

**KLASTERISASI MAHASISWA UNIVERSITAS SULTAN
AGENG TIRTAYASA ANGKATAN 2021/2022 DENGAN
ALGORITMA *K-PROTOTYPE***

SKRIPSI



Oleh:

FINDI CHOERUN ANNISA

3333200111

**JURUSAN TEKNIK INDUSTRI
FAKULTAS TEKNIK
UNIVERSITAS SULTAN AGENG TIRTAYASA
CILEGON – BANTEN**

2024

**KLASTERISASI MAHASISWA UNIVERSITAS SULTAN
AGENG TIRTAYASA ANGKATAN 2021/2022 DENGAN
ALGORITMA *K-PROTOTYPE***

**“Skripsi ditulis untuk memenuhi sebagian persyaratan dalam mendapatkan
gelar Sarjana Teknik”**



Oleh:

FINDI CHOERUN ANNISA

3333200111

**JURUSAN TEKNIK INDUSTRI
FAKULTAS TEKNIK
UNIVERSITAS SULTAN AGENG TIRTAYASA
CILEGON – BANTEN
2024**

PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan dibawah ini:

NAMA : FINDI CHOERUN ANNISA

NIM : 3333200111

JURUSAN : TEKNIK INDUSTRI

JUDUL : KLASTERISASI MAHASISWA UNIVERSITAS SULTAN
AGENG TIRTAYASA ANGKATAN 2021/2022 DENGAN
ALGORITMA *K-PROTOTYPE*

Dengan ini menyatakan bahwa penelitian dengan judul tersebut diatas adalah benar karya saya sendiri dengan arahan dari pembimbing I, dan tidak ada duplikasi dengan karya orang lain kecuali yang telah disebutkan sumbernya.

Apabila dikemudian hari terbukti terdapat plagiat dalam penelitian ini, maka saya bersedia menerima sanksi sesuai peraturan perundang-undangan yang berlaku.

Cilegon, 3 Januari 2024



FINDI CHOERUN ANNISA

HALAMAN PENGESAHAN

Skripsi ini diajukan Oleh:

NAMA : FINDI CHOERUN ANNISA
NIM : 3333200111
JURUSAN : TEKNIK INDUSTRI
JUDUL : KLASTERISASI MAHASISWA UNIVERSITAS SULTAN
AGENG TIRTAYASA ANGKATAN 2021/2022 DENGAN
ALGORITMA *K-PROTOTYPE*

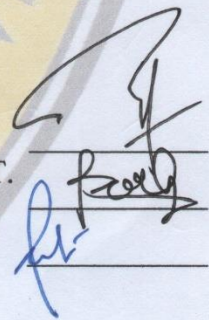
Telah berhasil dipertahankan di hadapan Dewan Penguji dan Diterima
sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar
Sarjana Teknik pada Jurusan Teknik Industri, Fakultas Teknik,
Universitas Sultan Ageng Tirtayasa

Pada hari : Rabu

Tanggal : 3 Januari 2024

DEWAN PENGUJI

Pembimbing 1 : Dr. H. M. Adha Ilhami, ST., MT.
Penguji 1 : Dr. Eng. Ir. H. Bobby Kurniawan ST., MT.
Penguji 2 : Dr. Ir. Ratna Ekawati, ST., MT., IPP.



Mengetahui,
Ketua Jurusan Teknik Industri
Ade Irman Saeful Muttaqin S. ST., MT.
NIP. 1982061520121211002

PRAKATA

Puji syukur penulis panjatkan kepada Allah SWT, yang telah memberikan rahmat, petunjuk, dan hidayah-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul “Klasterisasi Mahasiswa Universitas Sultan Ageng Tirtayasa Angkatan 2021/2022 Dengan Algoritma *K-Prototype*” sebagai persyaratan dalam mendapatkan gelar Sarjana Teknik pada Jurusan Teknik Industri Fakultas Teknik Universitas Sultan Ageng Tirtayasa. Penulisan skripsi ini bertujuan untuk menyajikan hasil dari penelitian yang telah dilakukan. Selama penyelesaian skripsi ini, kami mendapat bantuan, arahan dan dorongan dari berbagai pihak. Oleh karena itu, penulis mengucapkan rasa syukur dan terima kasih kepada:

1. Kedua orang tua dan kakak-kakak penulis yang selalu memberikan dukungan baik moril maupun materil, motivasi, dan doa demi keberhasilan skripsi ini.
2. Bapak H. Syafik Fahmi selaku keluarga yang telah membantu secara materil dan memberikan inspirasi agar menjadi orang sukses.
3. Bapak Dr. H. Muhammad Adha Ilhami S.T., M.T. dan Bapak Aditya Rahadian Fachrur S.Si., M.MT. selaku dosen yang membimbing dan memberikan nasihat dalam kelancaran skripsi ini.
4. Bapak Dr. Eng. Ir. H. Bobby Kurniawan S.T., M.T. dan Ibu Dr. Ir. Ratna Ekawati, S.T., M.T., IPP. selaku dosen yang penguji yang memberikan nasihat serta saran dalam penyusunan skripsi.
5. Dosen pembimbing mata kuliah metodologi penelitian Ibu Dr. Lely Herlina S.T., M.T.
6. Bapak Ade Irman Saeful Mutaqin S,S.T., M.T. selaku ketua Jurusan Teknik Industri Fakultas Teknik Universitas Sultan Ageng Tirtayasa.
7. Ibu Yusraini Muharni S.T., M.T. selaku koordinator tugas akhir.
8. Ibu Dr. Ade Sri Mariawati ST., MT. selaku dosen wali yang telah membantu berbagai hal Ketika penulis membutuhkan perizinan atau saran serta nasihat.

9. Ibu Kulsum ST., MT. dan Ibu Nustin Merdiana Dewantari S.T., M.T. selaku sosok dosen yang telah memberikan kesempatan penulis untuk belajar banyak hal.
10. Ibu Dr. Faula Arina, S.Si., M.Si. selaku dosen yang telah memberikan kesempatan penulis untuk bergabung pada kelas analisis multivariat yang membantu dalam penyusunan skripsi ini.
11. Ibu Dr. Shanti Kirana Anggraeni, SP., MT. selaku dosen yang telah memberikan kesempatan penulis untuk bergabung pada kelas metode penelitian yang membantu dalam penyusunan skripsi ini.
12. Bapak dan Ibu Dosen Jurusan Teknik Industri.
13. Hafiz Ramadhani selaku seseorang yang sabar dan pengertian yang memiliki peran dalam menemani, memberikan dukungan, dan membantu pada saat suka dan duka keadaan penulis, dalam menyelesaikan skripsi ini.
14. Keluarga cemara yang membantu, menghibur, dan pemberi semangat dalam kelancaran skripsi ini.
15. Abang Bernard yang bersedia bertukar pikiran di berbagai hal.
16. Teman-teman angkatan 2020 Jurusan Teknik Industri Fakultas Teknik Universitas Sultan Ageng Tirtayasa dan rekan-rekan telah berbagi pengalaman, diskusi, dan dukungan dalam menyelesaikan skripsi ini.
17. Serta seluruh pihak yang tidak dapat disebutkan satu per satu yang telah membantu penulis dalam pengerjaan skripsi ini.

Penulis menyadari bahwa skripsi ini tidak luput dari kesalahan. Oleh karena itu, penulis mengharapkan kritik dan saran yang akan datang untuk perbaikan skripsi ini. Akhir kata penulis mengucapkan terima kasih dan harapan semoga skripsi ini dapat bermanfaat bagi semua pihak yang berkepentingan dan memberikan kontribusi bagi kemajuan keilmuan Teknik Industri.

Cilegon, 3 Januari 2024



FINDI CHOERUN ANNISA

ABSTRAK

FINDI CHOERUN ANNISA. KLASTERISASI MAHASISWA UNIVERSITAS SULTAN AGENG TIRTAYASA ANGGARAN 2021/2022 DENGAN ALGORITMA *K-PROTOTYPE*. DIBIMBING OLEH DR. MUHAMMAD ADHA ILHAMI, S.T., M.T.

Kualitas suatu perguruan tinggi dapat dilihat dari peringkat perguruan tinggi tersebut. Universitas Sultan Ageng Tirtayasa (UNTIRTA) merupakan perguruan tinggi yang berada di Provinsi Banten. yang berupaya bersaing dengan perguruan tinggi lainnya dengan berusaha meningkatkan peringkatnya. Salah satu upaya yang dilakukan UNTIRTA dengan cara melakukan identifikasi potensi mahasiswa UNTIRTA untuk mengetahui faktor-faktor yang dapat digunakan dalam menentukan karakteristik calon mahasiswa yang diharapkan memiliki peluang sukses dalam menjalankan studinya berdasarkan karakteristik mahasiswa yang menempuh studi di UNTIRTA dalam memberikan rekomendasi usulan yang berguna sebagai media evaluasi penerimaan mahasiswa baru di tahun selanjutnya. terkhususnya pada jalur penerimaan Seleksi Nasional Berdasarkan Prestasi (SNBP). Metode yang digunakan dalam mengidentifikasi faktor-faktor yang direkomendasikan dalam evaluasi penerimaan mahasiswa baru adalah dengan menggunakan metode seperti one-way ANOVA, K-prototype, dan General Linear Model (GLM). Hasil pengujian dari beberapa uji yang telah dilakukan didapatkan faktor-faktor yang perlu dipertimbangkan menjadi rekomendasi untuk penerimaan SNBP adalah kategori SMTA, kWh listrik, dan bidikmisi.

Kata kunci: ANOVA, K-Prototype, General Linear Model, SNBP.

ABSTRACT

FINDI CHOERUN ANNISA. CLUSTERING OF SULTAN AGENG TIRTAYASA UNIVERSITY STUDENTS CLASS OF 2021/2022 WITH THE K-PROTOTYPE ALGORITHM. GUIDED BY DIBIMBING OLEH DR. MUHAMMAD ADHA ILHAMI, S.T., M.T.

The quality of a university can be seen from its ranking. Sultan Ageng Tirtayasa University (UNTIRTA) is a university located in Banten Province. that seeks to compete with other universities by trying to improve its ranking. One of the efforts made by UNTIRTA by identifying the potential of UNTIRTA students to determine the factors that can be used in determining the characteristics of prospective students who are expected to have a chance of success in carrying out their studies based on the characteristics of students studying at UNTIRTA is providing recommendations for proposals that are useful as a medium for evaluating new student admissions in the following year, especially in the Seleksi Nasional Berdasarkan Prestasi (SNBP) admission path. The method used in identifying recommended factors in evaluating new student admissions is to use methods such as one-way ANOVA, K-prototype, and General Linear Model (GLM). The test results from several tests that have been carried out indicate that factors that need to be considered as recommendations for SNBP admission are high school category, electricity kWh, and bidikmisi.

Keyword: ANOVA, K-Prototype, General Linear Model, SNBP

RINGKASAN

FINDI CHOERUN ANNISA. KLASIFIKASI MAHASISWA UNIVERSITAS SULTAN AGENG TIRTAYASA ANGKATAN 2021/2022 DENGAN ALGORITMA *K-PROTOTYPE*. DIBIMBING OLEH MUHAMMAD ADHA ILHAMI

Latar belakang; Kualitas suatu perguruan tinggi dapat dilihat dari peringkat perguruan tinggi tersebut. Universitas Sultan Ageng Tirtayasa (UNTIRTA) merupakan perguruan tinggi yang memiliki lokasi pusat di Banten, yang berupaya bersaing dengan perguruan tinggi lainnya dengan berusaha meningkatkan peringkatnya. Upaya yang dilakukan UNTIRTA dalam berusaha meraih peringkat perguruan tinggi yang lebih tinggi dengan cara melakukan identifikasi potensi mahasiswa UNTIRTA untuk mengetahui faktor-faktor yang dapat digunakan dalam menentukan karakteristik calon mahasiswa yang diharapkan memiliki peluang sukses dalam menjalankan studinya berdasarkan karakteristik mahasiswa yang menempuh studi di UNTIRTA. Data yang digunakan dalam mengidentifikasi potensi mahasiswa UNTIRTA adalah data mahasiswa UNTIRTA angkatan 2021/2022. Data mahasiswa UNTIRTA angkatan 2021/2022. Algoritma *K-Prototype* bertujuan untuk mengelompokkan mahasiswa UNTIRTA berdasarkan beberapa variabel seperti Indeks Prestasi Akademik, jenis kelamin, jalur masuk, bidikmisi dan lain sebagainya agar mengetahui karakteristik mahasiswa yang dengan kelompok kualitas yang baik sehingga dapat menjadi rekomendasi usulan yang berguna sebagai media evaluasi penerimaan mahasiswa baru terkhususnya pada jalur usulan untuk penerimaan mahasiswa baru di tahun selanjutnya terkhususnya untuk jalur Seleksi Nasional Berdasarkan Prestasi (SNBP).

Perumusan masalah; Penelitian ini memiliki rumusan masalah berupa peningkatan kualitas UNTIRTA dalam persaingan dengan perguruan tinggi lainnya

sehingga memerlukan faktor-faktor yang penting untuk dipertimbangkan dalam seleksi penerimaan jalur masuk SNBP.

Tujuan penelitian; Penelitian ini bertujuan agar UNTIRTA mengetahui faktor-faktor yang penting untuk dipertimbangkan dalam seleksi penerimaan jalur masuk SNBP.

Metode penelitian; Penelitian ini menggunakan jenis penelitian kuantitatif dengan menggunakan data sekunder dengan teknik pengambilan data dari keseluruhan mahasiswa UNTIRTA Angkatan 2021/2022.

Hasil penelitian; Penelitian ini mendapatkan hasil berupa variabel-variabel signifikan berdasarkan *one-way* ANOVA seperti jenis kelamin, jalur masuk, kategori provinsi, kategori SMTA, kWh listrik, dan bidikmisi. Selanjutnya hasil klusterisasi dengan *K-Prototype* didapatkan 2 kluster optimal. Hasil dari klusterisasi diketahui bahwa faktor-faktor yang direkomendasikan dari hasil GLM adalah kategori SMTA, kategori kWh listrik, dan penerima kategori bidikmisi.

