	65
Gambar 5.11 Pola <i>output cluster Bar (Horizontal)</i>	66

DAFTAR TABEL

Tabel 3.1 Inisialisasi Jurusan	30
Tabel 3.2 Inisialisasi Hobi	30
Tabel 3.3 Inisialisasi Minat	31
Tabel 3.4 Inisialisasi Sifat	33
Tabel 3.5 Inisialisasi Bakat	34
Tabel 4.1 Sampel Data Siswa	36
Tabel 4.2 Centroid Awal	38
Tabel 4.3 Hasil Perhitungan Jarak Data Terhadap Titik Pusat Pada Iterasi 1	40
Tabel 4.4 Centroid baru pada iterasi ke-2	44
Tabel 4.5 Hasil Perhitungan Jarak Data Terhadap Centroid Pada Iterasi 3	44
Tabel 4.6 Hasil Perhitungan Jarak Data Terdekat terhadap Centroid Pada Iterasi 3	46
Tabel 4.7 Hasil Perhitungan Jarak Data Terhadap Centroid Pada Iterasi 4	49
Tabel 4.8 Hasil Perhitungan Jarak Data Terdekat Terhadap Centroid Pada Iterasi 4	50
Tabel 4.9 Hasil Akhir Pengelompokkan Siswa Terhadap Masing-Masing Cluster.....	52
Tabel 5.1 Hasil Pengelompokkan Cluster dengan RapidMiner Berdasarkan ID	61