Kata kunci: *ANOVA, K-Prototype, General Linear Model, SNBP.*

DAFTAR ISI

	Halaman
HALAMAN SAMPUL.....	i
HALAMAN JUDUL.....	ii
PERNYATAAN KEASLIAN SKRIPSI.....	iii
HALAMAN PENGESAHAN.....	iv
PRAKATA.....	v
ABSTRAK.....	vii
<i>ABSTRACT</i>	viii
RINGKASAN.....	ix
DAFTAR ISI.....	xi
DAFTAR TABEL.....	xv
DAFTAR GAMBAR.....	xvi
DAFTAR ARTI LAMBANG, SINGKATAN, DAN ISTILAH.....	xvii
DAFTAR LAMPIRAN.....	xx
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	2
1.3 Tujuan Penelitian.....	3
1.4 Batasan Masalah.....	3
1.5 Sistematika Penulisan.....	3
1.6 Penelitian Terdahulu.....	4
BAB II TINJAUAN PUSTAKA.....	7
2.1 <i>Data Mining</i>	7

2.2	Statistika Deskriptif.....	8
2.3	Klasterisasi	9
2.4	Teorema Limit Pusat	9
2.5	<i>One-Way Analyze of Variance (One-Way ANOVA)</i>	10
2.6	Algoritma <i>K-Prototype</i>	10
2.7	<i>Preprocessing Data Mining</i>	11
2.8	Evaluasi Pengelompokan	11
2.8.1	Uji Validitas Kelompok.....	12
2.8.1.1	Ukuran Internal.....	12
2.8.1.2	Ukuran Eksternal	13
2.9	<i>General Linear Model (GLM)</i>	13
BAB III METODE PENELITIAN		15
3.1	Rancangan Penelitian	15
3.2	Lokasi Penelitian dan Waktu Penelitian	15
3.3	Cara Pengambilan Data.....	15
3.4	Alur Pemecahan Masalah.....	16
3.4.1	<i>Flowchart</i> Pemecahan Masalah.....	16
3.4.2	<i>Flowchart</i> Pengolahan Data	17
3.5	Deskripsi Pemecahan Masalah.....	18
3.5.1	Deskripsi <i>Flowchart</i> Pemecahan Masalah	18
3.5.2	Deskripsi <i>Flowchart</i> Pengolahan Data.....	19
3.6	Analisis Data	21
BAB IV HASIL PENELITIAN.....		23
4.1	Pengumpulan Data	23
4.1.1	Deskripsi Karakteristik Variabel	23

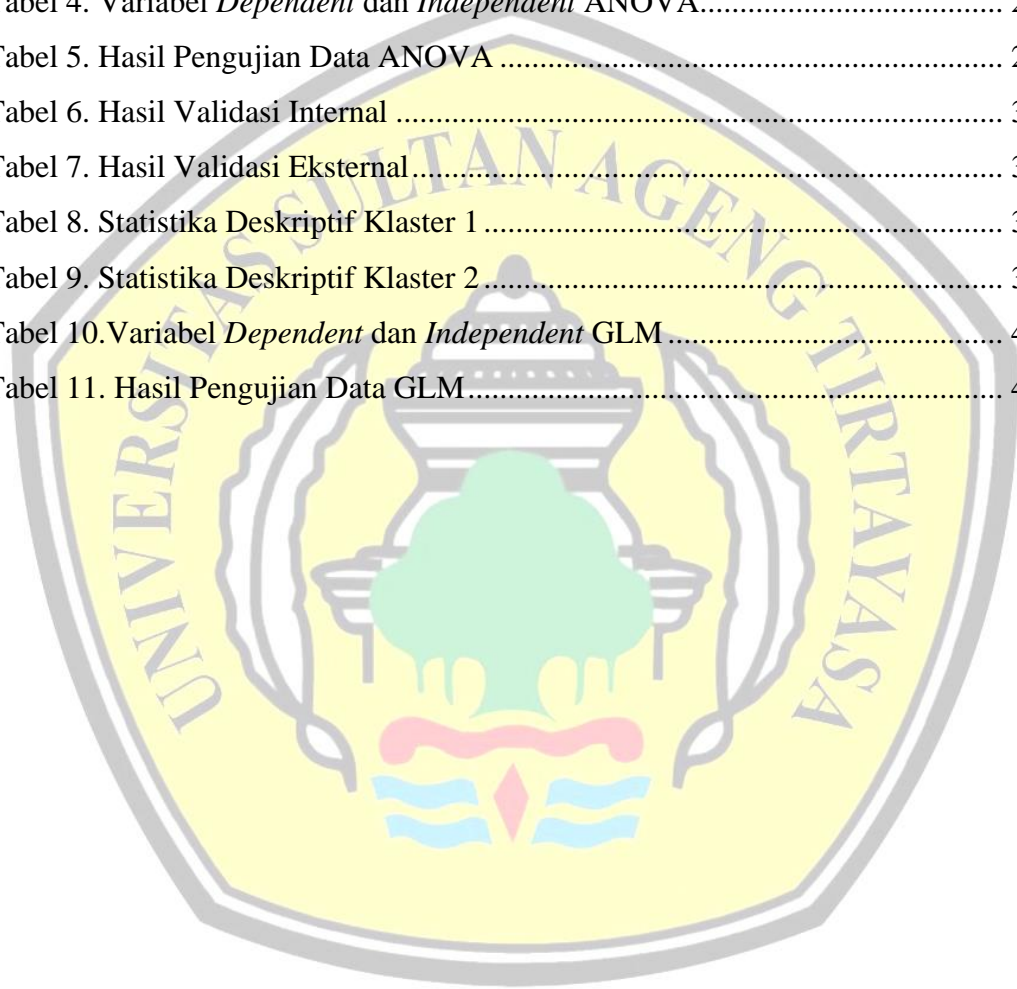
4.1.1.1	Deskripsi Karakteristik Variabel Identitas.....	23
4.1.1.2	Deskripsi Karakteristik Variabel Numerik	24
4.1.1.3	Deskripsi Karakteristik Variabel Kategori	25
4.2	Pengolahan Data.....	25
4.2.1	<i>Preprocessing</i> Data	25
4.2.2	Menentukan Variabel Signifikan dengan <i>One -Way ANOVA</i>	27
4.2.2.1	Menentukan Jenis Variabel Pengujian pada <i>One-Way ANOVA</i>	27
4.2.2.2	Hasil Pengujian Data <i>One-Way ANOVA</i>	27
4.2.3	Algoritma <i>K-Prototype</i>	28
4.2.4	Karakteristik Hasil Pengelompokan Metode <i>K-Prototype</i>	31
4.2.5	Perhitungan Indeks Validasi Internal Metode <i>K-Prototype</i>	31
4.2.6	Perhitungan Indeks Validasi Eksternal Metode <i>K-Prototype</i>	31
4.2.7	Statistika Deskriptif Kluster	32
4.2.7.1	Klaster 1	32
4.2.7.2	Klaster 2.....	33
4.2.8	Perbandingan Kluster.....	34
4.2.8.1	Karakteristik Jenis Kelamin.....	34
4.2.8.2	Karakteristik Jalur Masuk.....	36
4.2.8.3	Karakteristik Kategori Provinsi	38
4.2.8.4	Karakteristik Kategori SMTA	40
4.2.8.5	Karakteristik Kategori kWh Listrik	43
4.2.8.6	Karakteristik Kategori Bidikmisi.....	45
4.2.9	Menentukan Variabel Signifikan dengan <i>General Linear Model (GLM)</i> 46	
4.2.9.1	Menentukan Jenis Variabel Pengujian pada <i>General Linear Model (GLM)</i>	47

4.2.9.2	Hasil Pengujian Data <i>General Linear Model</i> (GLM).....	47
BAB V ANALISIS DAN PEMBAHASAN.....		49
5.1	Analisis Variabel-Variabel yang Signifikan terhadap Variabel IPK	49
5.2	Analisis Jumlah Klasterisasi Optimal dan Karakteristiknya.....	50
5.3	Rekomendasi Faktor-Faktor yang Penting untuk Dipertimbangkan dalam Seleksi Penerimaan Jalur Masuk SNBP.....	52
BAB V I KESIMPULAN DAN SARAN		54
6.1	Kesimpulan	54
6.2	Saran.....	54
DAFTAR PUSTAKA		56
LAMPIRAN.....		59
DAFTAR RIWAYAT HIDUP PENULIS		66



DAFTAR TABEL

Tabel 1. Ringkasan Data	23
Tabel 2. Deskripsi Karakteristik Data Numerik.....	24
Tabel 3. Variabel Kategori.....	26
Tabel 4. Variabel <i>Dependent</i> dan <i>Independent</i> ANOVA.....	27
Tabel 5. Hasil Pengujian Data ANOVA	28
Tabel 6. Hasil Validasi Internal	31
Tabel 7. Hasil Validasi Eksternal.....	31
Tabel 8. Statistika Deskriptif Klaster 1	32
Tabel 9. Statistika Deskriptif Klaster 2.....	33
Tabel 10. Variabel <i>Dependent</i> dan <i>Independent</i> GLM.....	47
Tabel 11. Hasil Pengujian Data GLM.....	47



DAFTAR GAMBAR

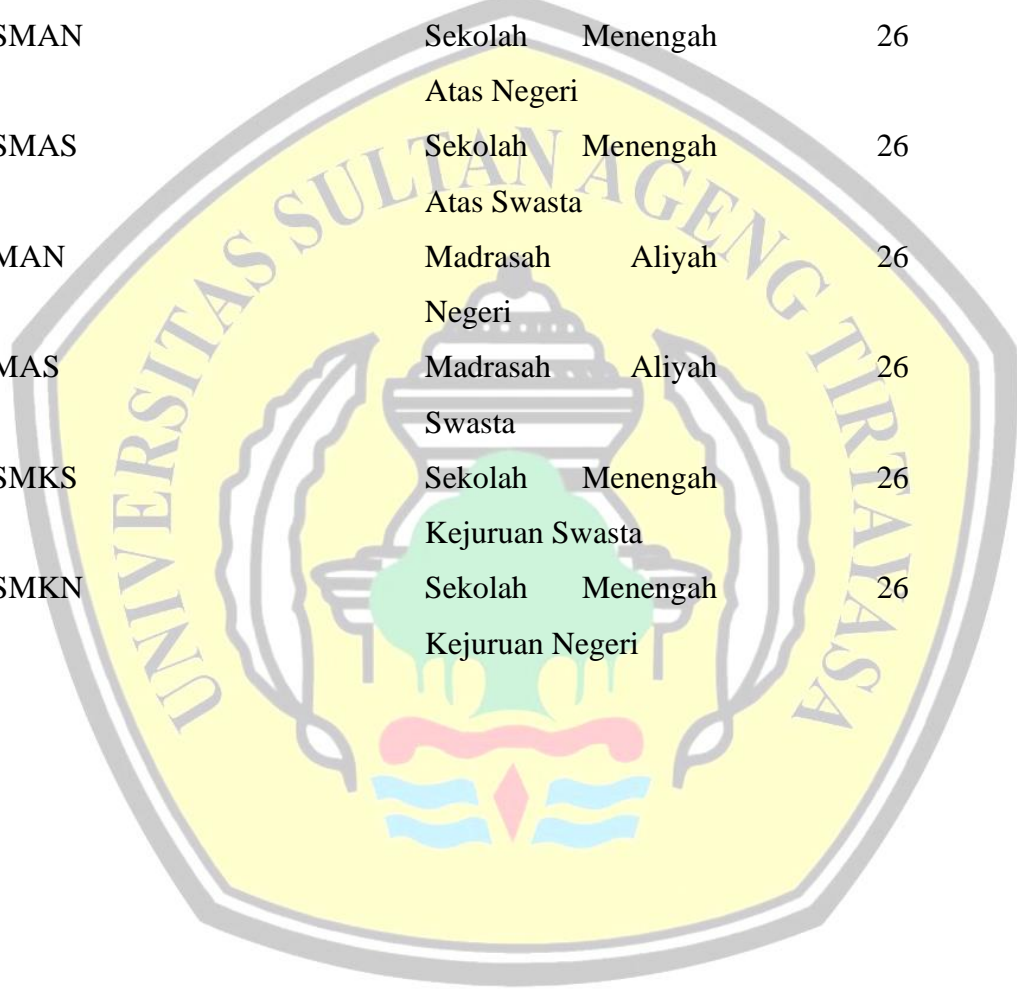
Gambar 1. <i>Flowchart</i> Pemecahan Masalah	16
Gambar 2. <i>Flowchart</i> Pengolahan Data	17
Gambar 3. Karakteristik Jenis Kelamin Berdasarkan Keseluruhan Data	35
Gambar 4. Karakteristik Jenis Kelamin Berdasarkan Anggota Klaster	36
Gambar 5. Karakteristik Jalur Masuk Berdasarkan Keseluruhan Data	37
Gambar 6. Karakteristik Jalur Masuk Berdasarkan Anggota Klaster	38
Gambar 7. Karakteristik Kategori Provinsi Berdasarkan Keseluruhan Data	39
Gambar 8. Karakteristik Kategori Provinsi Berdasarkan Anggota Klaster	40
Gambar 9. Karakteristik Kategori SMTA Berdasarkan Keseluruhan Data	41
Gambar 10. Karakteristik Kategori SMTA Berdasarkan Anggota Klaster	42
Gambar 11. Karakteristik Kategori kWh Listrik Berdasarkan Keseluruhan Data	43
Gambar 12. Karakteristik Kategori kWh Listrik Berdasarkan Anggota Klaster ..	44
Gambar 13. Karakteristik Kategori Bidikmisi Berdasarkan Keseluruhan Data ...	45
Gambar 14 Karakteristik Kategori Bidikmisi Berdasarkan Anggota Klaster	46

DAFTAR ARTI LAMBANG, SINGKATAN, DAN ISTILAH

LAMBANG/ SINGKATAN	Nama	Pemakaian Pertama kali pada halaman
UNTIRTA	Universitas Sultan Ageng Tirtayasa	vii
SNBP	Seleksi Nasional Berdasarkan Prestasi	vii
ANOVA	<i>Analyze of Variance</i>	vii
GLM	<i>General Linear Model</i>	vii
SMTA	Sekolah Menengah Tinggi Atas	vii
kWh	Kilowatt Hour	vii
PTN	Perguruan Tinggi Negeri	1
IPK	Indeks Prestasi Kumulatif	2
SNMPTN,	Seleksi Nasional Masuk Perguruan Tinggi Negeri	2
SBMPTN	Seleksi Bersama Masuk Perguruan Tinggi Negeri	3
SMMPTN	Seleksi Mandiri Masuk Perguruan Tinggi Negeri	3
S1	Strata 1	3
\bar{X}	Rata-rata	8
X	Jumlah nilai	8

LAMBANG/ SINGKATAN	Nama	Pemakaian Pertama kali pada halaman
n	Banyaknya data	8
S	Standar deviasi	9
$a(i)$	Rata-rata perbedaan dari i -objek dengan semua objek lain pada kelompok yang sama	12
$b(i)$	Objek pada kelompok lain (di kelompok terdekat)	12
GS_n	Indeks <i>global silhouette</i>	12
$S(i)$	<i>Silhouette</i> kelompok ke- i ,	12
D	Indeks validasi <i>dunn</i> ,	13
$d'(c_k)$	Jarak dalam kelompok (c_k).	13
E	<i>Entropy</i> ,	13
n_{ik}	Banyaknya anggota pada kelas ke- i yang terdapat pada kelompok k	13
n_k	Banyaknya anggota kelompok ke- k .	13
g	Banyaknya kelompok	13
NIM	Nomor Induk Mahasiswa	15
Me	Median	24
L	Laki-laki	26
P	Perempuan	26
PNS	Pegawai Negeri Sipil	26

LAMBANG/ SINGKATAN	Nama	Pemakaian Pertama kali pada halaman
TNI	Tentara Nasional Indonesia	26
ABRI	Angkatan Bersenjata Republik Indonesia	26
Prof	Profesional	26
SMAN	Sekolah Menengah Atas Negeri	26
SMAS	Sekolah Menengah Atas Swasta	26
MAN	Madrasah Aliyah Negeri	26
MAS	Madrasah Aliyah Swasta	26
SMKS	Sekolah Menengah Kejuruan Swasta	26
SMKN	Sekolah Menengah Kejuruan Negeri	26



DAFTAR LAMPIRAN

Lampiran 1. Data Mahasiswa UNTIRTA 2021/2022	60
Lampiran 2. Hasil <i>One-Way</i> ANOVA pada Variabel Jenis Kelamin dengan Variabel IPK.....	60
Lampiran 3. Hasil <i>One-Way</i> ANOVA pada Variabel Jalur Masuk dengan Variabel IPK	60
Lampiran 4. Hasil <i>One-Way</i> ANOVA pada Variabel Kategori Provinsi dengan Variabel IPK.....	61
Lampiran 5. Hasil <i>One-Way</i> ANOVA pada Variabel Kategori Pulau dengan Variabel IPK.....	61
Lampiran 6. Hasil <i>One-Way</i> ANOVA pada Variabel Pekerjaan Ayah dengan Variabel IPK.....	61
Lampiran 7. Hasil <i>One-Way</i> ANOVA pada Variabel Tahun Lulus SMTA dengan Variabel IPK.....	61
Lampiran 8. Hasil <i>One-Way</i> ANOVA pada Variabel kWh Listrik dengan Variabel IPK	62
Lampiran 9. Hasil <i>One-Way</i> ANOVA pada Variabel Bidikmisi dengan Variabel IPK	62
Lampiran 10. Hasil <i>One-Way</i> ANOVA pada Variabel Kategori SMTA dengan Variabel IPK.....	62
Lampiran 11. <i>Syntax K-Prototype</i>	63
Lampiran 12. Hasil Klaster	65
Lampiran 13. Hasil <i>General Linear Model</i>	65

BAB I

PENDAHULUAN

1.1 Latar Belakang

Kualitas suatu perguruan tinggi dapat dilihat dari peringkat perguruan tinggi tersebut, oleh karena itu perguruan tinggi perlu membuktikan kualitasnya dengan meraih peringkat yang lebih baik. Perguruan tinggi perlu memperhatikan bahwa peringkat pada perguruan tinggi memiliki pengaruh pada minat mahasiswa baru maupun dunia akademis oleh karena itu seluruh perguruan tinggi di dunia termasuk di Indonesia berkompetisi memperoleh peringkat yang lebih baik (Hadjon, 2022).

Universitas Sultan Ageng Tirtayasa (UNTIRTA) merupakan sebuah Perguruan Tinggi Negeri (PTN) yang berada di Jalan Raya Palka No.Km 3, Sindangsari, Kec. Pabuaran, Kota Serang, Provinsi Banten. Berdasarkan pemeringkatan yang dilakukan Edurank, UNTIRTA meraih peringkat ke-62 dalam peringkat nasional dan peringkat ke-4068 dalam peringkat dunia (Edurank, 2023). UNTIRTA berupaya bersaing dengan perguruan tinggi lainnya dengan berusaha meningkatkan peringkatnya. Upaya yang dilakukan UNTIRTA dalam berusaha meraih peringkat perguruan tinggi yang lebih tinggi dengan cara melakukan identifikasi potensi mahasiswa UNTIRTA untuk mengetahui faktor-faktor yang dapat digunakan dalam menentukan karakteristik calon mahasiswa yang diharapkan memiliki peluang sukses dalam menjalankan studinya berdasarkan karakteristik mahasiswa yang menempuh studi di UNTIRTA. Data yang digunakan dalam mengidentifikasi potensi mahasiswa UNTIRTA adalah data mahasiswa UNTIRTA angkatan 2021/2022. Data mahasiswa UNTIRTA angkatan 2021/2022 dipilih sebagai perwakilan dari keseluruhan mahasiswa UNTIRTA serta data tersebut memiliki karakteristik data yang lebih rapi dan lengkap dibandingkan data mahasiswa untuk angkatan lainnya sehingga akan memberikan hasil analisis yang lebih baik.

Algoritma *K-Prototype* telah membantu dalam melakukan klusterisasi asal sekolah pada tingkat Sekolah Menengah Atas (SMA) dalam melihat asal sekolah yang baik yang dapat memberikan calon mahasiswa baru yang memiliki potensi dalam menyelesaikan studinya di perguruan tinggi (Sulastri, et al., 2021). Klusterisasi dengan algoritma *K-Prototype* membagi data menjadi kluster-kluster sehingga data yang memiliki kesamaan karakteristik dikelompokkan, dan data yang berbeda karakteristiknya dikelompokkan dalam kluster lainnya dengan melihat kesamaan dari data dapat tersebut dapat mengetahui faktor seperti variabel yang signifikan (Virtusena, et al., 2021).

Penentuan faktor-faktor yang signifikan dibantu dengan uji *Analyze of Variance* (ANOVA), yang di mana uji ini membantu mengetahui variabel yang signifikan agar ketika dilakukan klusterisasi hasil yang pengelompokan yang didapatkan baik. Algoritma *K-Prototype* bertujuan untuk mengelompokkan mahasiswa UNTIRTA berdasarkan beberapa variabel seperti Indeks Prestasi Kumulatif (IPK), jenis kelamin, jalur masuk, bidikmisi dan lain sebagainya agar mengetahui karakteristik mahasiswa dengan kelompok kualitas yang baik. Hasil dari pengolahan *K-Prototype* dilakukan pengujian dengan *General Linear Model* (GLM) dalam memberikan rekomendasi usulan yang berguna sebagai media evaluasi penerimaan mahasiswa baru di tahun selanjutnya. terkhususnya pada jalur penerimaan Seleksi Nasional Berdasarkan Prestasi (SNBP). SNBP merupakan salah satu jalur dalam seleksi penerimaan mahasiswa baru di PTN yang mengganti jalur sebelumnya yang diberi nama Seleksi Nasional Masuk Perguruan Tinggi Negeri (SNMPTN).

1.2 Rumusan Masalah

Rumusan masalah berdasarkan latar belakang dalam penelitian ini adalah sebagai berikut:

1. Apakah variabel-variabel yang signifikan terhadap variabel IPK?
2. Bagaimana mengetahui jumlah klusterisasi optimal dan karakteristiknya?
3. Bagaimana memberikan rekomendasi faktor-faktor yang penting untuk dipertimbangkan dalam seleksi penerimaan jalur masuk SNBP?

1.3 Tujuan Penelitian

Tujuan penelitian berdasarkan rumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Mengetahui variabel-variabel yang signifikan terhadap variabel IPK.
2. Mengetahui jumlah klusterisasi optimal dan karakteristiknya.
3. Memberikan rekomendasi faktor-faktor yang penting untuk dipertimbangkan dalam seleksi penerimaan jalur masuk SNBP.

1.4 Batasan Masalah

Batasan masalah yang terdapat pada penelitian ini adalah sebagai berikut:

1. Data mahasiswa dengan status akademik “Aktif” dan “Cuti” angkatan 2021/2022.
2. Data mahasiswa strata-1 (S1) di Universitas Sultan Ageng Tirtayasa.
3. Data mahasiswa lulusan SMTA 2019, 2020, dan 2021.
4. Data mahasiswa jalur masuk SNMPTN, SBMPTN, dan SMMPTN.
5. Mahasiswa memiliki kelengkapan data.
6. Mahasiswa yang tidak memiliki IPK bernilai “0”.
7. Transkrip nilai IPK mahasiswa 2021/2022 terakhir pada semester ganjil 2022/2023.

1.5 Sistematika Penulisan

Sistematika penulisan yang terdapat dalam penelitian ini adalah sebagai berikut:

BAB I PENDAHULUAN

Bab pendahuluan berfungsi sebagai acuan dalam penelitian ini. Pendahuluan bab ini meliputi latar belakang, rumusan masalah, tujuan penelitian, batasan masalah, sistematika penulisan, dan penelitian terdahulu.

BAB II TINJAUAN PUSTAKA

Bab tinjauan pustaka menjelaskan teori-teori yang akan digunakan untuk mengatasi permasalahan penelitian. Teori yang digunakan didukung oleh kutipan berbagai buku, jurnal, publikasi ilmiah, dan penelitian sebelumnya yang akan membantu dalam memecahkan masalah.

BAB III METODE PENELITIAN

Bab metodologi penelitian meliputi *flowchart* dan deskripsi penelitian. Bab ini memuat informasi mengenai rancangan penelitian, lokasi dan lama penelitian, teknik pengumpulan data, alur penelitian, gambaran alur penelitian, dan analisis data.

BAB IV HASIL PENELITIAN

Bab ini menjelaskan tentang data yang dikumpulkan dan diolah oleh peneliti untuk mencapai tujuan penelitian yang telah ditentukan.

BAB V ANALISIS DAN PEMBAHASAN

Bab ini akan memberikan gambaran dan penjelasan yang komprehensif tentang analisis data yang telah diolah secara menyeluruh berdasarkan penggunaan kerangka teori, dilanjutkan dengan membandingkan hasil dengan penelitian sebelumnya.

BAB VI KESIMPULAN DAN SARAN

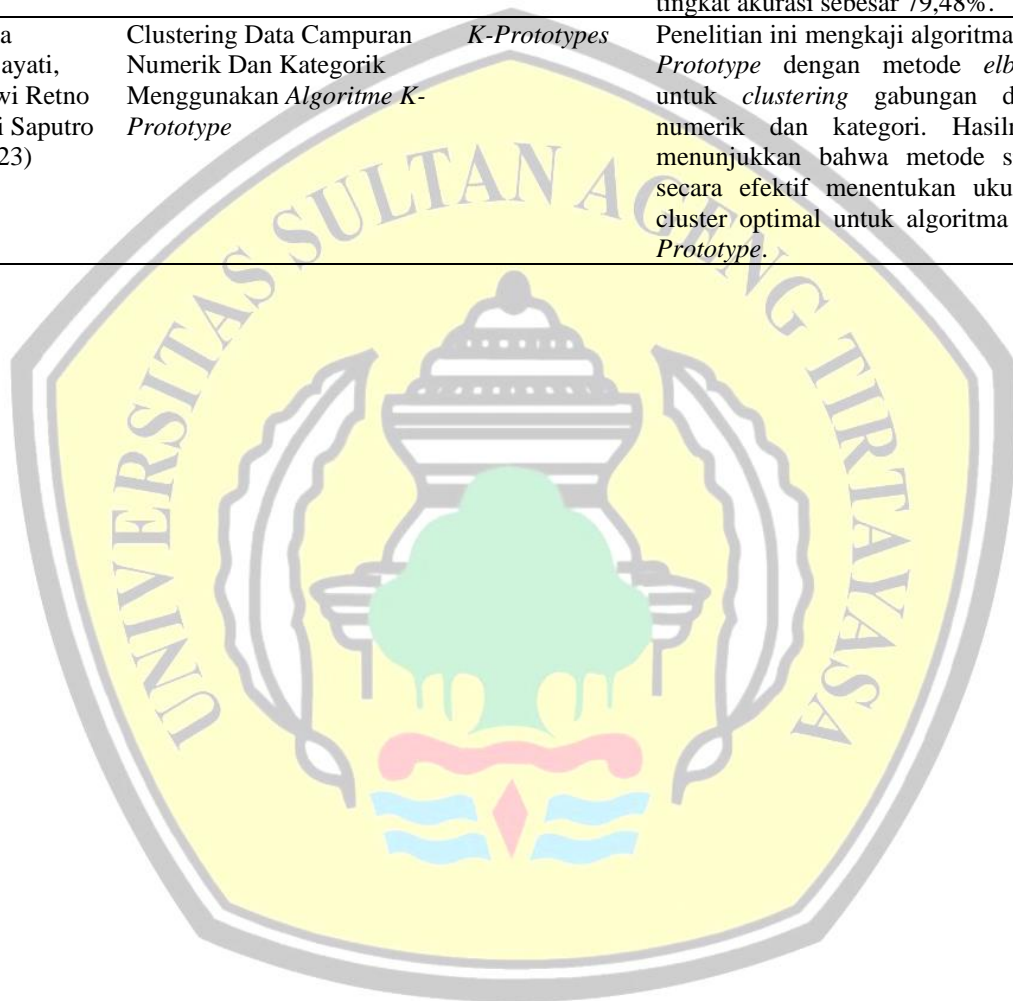
Bab ini memberikan rekomendasi untuk penelitian masa depan dan mengungkapkan kesimpulan penelitian.

1.6 Penelitian Terdahulu

Adapun penelitian terdahulu dengan topik yang memiliki kesamaan pada penelitian ini adalah sebagai berikut:

No	Nama Peneliti	Judul Penelitian	Metode	Hasil Penelitian
1.	Josi Aranda, W birda Astari Galvani Natasya (2016)	Penerapan Metode <i>K-Means Cluster Analysis</i> Pada Sistem Pendukung Keputusan Pemilihan Konsentrasi Untuk Mahasiswa <i>International Class</i> STMIK Amikom Yogyakarta	<i>K-Means</i>	Pada penelitian ini dilakukan dengan metode <i>K-Means Cluster Analysis</i> untuk mengetahui kecenderungan mahasiswa memilih konsentrasi berdasarkan nilai mata kuliah yang ada. Setelah dilakukan pengolahan data dan tiga iterasi <i>clustering</i> , hasil penelitian ini menunjukkan bahwa empat dari dua belas mahasiswa diarahkan ke konsentrasi Pemrograman, empat mahasiswa ke konsentrasi Multimedia, dan tiga sampai lima mahasiswa ke konsentrasi Jaringan Komputer. Hasil ini dipengaruhi oleh nilai awal <i>centroid</i> yang digunakan dan jumlah data yang digunakan. Jika ada dua pilihan, disarankan untuk memilih konsentrasi Multimedia. pembuat pupuk cair organik dengan tepat dan akurat.
2.	Sri Sulastri, Lismayani Usman, Utami Dyah Syafitri (2021)	<i>K-Prototypes Algorithm for Clustering Schools Based on The Student Admission Data in IPB University</i>	<i>K-Prototypes</i>	Kajian ini bertujuan untuk mengkaji data penerimaan mahasiswa di IPB <i>University</i> dan lembaga kluster berdasarkan efektifitas dalam memberangkatkan lulusan IPB. Algoritma <i>K-Prototype</i> diterapkan pada variabel kategori dan numerik yang merupakan tipe data campuran. Berdasarkan hasil tersebut, terdapat empat kluster optimal. Kluster keempat yang terdiri dari 421 sekolah dinilai terbaik dalam penerimaan mahasiswa ke IPB <i>University</i> , sedangkan kluster ketiga yang terdiri dari 391 sekolah dinilai paling buruk. Kluster keempat memiliki lulusan terbanyak yang efektif matrikulasi dan lulus dari IPB <i>University</i> , sedangkan kluster ketiga paling sedikit. Mengenai penerimaan mahasiswa di IPB <i>University</i> , kluster keempat menunjukkan kinerja tertinggi, sedangkan kluster ketiga menunjukkan kinerja terendah.
3.	Millah Azkiyyah (2017)	Pengelompokan Kabupaten/Kota Di Nusa Tenggara Timur Berdasarkan Pelayanan Kesehatan Ibu Dan Anak Menggunakan <i>K-Means</i> Dan <i>Fuzzy C-Means Cluster</i>	<i>K-Means</i> dan <i>Fuzzy C-Means Cluster</i>	Penelitian ini mengkaji pelaporan pelayanan KIA di NTT, metode klasifikasi yang digunakan, dan karakteristik kategori berdasarkan pelayanan KIA..

No	Nama Peneliti	Judul Penelitian	Metode	Hasil Penelitian
4.	Viryandra Virtusena, Asahar Johar, Andang Wijanarko (2021)	Pengelompokan Potensi Kelulusan Mahasiswa Fakultas Teknik Unib Menggunakan <i>Algoritme K-Means</i> (Studi Kasus: Fakultas Teknik Universitas Bengkulu	<i>K-Means</i>	Penelitian ini berhasil mengimplementasikan algoritma K-Means untuk mengklasifikasikan mahasiswa menjadi dua kelompok yaitu yang lulus tepat waktu dan yang tidak lulus. Sistem pengklasifikasian mahasiswa menurut peluang kelulusannya menggunakan website dengan tingkat akurasi sebesar 79,48%.
5.	Rika Wijayati, Dewi Retno Sari Saputro (2023)	Clustering Data Campuran Numerik Dan Kategorik Menggunakan <i>Algoritme K-Prototype</i>	<i>K-Prototypes</i>	Penelitian ini mengkaji algoritma <i>K-Prototype</i> dengan metode <i>elbow</i> untuk <i>clustering</i> gabungan data numerik dan kategori. Hasilnya menunjukkan bahwa metode siku secara efektif menentukan ukuran cluster optimal untuk algoritma <i>K-Prototype</i> .



BAB II

TINJAUAN PUSTAKA

2.1 *Data Mining*

Data mining memiliki definisi sederhana sebagai kegiatan dalam melakukan ekstraksi informasi atau pola penting atau menarik dari data dalam *database* besar. *Data mining* adalah penemuan informasi baru dengan mencari pola atau aturan dari sejumlah besar data. Hasil pengolahan data dengan metode *data mining* dapat digunakan untuk mengambil keputusan di kemudian hari. Berdasarkan penjelasan beberapa penelitian sebelumnya, hasil dari proses *data mining* harus berupa informasi baru dan berguna untuk kebutuhan masa depan dan mudah dipahami (Qadrini, 2018). Berikut merupakan tahap-tahapan *data mining* (Sudarsono, et al., 2021):

1. Pembersihan data

Tahap pembersihan data dilakukan pembersihan agar data-data yang tidak diperlukan lebih baik dihilangkan karena keberadaan data tersebut dapat mengurangi kualitas atau akurasi dari hasil *data mining*.

2. Integrasi data

Tahap integrasi data dilakukan pada atribut-atribut yang diidentifikasi sebagai entitas-entitas unik seperti nama, nomor identitas diri dan lain sebagainya. Tahap integrasi data membantu menggabungkan sebuah data berdasarkan entitas-entitas unik.

3. Transformasi data

Tahap transformasi data dilakukan untuk merubah format data menjadi format khusus sebelum dapat diaplikasikan sehingga didapatkan kualitas dari hasil *data mining*.

4. Aplikasi teknik *data mining*

Tahap aplikasi teknik *data mining* merupakan salah satu bagian dari alur proses *data mining* yang dilakukan untuk membantu melaksanakan *data mining* di bidang tertentu atau untuk data tertentu.

5. Evaluasi pola

Tahap evaluasi pola dilakukan untuk melihat hasil dari teknik *data mining* dalam menemukan pola-pola yang unik.

6. Presentasi pola

Tahap presentasi pola dilakukan untuk memformulasikan keputusan dari hasil analisis yang didapat.

2.2 Statistika Deskriptif

Metode mengenai pengumpulan dan penyajian kumpulan data untuk memberikan informasi yang sesuai dengan kebutuhan, yang dikenal sebagai statistik deskriptif. Pada metode ini menggunakan ukuran pemusatan data berupa rata-rata untuk mengetahui karakteristik setiap variabel. Statistika deskriptif bertujuan untuk menemukan karakteristik yang terkait pada setiap variabel (Azkiyah, 2017). Berikut adalah rumus dari rata-rata.

$$\bar{X} = \frac{\sum_{i=1}^n X}{n} \dots\dots\dots(1)$$

Keterangan :

- \bar{X} = Rata-rata,
- X = Jumlah nilai,
- n = Banyaknya data.

Standar deviasi merupakan akar dari varians yang digunakan untuk mengevaluasi ukuran keseragaman data setiap variabel. Varians adalah nilai yang menunjukkan tingkat variasi dalam kumpulan data dengan keterangan yang sama dengan rata-rata (Azkiyah, 2017). Berikut rumus standar deviasi adalah sebagai berikut:

$$S = \sqrt{S^2} = \sqrt{\frac{\sum_{i=1}^n (X - \bar{X})^2}{n-1}} \dots\dots\dots(2)$$

Keterangan :

- \bar{X} = Rata-rata,

- X = Jumlah nilai,
 n = Banyaknya data,
 S = Standar deviasi.

2.3 Klasterisasi

Klasterisasi (*clustering*) adalah metode yang mengklasifikasikan objek atau data ke dalam kategori yang mirip berdasarkan tingkat kesamaan atau jarak di antara atribut. Tujuan pengelompokan adalah untuk mengelompokkan objek yang mirip ke dalam satu kelompok dan memisahkan objek yang berbeda ke dalam kelompok yang terpisah. Klasterisasi akan mencari pola atau struktur pada data untuk membentuk kelompok dengan karakteristik yang mirip. Atribut ini mungkin atribut numerik atau kategori dari objek yang dianalisis. Klasterisasi pada umumnya menggunakan perhitungan seperti jarak *euclidean* dan lain sebagainya yang digunakan dengan metode pengelompokan untuk mengukur kesamaan antar objek. Klasterisasi berlaku untuk berbagai disiplin ilmu, termasuk *data mining*, analisis, pengenalan pola, dan sistem rekomendasi. Klasterisasi bertujuan utama pengelompokan adalah untuk mengungkap struktur tersirat dalam data, mengidentifikasi pengelompokan yang saling terkait, dan mendapatkan pemahaman dan pendekatan analisis data yang lebih terstruktur dan terorganisir (Tan, et al., 2005).

2.4 Teorema Limit Pusat

Teori teorema limit pusat atau *teorema limit central* adalah salah satu bagian statistika yang paling penting karena mempunyai jangkauan terjauh baik dari segi teori maupun penerapannya dan merupakan sebuah kontribusi modern yang besar, tidak hanya pada statistik tetapi juga pada semua bidang matematika. Teori ini memiliki asumsi-asumsinya langsung bersifat umum, sehingga mudah diterapkan. Teorema limit pusat merupakan teori yang menyatakan bahwa jika ukuran sampel bertambah maka sifat-sifat distribusi *mean* sampel akan semakin mendekati distribusi normal. Hasil pada sampel ini juga mempunyai ciri-ciri yang juga terdapat pada distribusi normal. Semakin banyak data yang diambil maka semakin mendekati distribusi normal (Andini & Trianasari, 2021).

2.5 *One-Way Analyze of Variance (One-Way ANOVA)*

One-Way Analysis of Variance (ANOVA) adalah salah jenis dari metode ANOVA yang dimana alat uji ini digunakan untuk menguji perbedaan mean (rata-rata) data lebih dari dua kelompok. Uji *one-way* ANOVA menguji kemampuan dari signifikansi hasil penelitian. Uji ini jika terbukti berbeda dua atau lebih sampel tersebut dianggap dapat mewakili populasi. Uji *one-way* ANOVA dilakukan bertujuan dan pengujian *one-way* ANOVA adalah untuk mengetahui apakah terdapat perbedaan antara berbagai kriteria dengan hasil yang diinginkan. Syarat dalam melakukan uji *one-way* ANOVA pada saat pengambilan sampel yang dilakukan secara *random* terhadap beberapa (> 2) kelompok yang independen, di mana nilai pada satu kelompok tidak tergantung pada nilai di kelompok lain. Uji *one-way* ANOVA harus memenuhi beberapa asumsi sebagai berikut (Palupi & Prasetya, 2022).

1. Sampel terdiri dari kelompok yang independen.
2. Varian antar kelompok harus homogen.
3. Data masing-masing kelompok berdistribusi normal.

2.6 *Algoritma K-Prototype*

K-Prototype adalah algoritma dasar untuk melakukan pengelompokan gabungan data numerik dan kategori. Algoritma ini menggabungkan algoritma *K-Means* dengan algoritma *K-Modes*. Algoritma *K-Prototype* adalah teknik pengelompokan berdasarkan partisi. Algoritma ini berasal dari algoritma *K-Means* untuk mengelola kluster data dengan tipe atribut numerik dan atribut kategori. Metode ini mempertahankan efektivitas *K-Means* dalam menangani kumpulan data besar dan dapat diterapkan pada data numerik dan kategorikal. Inovasi inti algoritma *K-Prototype* adalah ukuran kesamaan antara objek dan *centroid*-nya (*prototype*) (Subhan, et al., 2022). Berikut tahap-tahap dalam menggunakan algoritma *K-Prototype* (Qadrini, 2018).

1. Tahap 1
Menentukan nilai k dengan inisial kluster z_1, z_2, \dots, z_k secara acak (*random*) dari n ke sebuah titik $\{x_1, x_2, \dots, x_n\}$.
2. Tahap 2

Melakukan perhitungan jarak pada seluruh data *point* pada dataset terhadap kluster awal yang telah ditentukan dengan persamaan $d(X_j, Z_i = \pi r^2 = (\sum_{l=1}^{m_r} (x_{jl}^r - z_{il}^r)^2)^{\frac{1}{2}}$, selanjutnya melakukan alokasi titik data ke dalam kluster yang memiliki jarak terdekat dengan objek yang diukur (*prototypes*).

3. Tahap 3

Melakukan perhitungan untuk titik pusat kluster yang baru setelah semua objek dialokasikan, selanjutnya melakukan relokasi pada semua titik data pada dataset terhadap *prototype* yang baru.

4. Tahap 4

Melakukan pemeriksaan ulang jika titik pusat kluster tidak berubah atau telah konvergen menandakan bahwa proses algoritma telah selesai tetapi jika titik pusat dari kluster masih mengalami perubahan secara signifikan menandakan bahwa proses akan kembali pada tahap 2 dan 3 yang bertujuan untuk mendapatkan iterasi maksimum atau objek tidak mengalami perpindahan.

2.7 *Preprocessing Data Mining*

Pengolahan *data mining* memerlukan sebuah tahap yaitu *preprocessing data mining* untuk menemukan data yang memiliki nilai yang hilang (*missing value*), distorsi nilai, tidak tersimpannya nilai (*misrecording*), *sampling* yang tidak cukup bagus dan sebagainya. Permasalahan tersebut memerlukan sebuah solusi untuk meningkatkan kualitas pengolahan data dengan melakukan sebuah penyiapan data (*preprocessing*). Hal yang menjadi penyebab dari kurang baiknya kualitas dari sebuah data mentah adalah dikarenakan sebuah kesalahan dalam penyimpanan dan pengukuran, tetapi penyebab lain adalah karena tidak adanya nilai yang mewakili nilai yang tersedia (Qadrini, 2018).

2.8 *Evaluasi Pengelompokan*

Parameter dalam mengetahui kelayakan dari hasil pengelompokan dilakukan evaluasi sebagai berikut.

2.8.1 Uji Validitas Kelompok

Permasalahan utama dalam analisis kelompok adalah menentukan jumlah kelompok oleh peneliti karena belum memiliki dasar yang kuat mengenai jumlah kelompok terbaik. Uji validitas kelompok dilakukan dalam menyelesaikan permasalahan sebelumnya perlu dilakukan uji validitas kelompok dalam menguji hasil dari analisis kelompok secara kuantitatif sehingga dapat dihasilkan kelompok yang optimal (Qadrini, 2018). Kelompok optimal adalah sebuah kelompok yang memiliki jarak yang padat antar individu dalam kelompok dan terisolasi dari kelompok lainnya dengan baik (Jain & Dubes, 1988). Adapun cara dalam menentukan validitas suatu data diukur dengan ukuran internal dan ukuran eksternal (Steinbach, et al., 2000).

2.8.1.1 Ukuran Internal

Ukuran internal digunakan dalam mengukur kelompok data yang terbentuk tanpa pertimbangan informasi dari luar (Steinbach, et al., 2000).

1. Indeks *Global Silhoutte*

Adapun mendapatkan indeks kelompok $S(i)$ adalah berikut

$$S(i) = \frac{b(i)-a(i)}{\max(a(i),b(i))} \dots\dots\dots(4)$$

Keterangan :

$a(i)$ = Rata-rata perbedaan dari i -objek dengan semua objek lain pada kelompok yang sama,

$b(i)$ = Objek pada kelompok lain (di kelompok terdekat).

Nilai yang paling besar dari indeks *global silhouette* atau rata-rata koefisien *silhouette* menentukan jumlah kelompok terbaik yang akan diambil menjadi kelompok optimal.

Adapun rumus *global silhouette* adalah sebagai berikut.

$$GS_n = \frac{1}{n} \sum_{i=1}^n S(i) \dots\dots\dots(5)$$

Keterangan :

GS_n = Indeks *global silhouette*,

$S(i)$ = *Silhoutte* kelompok ke- i .

2. Indeks *Dunn*

Adapun indeks *dunn* dirumuskan sebagai berikut.

$$D = \min_{1 \leq l \leq n} \left\{ \min_{1 \leq j \leq n, i \neq j} \left\{ \frac{d(c_i, c_j)}{\max_{1 \leq k \leq n} (d'(c_k))} \right\} \right\} \dots \dots \dots (6)$$

Keterangan :

D = Indeks validasi *dunn*,

$d'(c_k)$ = Jarak dalam kelompok (c_k).

Nilai terbesar dari D dipilih menjadi jumlah kelompok optimal (Bolshakova & Azuaje, 2003).

2.8.1.2 Ukuran Eksternal

Ukuran eksternal digunakan dalam mengukur tingkat kecocokan kelompok yang telah terbentuk dengan informasi kelas data. Dalam ukuran validasi eksternal terdapat ukuran kualitas eksternal adalah *entropy* (Steinbach, et al., 2000). Dalam menentukan *entropy* melakukan perhitungan kelas distribusi untuk setiap kelompok. Selanjutnya hitung p_{ik} peluang bahwa kelompok k memuat anggota kelas ke- i , dengan $p_{ik} = \frac{n_{ik}}{n_k}$. Adapun nilai *entropy* setiap kelompok adalah sebagai berikut (Qadrini, 2018).

$$E_k = - \sum_i p_{ik} \log(p_{ik}) \dots \dots \dots (7)$$

Selanjutnya jumlah total *entropy* dinyatakan dengan persamaan sebagai berikut.

$$E = \sum_{k=1}^g \frac{n_k \times E_k}{n} \dots \dots \dots (8)$$

Keterangan :

E = *Entropy*,

N_{ik} = Banyaknya anggota pada kelas ke- i yang terdapat pada kelompok k ,

N_k = Banyaknya anggota kelompok ke- k ,

g = Banyaknya kelompok.

2.9 General Linear Model (GLM)

General Linear Model (GLM) adalah model statistik yang dapat digunakan untuk menganalisis data yang berskala interval atau rasio. GLM merupakan metode

hasil perkembangan dari model linear dengan asumsi prediksi yang memiliki pengaruh linier tetapi tidak melakukan asumsi untuk distribusi tertentu dari variabel *respond* (Caraka, et al., 2018). GLM merupakan metode yang didukung dengan perhitungan *Root Mean Square Error* (RMSE) yang memiliki fungsi untuk mendapatkan pendekatan berdasarkan tingkat kesalahan hasil perkiraan, semakin kecil (mendekati 0) nilai RMSE menunjukkan bahwa semakin akurat nilai perkiraannya (Hamdanah & Fitriana, 2021).



BAB III

METODE PENELITIAN

3.1 Rancangan Penelitian

Penelitian ini menggunakan jenis penelitian kuantitatif. Metode kuantitatif adalah metode yang berhubungan dengan angka atau nominal yang digunakan pada penelitian dalam pengambilan data (Waruwu, 2023). Data yang digunakan berdasarkan data-data yang diperoleh dari basis data Universitas Sultan Ageng Tirtayasa dengan mengumpulkan berbagai data numerik dan kategorik untuk mengetahui variabel-variabel yang signifikan yang kemudian dianalisis untuk menentukan pengelompokan data.

3.2 Lokasi Penelitian dan Waktu Penelitian

Penelitian ini dilakukan di Universitas Sultan Ageng Tirtayasa. Jl. Jenderal Sudirman Km 3, Kotabumi, Kec. Purwakarta, Kota Cilegon, Banten. Penelitian ini dilakukan dari bulan Mei 2023 sampai Agustus 2023.

3.3 Cara Pengambilan Data

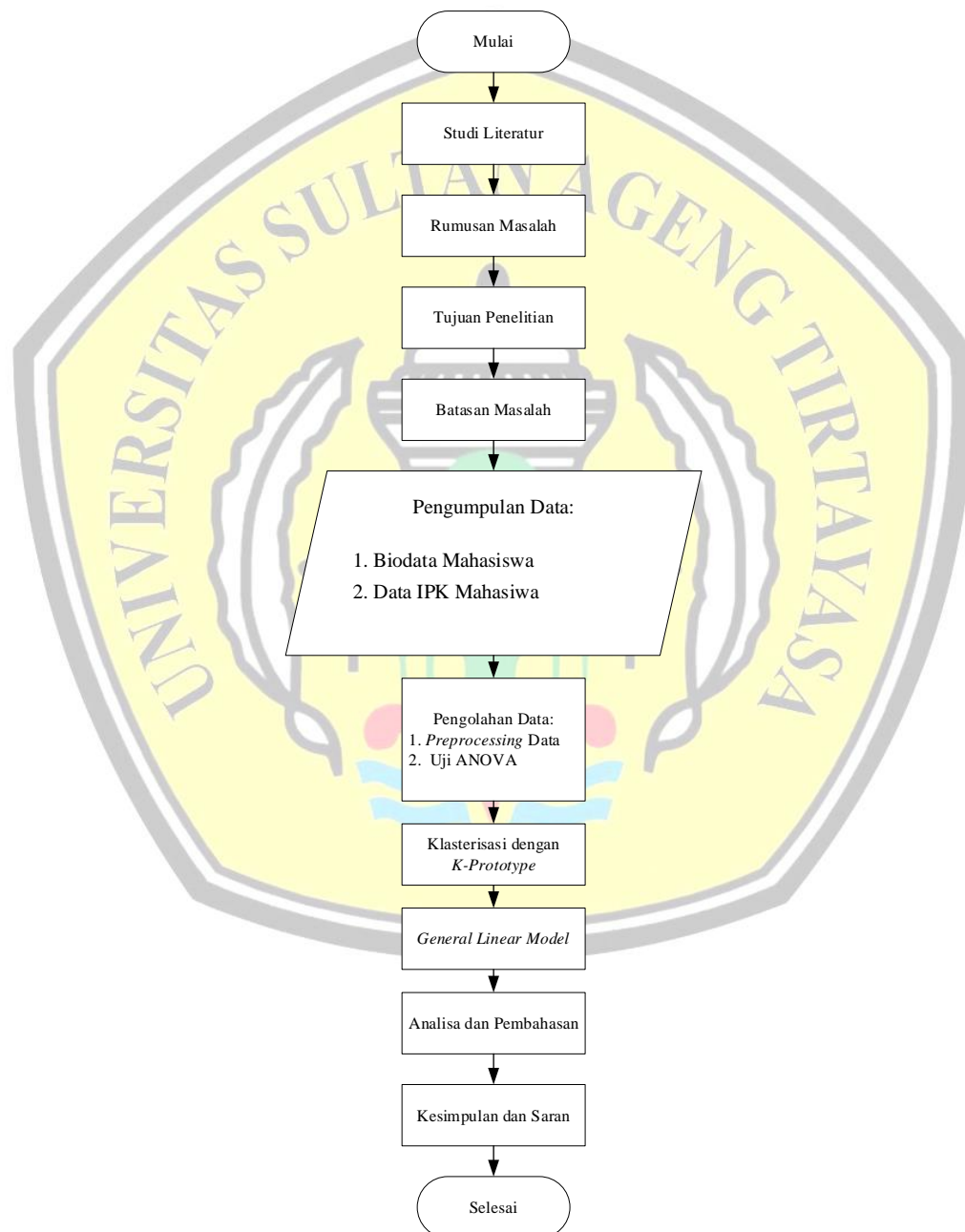
Penelitian ini menggunakan data sekunder yang dimana pengambilan data dilakukan dengan mengunduh data-data dari basis data Universitas Sultan Ageng Tirtayasa yang memiliki kesesuaian dan keterkaitan pada topik penelitian ini. Data yang diambil merupakan data-data dari mahasiswa Universitas Sultan Ageng Tirtayasa, dari data biodata mahasiswa dan data IPK mahasiswa angkatan 2021/2022 yang meliputi variabel-variabel seperti, Nomor Induk Mahasiswa (NIM), nama mahasiswa, kode prodi, angkatan, status mahasiswa, Status Kredit Semester (SKS) wajib, SKS pilihan, jumlah SKS, IPK, jenis kelamin, periode masuk, jalur masuk, status akademik, status mahasiswa, agama, kota tinggal, kecamatan, kelurahan, kategori provinsi, kategori pulau, pekerjaan ayah, pekerjaan ibu, tahun lulus SMTA, jurusan SMTA, asal SMTA, kWh listrik, bidikmisi, dan luas tanah.

3.4 Alur Pemecahan Masalah

Penelitian ini terdapat alur proses-proses yang bertujuan untuk mendapatkan hasil.

3.4.1 *Flowchart* Pemecahan Masalah

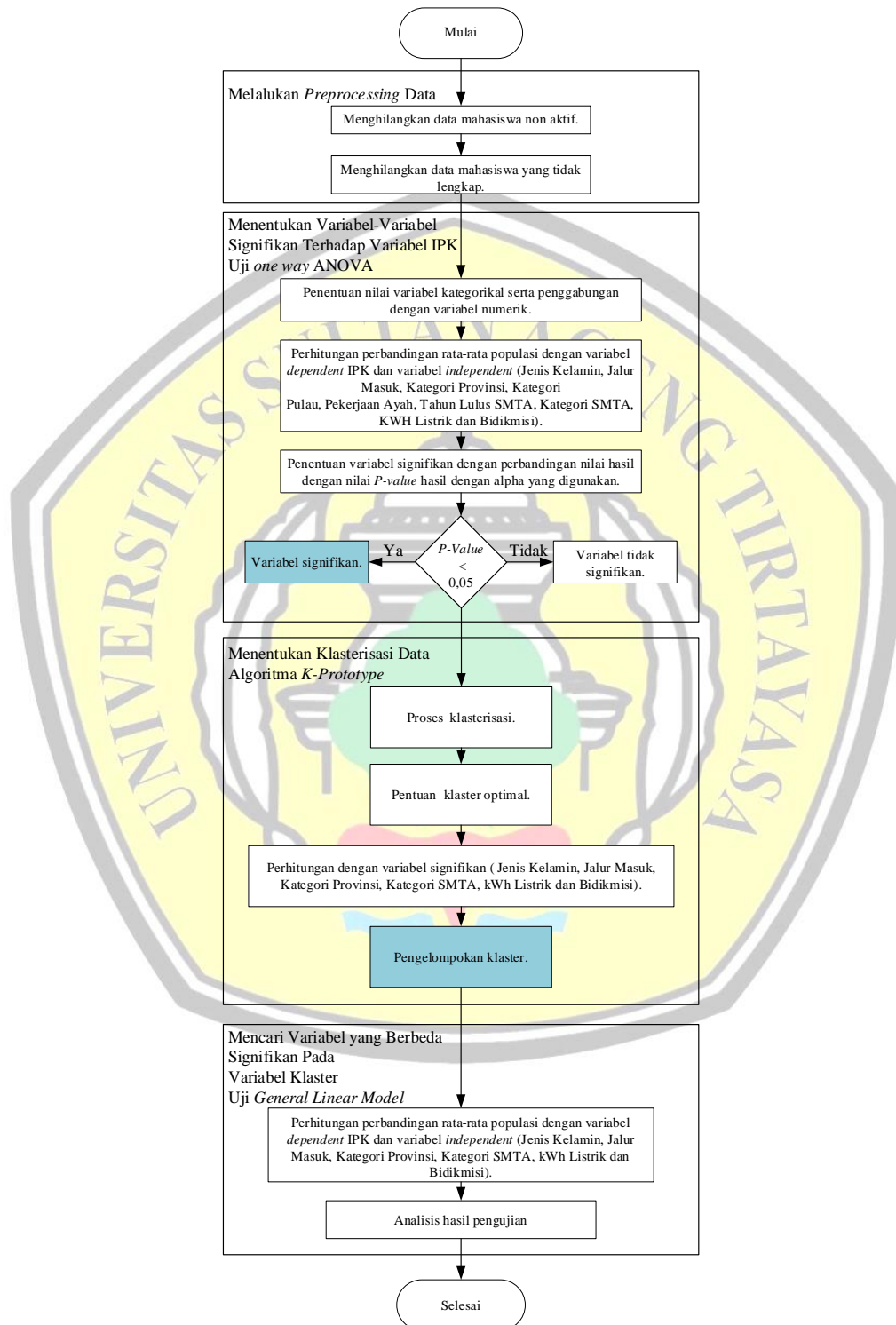
Gambar 1 merupakan *flowchart* alur pemecahan masalah yang dalam penelitian ini:



Gambar 1. *Flowchart* Pemecahan Masalah

3.4.2 Flowchart Pengolahan Data

Gambar 2 merupakan *flowchart* alur pengolahan data yang dalam penelitian ini:



Gambar 2. Flowchart Pengolahan Data

3.5 Deskripsi Pemecahan Masalah

Adapun pada penelitian ini terdapat deskripsi dari *flowchart* Gambar 1 dan Gambar 2 untuk menjelaskan alur pemecahan masalah dan pengolahan data.

3.5.1 Deskripsi *Flowchart* Pemecahan Masalah

Berikut ini merupakan deskripsi *flowchart* pemecahan masalah pada penelitian kali ini

1. Mulai
Melakukan identifikasi masalah berdasarkan fenomena di Universitas Sultan Ageng Tirtayasa
2. Studi Literatur
Menelaah studi literatur untuk mendukung penyelesaian masalah pada penelitian ini.
3. Rumusan Masalah
Menentukan rumusan masalah yang memiliki urgensi berdasarkan fenomena yang terjadi untuk diselesaikan
4. Tujuan Penelitian
Menentukan hasil yang ingin didapatkan dari penyelesaian masalah yang terjadi.
5. Batasan Masalah
Menentukan ruang lingkup masalah yang akan dibahas pada penelitian ini.
6. Pengumpulan Data
Melakukan pengumpulan data berkaitan data-data mahasiswa Universitas Sultan Ageng Tirtayasa Angkatan 2021/2022.
7. Penentuan Variabel
Melakukan pemilihan variabel yang diasumsikan mewakili karakteristik mahasiswa.
8. Pengolahan Data
Melakukan pembersihan data terhadap data-data yang telah dikumpulkan dengan menggunakan *preprocessing* data kemudian dilanjutkan dengan melakukan perhitungan dengan metode ANOVA.

9. Klasterisasi dengan *K-Prototype*

Menentukan klaster optimal sehingga mendapatkan variabel klaster.

10. Perhitungan GLM

Melakukan perhitungan GLM untuk mengetahui variabel apa saja yang signifikan terhadap klaster.

11. Analisis dan Pembahasan

Melakukan analisis mengenai data yang telah diolah dengan melihat permasalahan yang terjadi serta membahas langkah pemecahan masalah yang ada.

12. Kesimpulan dan Saran

Melakukan penarikan kesimpulan pada penelitian ini sesuai dengan tujuan yang diharapkan kemudian memberikan saran-saran dari solusi masalah yang ada.

13. Selesai

Penelitian selesai dilakukan

3.5.2 Deskripsi *Flowchart* Pengolahan Data

Berikut merupakan deskripsi *flowchart* pengolahan data pada Gambar 2 penelitian.

1. Mulai

Memulai proses tahap awal dari pengolahan data.

2. Melakukan *preprocessing* data

Menghilangkan data mahasiswa non aktif selanjutnya melakukan pembersihan data untuk data yang tidak lengkap.

3. Menentukan variabel-variabel signifikan

Hasil dari uji ANOVA digunakan untuk mencari variabel yang signifikan. Selanjutnya variabel yang signifikan dapat dilanjutkan ke tahap berikutnya.

4. Penentuan nilai variabel kategorikal serta penggabungan dengan variabel numerik

Melakukan konversi data pada variabel kategorik agar dapat diolah pada uji ANOVA kemudian data tersebut digabungkan dengan data variabel numerik.

5. Perhitungan perbandingan rata-rata populasi dengan variabel *dependent* IPK

Melakukan perhitungan uji ANOVA yang didasarkan pada data variabel IPK sebagai variabel *dependent*-nya dengan data lainnya sebagai variabel *independent*-nya.

6. Penentuan variabel signifikan dengan perbandingan nilai hasil dengan nilai *p-value* hasil dengan alpha yang digunakan

Melakukan penentuan untuk mencari tahu apa saja variabel yang sangat berpengaruh pada penelitian ini dengan membandingkan nilai *p-value* hasil dengan alpha.

7. Variabel signifikan

Menentukan variabel yang signifikan jika nilai *p-value* kurang dari nilai alpha yang digunakan. Jika didapatkan bahwa nilai dari *p-value* lebih kecil maka variabel merupakan variabel yang berpengaruh dan dapat dilanjutkan ke penelitian selanjutnya.

8. Variabel tidak signifikan

Menentukan variabel yang tidak signifikan jika nilai *p-value* lebih dari nilai alpha yang digunakan, jika didapatkan bahwa nilai dari *p-value* lebih besar maka variabel merupakan variabel yang tidak berpengaruh dan tidak dapat dilanjutkan ke penelitian selanjutnya.

9. Menentukan klasterisasi data

Hasil dari uji dengan menggunakan algoritma *K-Prototype* didapatkan klasterisasi data.

10. Proses Klasterisasi

Melakukan proses klasterisasi untuk mengetahui alternatif jumlah klaster.

11. Penentuan klaster optimal

Menentukan klaster yang optimal berdasarkan data yang ada untuk melihat penyebaran data terbaik.

12. Perhitungan dengan variabel signifikan

Melakukan perhitungan *K-Prototype* berdasarkan variabel yang signifikan.

13. Variabel kluster

Hasil akhir dari penelitian ini yaitu variabel mahasiswa Universitas Sultan Ageng Tirtayasa berdasarkan data kemiripan atributnya.

14. Perhitungan perbandingan rata-rata populasi dengan variabel *dependent* kluster dan variabel *independent*.

Melakukan perhitungan dengan uji GLM untuk mengetahui variabel signifikan terhadap kluster.

15. Selesai

Penelitian telah selesai dilakukan

3.6 Analisis Data

Dalam penelitian dilakukannya analisis data untuk menentukan faktor apa saja yang memiliki pengaruh yang sangat penting menggunakan uji ANOVA. Uji ANOVA bertujuan mengetahui karakteristik yang memiliki pengaruh signifikan terhadap IPK mahasiswa dengan signifikan (α) 0.05. Apabila $F_{hitung} < F_{tabel}$ atau $p-value > \alpha$ maka menunjukkan tidak ada perbedaan sedangkan jika $F_{hitung} > F_{tabel}$ atau $p-value < \alpha$ maka menunjukkan adanya perbedaan, sehingga hipotesis awal (H_0) ditolak sedangkan jika $F_{hitung} < F_{tabel}$ atau $p-value > \alpha$ maka menunjukkan tidak ada perbedaan, sehingga awal (H_0) diterima. Adapun hipotesis pada penelitian ini adalah sebagai berikut.

1. Jenis Kelamin

H_0 = Perbedaan jenis kelamin tidak berpengaruh pada nilai IPK mahasiswa.

H_1 = Perbedaan jenis kelamin berpengaruh pada nilai IPK mahasiswa.

2. Jalur Masuk

H_0 = Perbedaan jalur masuk tidak berpengaruh pada nilai IPK mahasiswa.

H_1 = Perbedaan jalur masuk berpengaruh pada nilai IPK mahasiswa.

3. Kategori Provinsi

H_0 = Perbedaan kategori provinsi tidak berpengaruh pada nilai IPK mahasiswa.

H_1 = Perbedaan kategori provinsi berpengaruh pada nilai IPK mahasiswa.

4. Kategori Pulau

H_0 = Perbedaan kategori pulau tidak berpengaruh pada nilai IPK mahasiswa.

H_1 = Perbedaan kategori pulau berpengaruh pada nilai IPK mahasiswa.

5. Pekerjaan Ayah

H_0 = Perbedaan pekerjaan ayah tidak berpengaruh pada nilai IPK mahasiswa.

H_1 = Perbedaan pekerjaan ayah berpengaruh pada nilai IPK mahasiswa.

6. Tahun Lulus SMTA

H_0 = Perbedaan tahun lulus SMTA tidak berpengaruh pada nilai IPK mahasiswa.

H_1 = Perbedaan tahun lulus SMTA berpengaruh pada nilai IPK mahasiswa.

7. Kategori SMTA

H_0 = Perbedaan kategori SMTA tidak berpengaruh pada nilai IPK mahasiswa.

H_1 = Perbedaan kategori SMTA berpengaruh pada nilai IPK mahasiswa.

8. kWh Listrik

H_0 = Perbedaan kWh listrik tidak berpengaruh pada nilai IPK mahasiswa.

H_1 = Perbedaan kWh listrik berpengaruh pada nilai IPK mahasiswa.

9. Bidikmisi

H_0 = Perbedaan bidikmisi tidak berpengaruh pada nilai IPK mahasiswa.

H_1 = Perbedaan bidikmisi berpengaruh pada nilai IPK mahasiswa.

Hasil dari variabel signifikan tersebut diuji kembali untuk mengetahui klusterisasi mahasiswa Universitas Sultan Ageng Tirtayasa Angkatan 2021/2022 dengan menggunakan uji algoritma *K-Prototype*.

BAB IV

HASIL PENELITIAN

4.1 Pengumpulan Data

Pengumpulan data pada penelitian ini mencakup data IPK mahasiswa UNTIRTA angkatan 2021/2022 dan data biodata mahasiswa UNTIRTA. Data-data tersebut adalah variabel identitas, variabel numerik, dan variabel kategori.

4.1.1 Deskripsi Karakteristik Variabel

Analisis deskriptif digunakan terhadap setiap jenis variabel yang digunakan pada penelitian ini dengan menggunakan 3 (tiga) karakteristik variabel yaitu variabel identitas, variabel numerik, dan variabel kategori. Pada Tabel 1 menampilkan ringkasan data yang dikumpulkan pada penelitian ini sesuai dengan Lampiran 1.

Tabel 1. Ringkasan Data

No	NIM	Jenis Kelamin	Tahun Lulus SMTA	Kategori SMTA	kWh Listrik	Bidikmisi	IPK
1	1111210001	P	2021	SMAN	1300	Tidak	3,98
2	1111210002	P	2021	SMAN	900	Tidak	3,76
3	1111210003	P	2021	SMAS	2200	Tidak	3,92
4	1111210005	P	2021	MAN	450	Ya	3,92
.....
4096	8884210039	P	2021	SMAN	450	Tidak	2,92

Berdasarkan data pada Tabel 1 menunjukkan data yang bersifat rahasia sehingga tidak ditampilkan untuk keseluruhan data. Variabel-variabel yang terdapat pada ringkasan data seperti NIM, jenis kelamin, tahun lulus SMTA, kWh listrik, bidikmisi, dan IPK.

4.1.1.1 Deskripsi Karakteristik Variabel Identitas

Variabel identitas pada penelitian ini menggunakan variabel NIM. Variabel NIM digunakan sebagai penanda identitas dari setiap mahasiswa UNTIRTA yang dimana variabel NIM yang dimiliki setiap mahasiswa akan berbeda-beda. Variabel

NIM digunakan untuk memberikan informasi seperti nomor fakultas yang terdiri dari 3 (tiga) digit angka pertama, nomor prodi yang terdiri dari 1 (satu) digit angka selanjutnya, tahun masuk yang terdiri dari 2 (dua) digit angka selanjutnya, dan nomor mahasiswa yang terdiri dari 4 (empat) angka terakhir.

4.1.1.2 Deskripsi Karakteristik Variabel Numerik

Deskripsi karakteristik variabel numerik terdapat pada Tabel 2 yang menunjukkan variabel IPK. Nilai-nilai pada variabel IPK antara lain seperti nilai minimal adalah 0,28 dan nilai maksimal adalah 3,99. Selanjutnya untuk nilai media adalah 3,69 dan standar deviasi adalah 0,460 serta terakhir untuk nilai rata-rata pada variabel IPK memiliki nilai 3,555.

Tabel 2. Deskripsi Karakteristik Data Numerik

Variabel	Minimal	Maksimal	Median	Standar Deviasi	Rata-rata	Jumlah
IPK	0,280	3,990	3,690	0,460	3,555	4096

Contoh Perhitung:

- Minimal (nilai terkecil)

$$\text{Minimal} = 2,28$$

- Maksimal (nilai terbesar)

$$\text{Maksimal} = 3,990$$

- Median (Me)

$$\text{Me} = \frac{X_n + X_{(\frac{n}{2}+1)}}{2}$$

$$\text{Me} = \frac{X_{\frac{4096}{2}} + X_{(\frac{4096}{2}+1)}}{2}$$

$$\text{Me} = \frac{X_{2048} + X_{2049}}{2}$$

$$\text{Me} = \frac{3,690 + 3,690}{2}$$

$$\text{Me} = 3,690$$

- Standar deviasi (S)

$$S = \sqrt{\frac{\sum(X_i - \bar{X})^2}{n}}$$

$$S = \sqrt{\frac{(3,980 - 3,555)^2 + (3,760 - 3,555)^2 + \dots + (2,920 - 3,555)^2}{4096}}$$

$$S = 0,460$$

5. Rata-rata (\bar{X})

$$\bar{X} = \frac{X_1 + X_2 + X_3 + \dots + X_{4096}}{4096}$$

$$\bar{X} = \frac{3,980 + 3,760 + 3,920 + \dots + 2,920}{4096}$$

$$\bar{X} = 3,555$$

6. Jumlah

$$\text{Jumlah} = 4096$$

4.1.1.3 Deskripsi Karakteristik Variabel Kategori

Deskripsi karakteristik variabel kategori ditunjukkan pada Tabel 3. Deskripsi karakteristik variabel kategori mencakup karakteristik variabel kategori yang berjumlah 9 (sembilan) yaitu, variabel jenis kelamin, jalur masuk, kategori provinsi, kategori pulau, pekerjaan ayah, tahun lulus SMTA, kategori SMTA, kWh listrik, dan bidikmisi.

4.2 Pengolahan Data

Pengolahan data dalam penelitian ini menggunakan beberapa metode seperti statistika deskriptif, ANOVA, dan klasterisasi menggunakan algoritma *K-Prototype*.

4.2.1 Preprocessing Data

Penelitian ini melakukan *preprocessing* data untuk variabel-variabel NIM, jenis kelamin, jalur masuk, kategori provinsi, kategori pulau, pekerjaan ayah, tahun lulus SMTA, kategori SMTA, kWh listrik, bidikmisi, dan IPK. Variabel tersebut jika memiliki *not available* (n/a) pada data perlu dilakukan perbaikan data dengan melakukan standarisasi data untuk menghilangkan *missing value*.

Tabel 3. Variabel Kategori

Variabel	Kategori
Jenis Kelamin	1 = Laki-laki (L) 2 = Perempuan (P)
Jalur Masuk	1 = SNMPTN 2 = SBMPTN 3 = SMMPTN
Kategori Provinsi	1 = Prov. Banten 2 = Luar Prov. Banten
Kategori Pulau	1 = Pulau Jawa 2 = Luar Pulau Jawa
Pekerjaan Ayah	1 = Karyawan Swasta 2 = Pedagang 3 = Lain-lain 4 = Petani/Nelayan 5 = Wiraswasta 6 = Guru/Dosen Swasta 7 = Buruh 8 = Pegawai Negeri 9 = Pensiunan 10 = Pensiunan PNS/TNI 11 = ABRI 12 = Guru/Dosen Negeri 13 = Pensiunan Peg.Swasta Non Guru/Dosen 14 = PNS Non Guru/Dosen 15 = TNI 16 = Ahli/Profesional Bekerja Perorangan
Tahun Lulus SMTA	1 = 2019 2 = 2020 3 = 2021
Kategori SMTA	1 = SMAN 2 = SMAS 3 = MAN 4 = MAS 5 = SMKS 6 = SMKN 7 = Lain-lain
kWh Listrik	1 = 450 2 = 900 3 = 1300 4 = 2200 5 = 2300
Bidikmisi	1 = Tidak 2 = Ya

4.2.2 Menentukan Variabel Signifikan dengan *One -Way* ANOVA

Pengujian dengan menggunakan metode *one-way* ANOVA dilakukan sebagai berikut.

4.2.2.1 Menentukan Jenis Variabel Pengujian pada *One-Way* ANOVA

Penelitian ini menggunakan asumsi untuk mengetahui variabel-variabel yang memiliki pengaruh pada variabel IPK. Penentuan jenis variabel pengujian yang dilakukan pada uji ANOVA dapat dilihat pada Tabel 4.

Tabel 4. Variabel *Dependent* dan *Independent* ANOVA

Variabel	Jenis Variabel
IPK	<i>Dependent</i>
Jenis Kelamin	<i>Independent</i>
Jalur Masuk	<i>Independent</i>
Kategori Provinsi	<i>Independent</i>
Kategori Pulau	<i>Independent</i>
Pekerjaan Ayah	<i>Independent</i>
Tahun Lulus SMTA	<i>Independent</i>
Kategori SMTA	<i>Independent</i>
kWh Listrik	<i>Independent</i>
Bidikmisi	<i>Independent</i>

Hasil pengujian dengan metode *one-way* ANOVA menggunakan 10 variabel. Selanjutnya jenis variabel terbagi menjadi 2 (dua) jenis yaitu, *dependent* dan *independent*. Jenis variabel *dependent* merupakan variabel yang dipengaruhi oleh jenis variabel *independent*. Penelitian ini menggunakan 1 (satu) variabel *dependent* dan 9 (Sembilan) variabel *independent*. Tabel 4 memberikan informasi mengenai variabel serta jenis variabel yang digunakan pada penelitian ini.

4.2.2.2 Hasil Pengujian Data *One-Way* ANOVA

Hasil pengujian *one-way* ANOVA dalam menentukan variabel signifikan pada Tabel 5.

Tabel 5. Hasil Pengujian Data ANOVA

Variabel	Nilai <i>P-Value</i>
Jenis Kelamin	0.000
Jalur Masuk	0.000
Kategori Provinsi	0.000
Kategori Pulau	0.540
Pekerjaan Ayah	0.054
Tahun Lulus SMTA	0.364
Kategori SMTA	0.013
kWh Listrik	0.003
Bidikmisi	0.000

Penelitian ini memberikan hasil variabel signifikan dalam menentukan karakteristik yang paling utama dalam mengetahui pengaruh variabel yang besar terhadap nilai IPK mahasiswa. Berdasarkan Tabel 5 memberikan informasi mengenai variabel-variabel serta hasil *p-value* yang didapatkan oleh setiap variabel. Hasil pengujian variabel yang dapat diterima sebagai variabel signifikan jika nilai dari *p-value*-nya bernilai $< 0,05$ serta didapatkan hasil variabel signifikan adalah jenis kelamin, jalur masuk, kategori provinsi, kategori SMTA, kWh listrik, dan bidikmisi.

4.2.3 Algoritma *K-Prototype*

Penelitian ini menggunakan algoritma *K-Prototype* dengan menggunakan *software* RStudio. Adapun *syntax* pemrograman dengan menggunakan *K-Prototype* dapat dilihat pada Lampiran 11. Berikut merupakan tahapan dalam melakukan pemrograman *K-Prototype* dengan menggunakan *software* RStudio.

1. Tahap 1

Melakukan input data dengan menggunakan variabel signifikan hasil pengujian ANOVA (Jenis Kelamin, Jalur Masuk, Kategori Provinsi, Kategori SMTA, kWh Listrik, dan Bidikmisi), variabel identitas (NIM), dan variabel numerik (IPK) dengan format *file* csv.

2. Tahap 2

Menetapkan nama baris dengan variabel NIM.

3. Tahap 3

Melakukan pengelompokan variabel-variabel terhadap karakteristiknya untuk karakteristik kategori dan numerik.

4. Tahap 4

Melakukan pemeriksaan data dengan perintah *str()* untuk melihat jenis data dan struktur objek.

5. Tahap 5

Mempersiapkan *package (ClustMixType)*, *package (data.table)*, *package (dplyr)*. Adapun fungsi dari *package* tersebut adalah sebagai berikut

a. *Package (ClustMixType)*

Package ini memiliki fungsi sebagai *Dist()* untuk membentuk matriks jarak (representasi dari jarak antara elemen-elemen dalam satu set data), *kproto()* untuk melakukan klusterisasi dengan menggunakan algoritma *K-Prototype* yang berfungsi untuk melakukan penggabungan data kategori dan numerik dalam sebuah analisis kluster sehingga memungkinkan untuk menemukan kelompok data yang serupa dalam *dataset* campuran yang memiliki berbagai jenis variabel, dan *split()* untuk membagi data menjadi subgrup berdasarkan variabel.

b. *Package (data.table)*

Package ini memiliki fungsi untuk melakukan manipulasi tabel seperti pembaca *file* cepat, penambahan, agregasi data besar, dan lain sebagainya.

c. *Package (dplyr)*

Package ini memiliki fungsi untuk melakukan manipulasi data *frame* (struktur data tabular yang digunakan dalam bahasa pemrograman).

6. Tahap 6

Menentukan pemilihan alternatif secara acak yang berfungsi untuk memilih sebuah pilihan dari sejumlah opsi yang tersedia secara acak, selanjutnya pemrograman diberikan perintah *set.seed()*.

7. Tahap 7

Menentukan jarak setiap objek (variabel) dengan menggunakan jarak *euclidean*.

8. Tahap 8

Menentukan kluster baru dengan melakukan pemilihan menggunakan rentang yang telah ditentukan.

9. Tahap 9

Melakukan analisis statistika pada setiap kluster dengan menggunakan *package (fpc)*.

10. Tahap 10

Melakukan visualisasi data pada setiap kluster dengan menggunakan *package (factoextra)*, *package (cluster)*, *package (NbClust)*. Adapun fungsi dari *package* tersebut adalah sebagai berikut.

a. *Package (factoextra)*

Package ini memiliki fungsi untuk melakukan ekstraksi dan melakukan visualisasi dari data hasil analisis multivariat.

b. *Package (cluster)*

Package ini memiliki fungsi untuk melakukan analisis kluster.

c. *Package (NbClust)*

Package ini memiliki fungsi untuk mencari kluster optimal.

11. Tahap 11

Melakukan perhitungan indeks validasi internal hasil pengelompokan dengan menggunakan fungsi-fungsi *cluster.stats()* dan *fviz_silhouette()* Adapun penjelasan dari fungsi-fungsi tersebut adalah sebagai berikut.

a. Fungsi *cluster.stats()* digunakan untuk menghitung indeks validasi internal hasil pengelompokan yaitu SSW, Koefisien Rata-Rata *Silhouette*, dan Indeks *Dunn*.

b. Fungsi *fviz_silhouette()* digunakan untuk memvisualisasikan hasil pengelompokan.

12. Tahap 12

Melakukan perhitungan indeks validasi eksternal hasil pengelompokan dengan menggunakan fungsi *entropy()* untuk menghitung nilai entropi.

13. Tahap 13

Menentukan kluster optimal dengan menggunakan hasil validasi internal dan validasi eksternal.

4.2.4 Karakteristik Hasil Pengelompokan Metode *K-Prototype*

Berdasarkan Lampiran 12 yang di mana dihasilkan 6 klaster. Hasil dari setiap klaster ditentukan dari variabel-variabel jenis kelamin, jalur masuk, kategori provinsi, kategori SMTA, kWh listrik bidikmisi, dan IPK. Dari hasil tersebut dapat diketahui bahwa data memiliki 6 pilihan alternatif pengelompokan.

4.2.5 Perhitungan Indeks Validasi Internal Metode *K-Prototype*

Hasil validasi internal data dapat dilihat pada Tabel 6.

**Tabel 6. Hasil Validasi Internal
Metode Algoritma *K-Prototype***

Indeks Validasi Internal	2	3	4	5	6
Rata-rata Koefisien <i>Silhouette</i>	0,267	0,038	0,079	-0,063	-0,100
Indeks <i>Dunn</i>	0,000005	0	0	0	0

Hasil validasi internal menampilkan untuk setiap klaster 2-6 klaster. Hasil validasi menampilkan rata-rata koefisien *silhouette*, dan Indeks *dunn*. Dalam menentukan klaster optimal dilihat dari nilai rata-rata koefisien *silhouette* terbesar dan indeks *dunn* terbesar.

4.2.6 Perhitungan Indeks Validasi Eksternal Metode *K-Prototype*

Hasil dari validasi eksternal dapat dilihat pada Tabel 7.

**Tabel 7. Hasil Validasi Eksternal
Metode Algoritma *K-Prototypes***

Indeks Validasi Eksternal	2	3	4	5	6
<i>Entropy</i>	0,692	0,845	133451861055805	15603550412191	166879368250958

Hasil validasi eksternal menampilkan untuk setiap klaster 2-6 klaster. Selanjutnya ditampilkan nilai ukuran kualitas eksternal (*entropy*) untuk setiap klaster. Dalam menentukan klaster optimal dilihat dari nilai *entropy* terkecil.

4.2.7 Statistika Deskriptif Klaster

Penelitian ini menampilkan statistika deskriptif dari klaster optimal yang didapatkan yaitu pengelompokan 2 klaster.

4.2.7.1 Klaster 1

Statistika Deskriptif untuk klaster 1 dapat dilihat pada Tabel 8. Klaster 1 memiliki jumlah data 2142 mahasiswa yang dikelompokan pada klaster ini. Klaster ini menunjukkan pada variabel IPK minimal pada klaster ini yaitu 0,280 dan nilai IPK maksimal adalah 3,980. Selanjutnya didapatkan nilai standar deviasi yaitu 0,340 dan rata-rata 3,590.

Tabel 8. Statistika Deskriptif Klaster 1

Minimal	Maksimal	Median	Standar Deviasi	Rata-rata	Jumlah
0,280	3,980	3,590	0,340	3,422	2142

Contoh Perhitung:

1. Minimal (nilai terkecil)

$$\text{Minimal} = 0,280$$

2. Maksimal (nilai terbesar)

$$\text{Maksimal} = 3,980$$

3. Median (Me)

$$\text{Me} = \frac{X_{\frac{n}{2}} + X_{(\frac{n}{2}+1)}}{2}$$

$$\text{Me} = \frac{X_{\frac{2142}{2}} + X_{(\frac{2142}{2}+1)}}{2}$$

$$\text{Me} = \frac{X_{1071} + X_{1072}}{2}$$

$$\text{Me} = \frac{3,590 + 3,590}{2}$$

$$\text{Me} = 3,590$$

4. Standar deviasi (S)

$$S = \sqrt{\frac{\sum(X_i - \bar{X})^2}{n}}$$

$$S = \sqrt{\frac{(3,760 - 3,422)^2 + (3,890 - 3,422)^2 + \dots + (3,550 - 3,422)^2}{2142}}$$

$$S = 0,340$$

5. Rata-rata (\bar{X})

$$\bar{X} = \frac{X_1 + X_2 + X_3 + \dots + X_{2142}}{1954}$$

$$\bar{X} = \frac{3,760 + 3,890 + 3,880 + \dots + 3,550}{1954}$$

$$\bar{X} = 3,422$$

6. Jumlah

$$\text{Jumlah} = 2142$$

4.2.7.2 Klaster 2

Statistika Deskriptif untuk klaster 2 dapat dilihat pada Tabel 9. Klaster 2 memiliki jumlah data 1954 mahasiswa yang dikelompokan pada klaster ini. Klaster ini menunjukkan pada variabel IPK minimal pada klaster ini yaitu 2,770 dan nilai IPK maksimal adalah 3,990. Selanjutnya didapatkan nilai standar deviasi yaitu 0,031 dan rata-rata 3,701.

Tabel 9. Statistika Deskriptif Klaster 2

Minimal	Maksimal	Median	Standar Deviasi	Rata-rata	Jumlah
2,770	3,990	3,740	0,031	3,701	1954

Contoh Perhitung:

7. Minimal (nilai terkecil)

$$\text{Minimal} = 2,770$$

8. Maksimal (nilai terbesar)

$$\text{Maksimal} = 3,990$$

9. Median (Me)

$$\text{Me} = \frac{X_{\frac{n}{2}} + X_{(\frac{n}{2}+1)}}{2}$$

$$\text{Me} = \frac{X_{\frac{1954}{2}} + X_{(\frac{1954}{2}+1)}}{2}$$

$$Me = \frac{X_{977} + X_{978}}{2}$$

$$Me = \frac{3,740 + 3,740}{2}$$

$$Me = 3,740$$

10. Standar deviasi (S)

$$S = \sqrt{\frac{\sum(X_i - \bar{X})^2}{n}}$$

$$S = \sqrt{\frac{(3,980 - 3,701)^2 + (3,920 - 3,701)^2 + \dots + (2,920 - 3,701)^2}{1954}}$$

$$S = 0,031$$

11. Rata-rata (\bar{X})

$$\bar{X} = \frac{X_1 + X_2 + X_3 + \dots + X_{1954}}{1954}$$

$$\bar{X} = \frac{3,980 + 3,920 + 3,920 + \dots + 2,920}{1954}$$

$$\bar{X} = 3,701$$

12. Jumlah

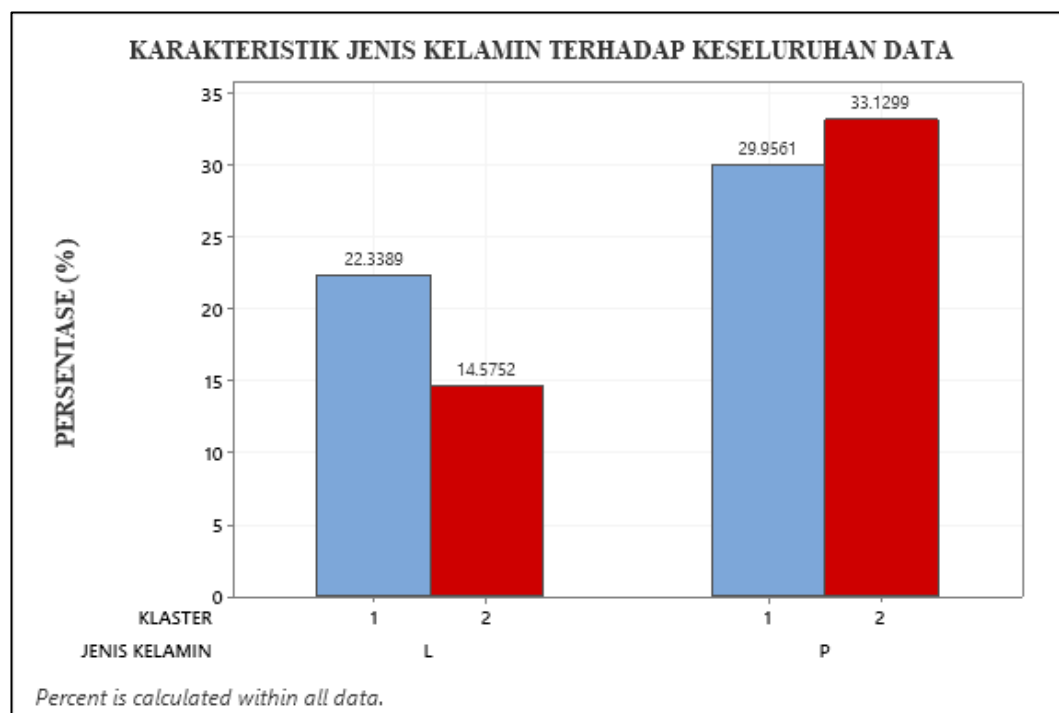
$$\text{Jumlah} = 1954$$

4.2.8 Perbandingan Klaster

Adapun perbandingan pada setiap klaster adalah sebagai berikut.

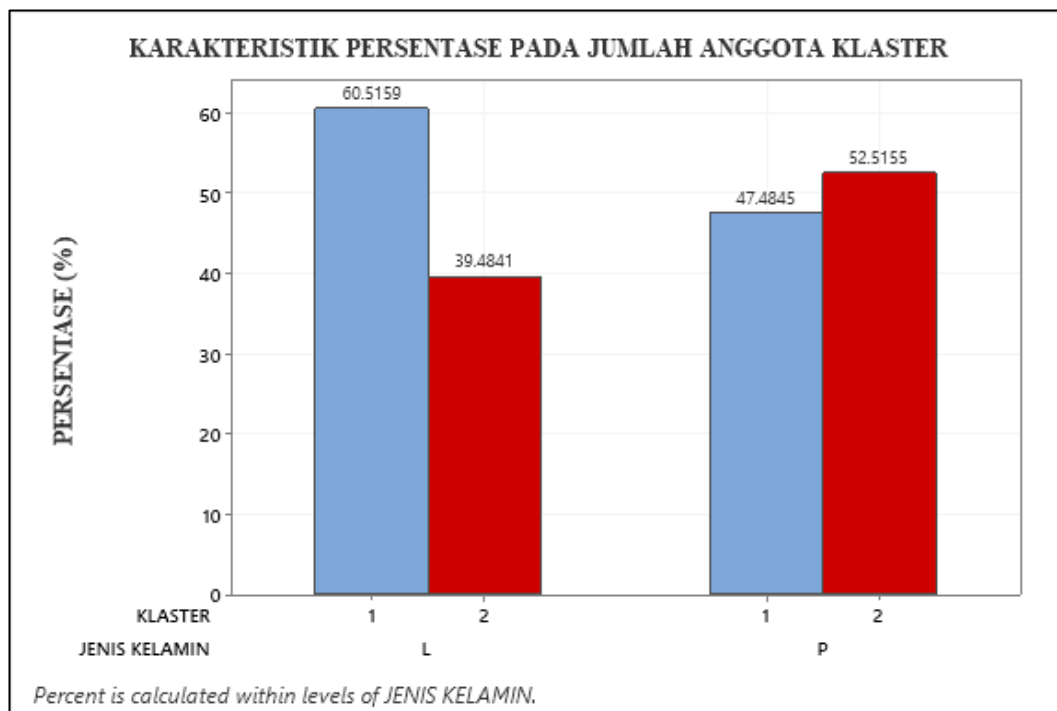
4.2.8.1 Karakteristik Jenis Kelamin

Perbandingan antara klaster 1 dengan klaster 2 dapat dilihat pada Gambar 3 dan Gambar 4 untuk karakteristik jenis kelamin.



Gambar 3. Karakteristik Jenis Kelamin Berdasarkan Keseluruhan Data

Gambar 3 menampilkan gambaran persentase karakteristik jenis kelamin berdasarkan keseluruhan data serta memberikan informasi berhubungan dengan perbandingan antara jumlah data mahasiswa yang berjenis kelamin laki-laki (L) dengan mahasiswa yang berjenis kelamin perempuan (P) dengan jumlah keseluruhan populasi adalah 4096 mahasiswa. Klaster 1 memiliki jumlah data yaitu 2142 mahasiswa dengan persentase populasi 52,294%. Selanjutnya rincian penyebaran populasi pada klaster 1 yaitu, jumlah mahasiswa laki-laki berjumlah 915 mahasiswa dengan persentase 22,338% sedangkan mahasiswa perempuan berjumlah 1227 mahasiswa dengan persentase 29,956%. Klaster 2 memiliki jumlah data yaitu 1954 mahasiswa dengan persentase populasi 47,705%. Selanjutnya rincian penyebaran populasi pada klaster 2 yaitu, jumlah mahasiswa laki-laki berjumlah 597 mahasiswa dengan persentase 14,575% sedangkan mahasiswa perempuan berjumlah 1357 mahasiswa dengan persentase 33,129%.

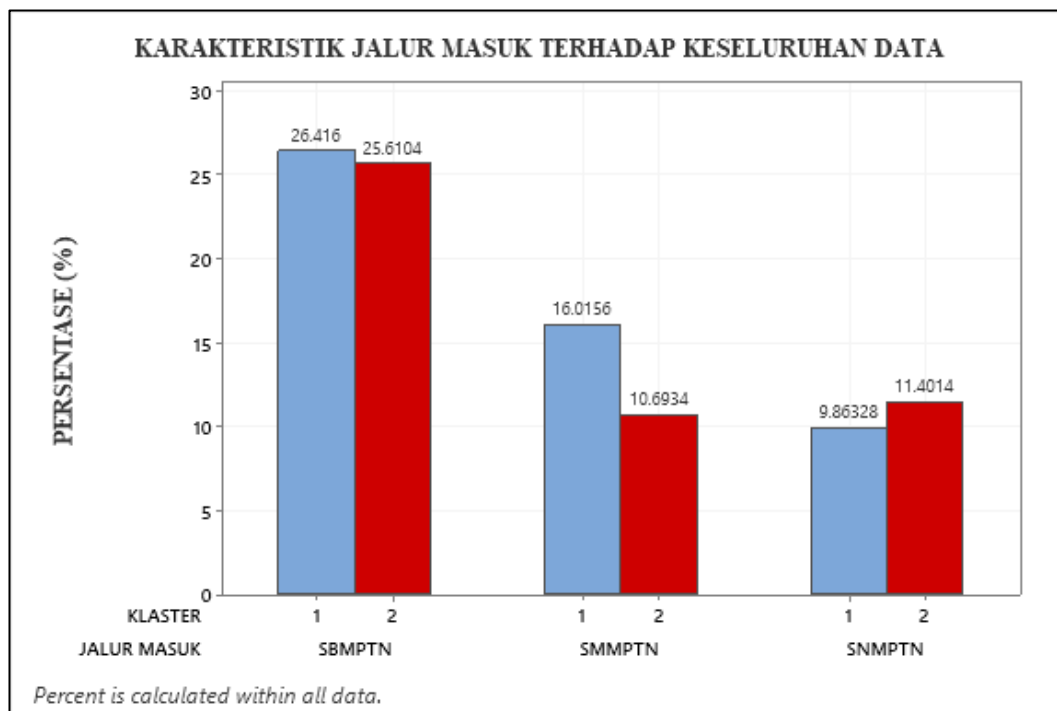


Gambar 4. Karakteristik Jenis Kelamin Berdasarkan Anggota Klaster

Gambar 4 menampilkan gambaran persentase karakteristik jenis kelamin berdasarkan klasternya serta memberikan informasi berhubungan dengan perbandingan antara jumlah data mahasiswa yang berjenis kelamin laki-laki dengan mahasiswa yang berjenis kelamin perempuan. Klaster 1 untuk mahasiswa berjenis kelamin laki-laki memiliki persentase 60,515% serta untuk mahasiswa perempuan memiliki persentase 47,484%. Selanjutnya pada klaster 2 untuk mahasiswa berjenis kelamin laki-laki memiliki persentase 39,484% serta untuk mahasiswa perempuan memiliki persentase 52,515%.

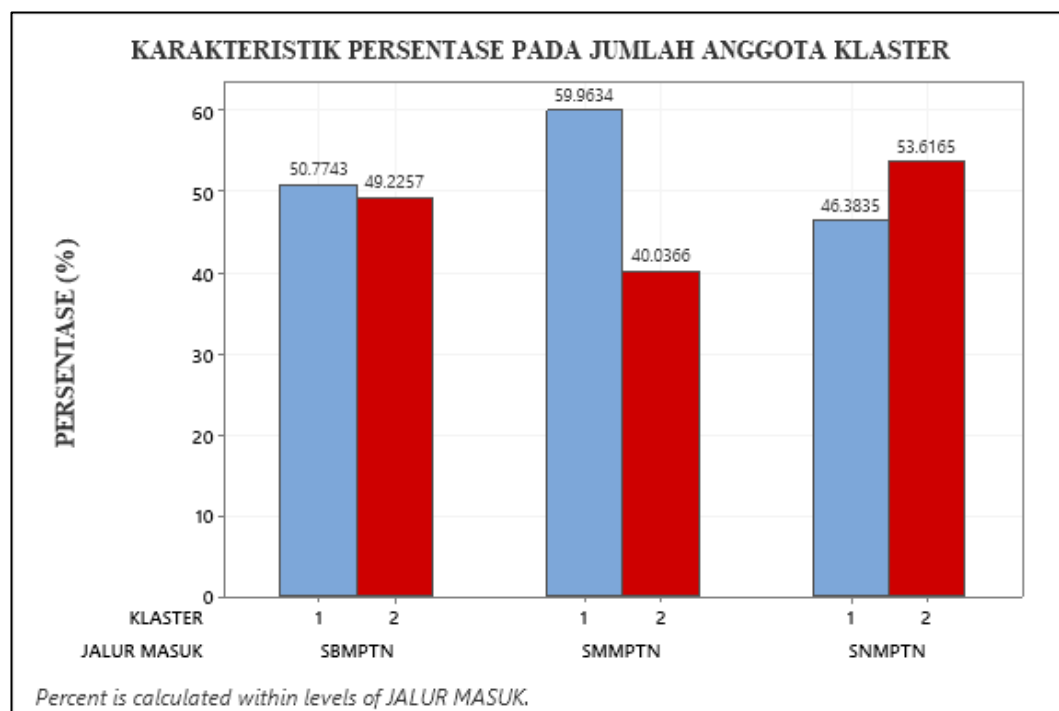
4.2.8.2 Karakteristik Jalur Masuk

Perbandingan antara klaster 1 dengan klaster 2 dapat dilihat pada Gambar 5 dan Gambar 6 untuk karakteristik jalur masuk.



Gambar 5. Karakteristik Jalur Masuk Berdasarkan Keseluruhan Data

Gambar 5 menampilkan gambaran persentase karakteristik jalur masuk berdasarkan keseluruhan data serta memberikan informasi berhubungan dengan perbandingan antara jumlah data mahasiswa dengan jalur masuk SNMPTN, SBMPTN, dan SMMPTN dengan jumlah keseluruhan populasi adalah 4096 mahasiswa. Klaster 1 memiliki jumlah data yaitu 2142 mahasiswa dengan persentase populasi 52,294%. Selanjutnya rincian penyebaran populasi pada klaster 1 yaitu, pada jalur masuk SNMPTN berjumlah 404 mahasiswa dengan persentase 9,86328%, mahasiswa pada jalur masuk SBMPTN berjumlah 1082 mahasiswa dengan persentase 26,416%, dan mahasiswa pada jalur masuk SMMPTN berjumlah 656 mahasiswa dengan persentase 16,015%. Klaster 2 memiliki jumlah data yaitu 1954 mahasiswa dengan persentase populasi 47,705%. Selanjutnya rincian penyebaran populasi pada klaster 2 yaitu, pada jalur masuk SNMPTN berjumlah 467 mahasiswa dengan persentase 11,401%, mahasiswa pada jalur masuk SBMPTN berjumlah 1049 mahasiswa dengan persentase 25,610%, dan mahasiswa pada jalur masuk SMMPTN berjumlah 438 mahasiswa dengan persentase 10,693%.

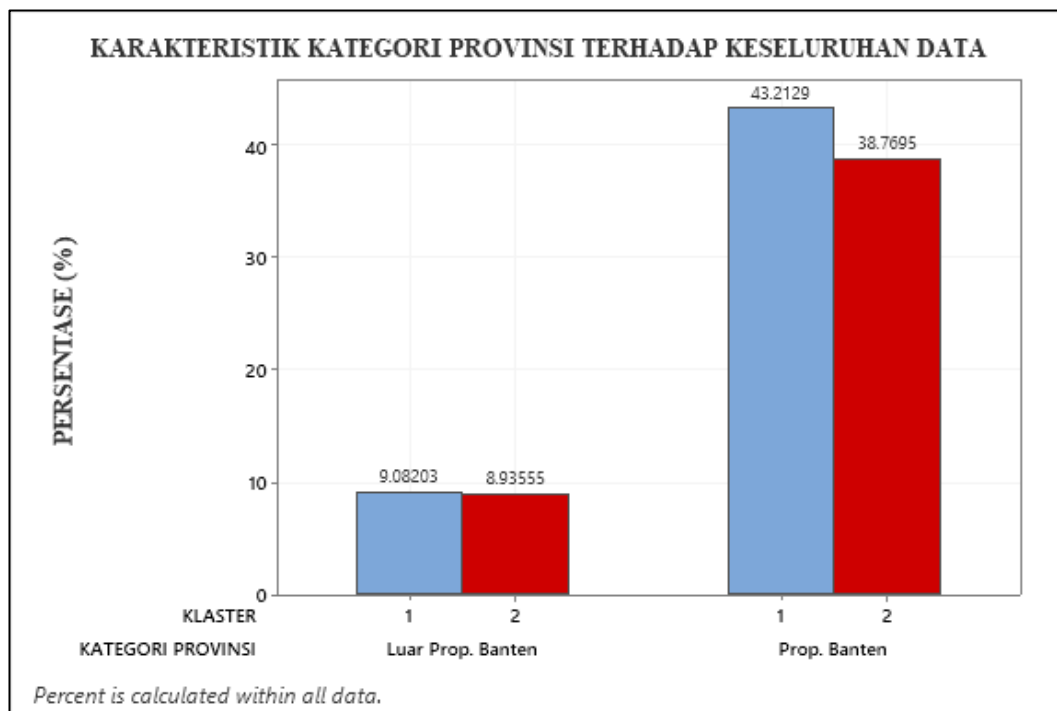


Gambar 6. Karakteristik Jalur Masuk Berdasarkan Anggota Klaster

Gambar 6 menampilkan gambaran persentase karakteristik jalur masuk berdasarkan klasternya serta memberikan informasi berhubungan dengan perbandingan antara jumlah data mahasiswa dengan jalur masuk SNMPTN, SBMPTN, dan SMMPTN. Klaster 1 untuk pada jalur masuk SNMPTN memiliki persentase 46,383%, mahasiswa pada jalur masuk SBMPTN memiliki persentase 50,774%, dan mahasiswa pada jalur masuk SMMPTN memiliki persentase 59,963%. Selanjutnya pada klaster 2 untuk pada jalur masuk SNMPTN memiliki persentase 53,616%, mahasiswa pada jalur masuk SBMPTN memiliki persentase 49,225%, dan mahasiswa pada jalur masuk SMMPTN memiliki persentase 40,036%.

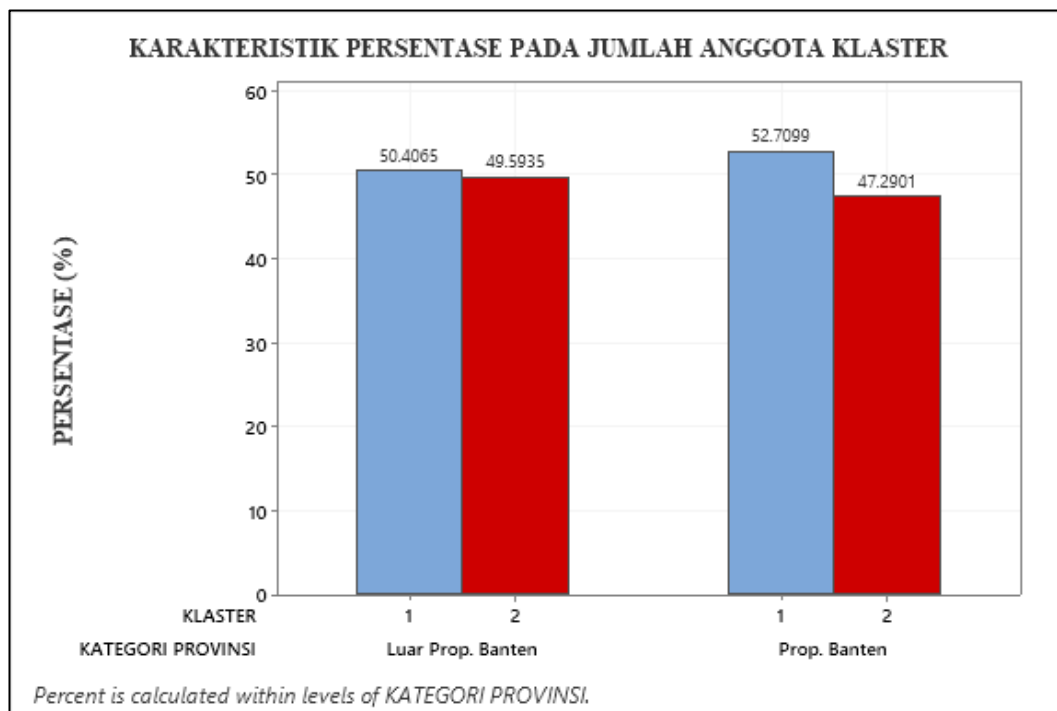
4.2.8.3 Karakteristik Kategori Provinsi

Perbandingan antara klaster 1 dengan klaster 2 dapat dilihat pada Gambar 7 dan Gambar 8 untuk karakteristik kategori provinsi.



Gambar 7. Karakteristik Kategori Provinsi Berdasarkan Keseluruhan Data

Gambar 7 menampilkan gambaran persentase karakteristik kategori provinsi berdasarkan keseluruhan data serta memberikan informasi berhubungan dengan perbandingan persentase antara jumlah mahasiswa yang berasal dari luar Provinsi Banten dengan mahasiswa yang berasal dari Provinsi Banten dengan jumlah keseluruhan populasi adalah 4096 mahasiswa. Klaster 1 memiliki jumlah data yaitu 2142 mahasiswa dengan persentase populasi 52,294%. Selanjutnya rincian penyebaran populasi pada klaster 1 yaitu, mahasiswa yang berasal dari luar Provinsi Banten berjumlah 372 mahasiswa dengan persentase 9,082% sedangkan mahasiswa yang berasal dari Provinsi Banten berjumlah 1770 mahasiswa dengan persentase 43,212%. Klaster 2 memiliki jumlah data yaitu 1954 mahasiswa dengan persentase populasi 47,705%. Selanjutnya rincian penyebaran populasi pada klaster 2 yaitu, mahasiswa yang berasal dari luar Provinsi Banten berjumlah 366 mahasiswa dengan persentase 8,935% sedangkan mahasiswa yang berasal dari Provinsi Banten berjumlah 1588 mahasiswa dengan persentase 38,769 %.

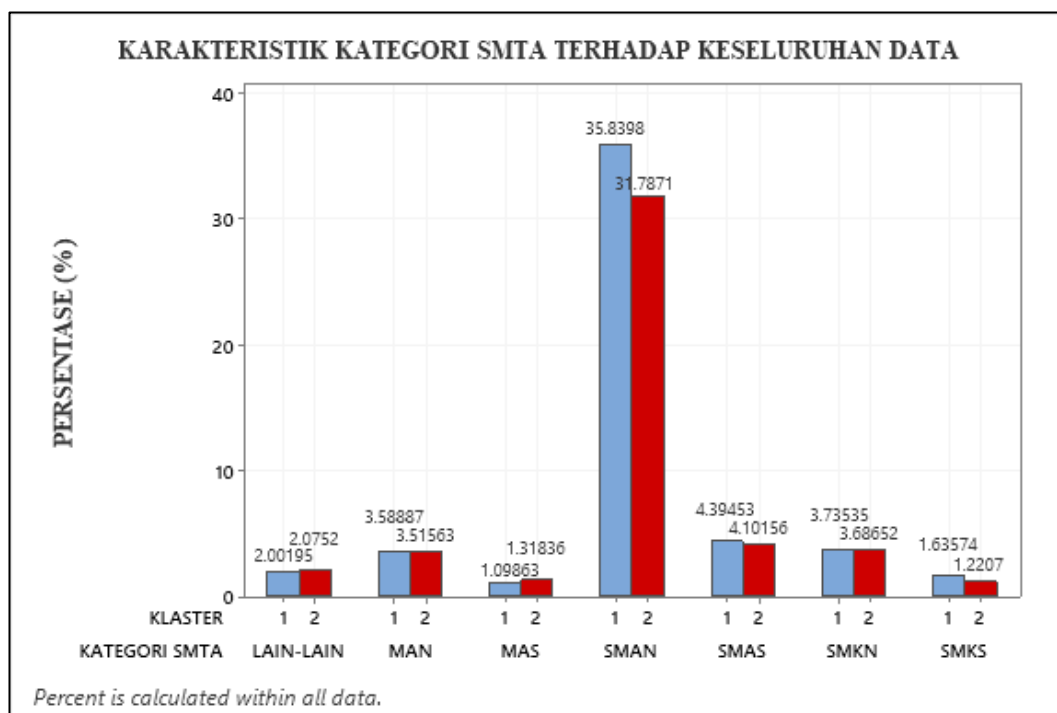


Gambar 8. Karakteristik Kategori Provinsi Berdasarkan Anggota Klaster

Gambar 8 menampilkan gambaran persentase karakteristik kategori provinsi berdasarkan klasternya serta memberikan informasi berhubungan dengan perbandingan persentase antara jumlah mahasiswa yang berasal dari luar Provinsi Banten dengan mahasiswa yang berasal dari Provinsi Banten. Klaster 1 untuk mahasiswa yang berasal dari luar Provinsi Banten memiliki persentase 50,406% sedangkan mahasiswa yang berasal dari Provinsi Banten memiliki persentase 43,212%. Selanjutnya pada klaster 2 untuk mahasiswa yang berasal dari luar Provinsi Banten memiliki persentase 49,593% sedangkan mahasiswa yang berasal dari Provinsi Banten memiliki persentase 47,290%.

4.2.8.4 Karakteristik Kategori SMTA

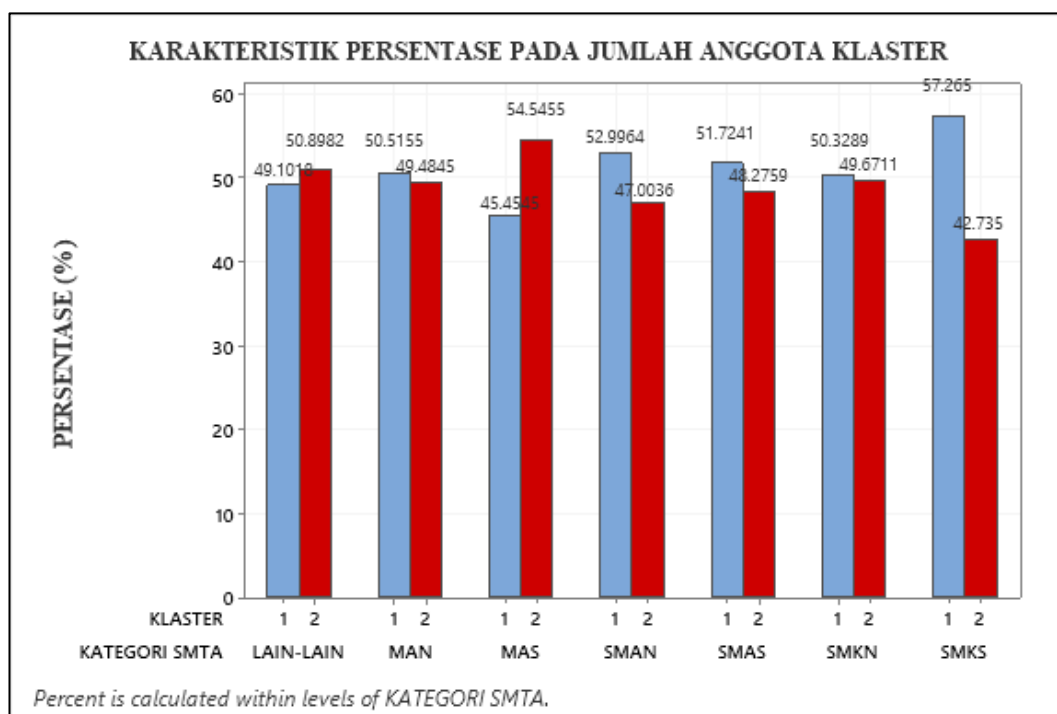
Perbandingan antara klaster 1 dengan klaster 2 dapat dilihat pada Gambar 9 dan Gambar 10 untuk karakteristik kategori SMTA.



Gambar 9. Karakteristik Kategori SMTA Berdasarkan Keseluruhan Data

Gambar 9 menampilkan gambaran persentase karakteristik kategori SMTA berdasarkan keseluruhan data serta memberikan informasi berhubungan dengan perbandingan antara jumlah mahasiswa yang berasal dari beberapa jenis SMTA seperti SMAN, MAN, SMAS, SMKS, SMKN, MAS, dan lain-lain dengan jumlah keseluruhan populasi adalah 4096 mahasiswa. Klaster 1 memiliki jumlah data yaitu 2142 mahasiswa dengan persentase populasi 52,294%. Selanjutnya rincian penyebaran populasi pada klaster 1 yaitu, mahasiswa yang berasal dari SMAN berjumlah 1468 mahasiswa dengan persentase 35,839%, MAN berjumlah 147 mahasiswa dengan persentase 3,588%, SMAS berjumlah 180 mahasiswa dengan persentase 4,394%, SMKS berjumlah 67 mahasiswa dengan persentase 1,635%, SMKN berjumlah 153 mahasiswa dengan persentase 3,735%, MAS berjumlah 45 mahasiswa dengan persentase 1,098%, dan lain-lain berjumlah 82 mahasiswa dengan persentase 2,001%. Klaster 2 memiliki jumlah data yaitu 1954 mahasiswa dengan persentase populasi 47,705%. Selanjutnya rincian penyebaran populasi pada klaster 2 yaitu, mahasiswa yang berasal dari SMAN berjumlah 1302 mahasiswa dengan persentase 31,787%, MAN berjumlah 144 mahasiswa dengan persentase 3,515%, SMAS berjumlah 168 mahasiswa dengan persentase 4,101%,

SMKS berjumlah 50 mahasiswa dengan persentase 1,220%, SMKN berjumlah 151 mahasiswa dengan persentase 1,220%, MAS berjumlah 54 mahasiswa dengan jumlah persentase 1,318%, dan lain-lain berjumlah 85 mahasiswa dengan persentase 2,075%.

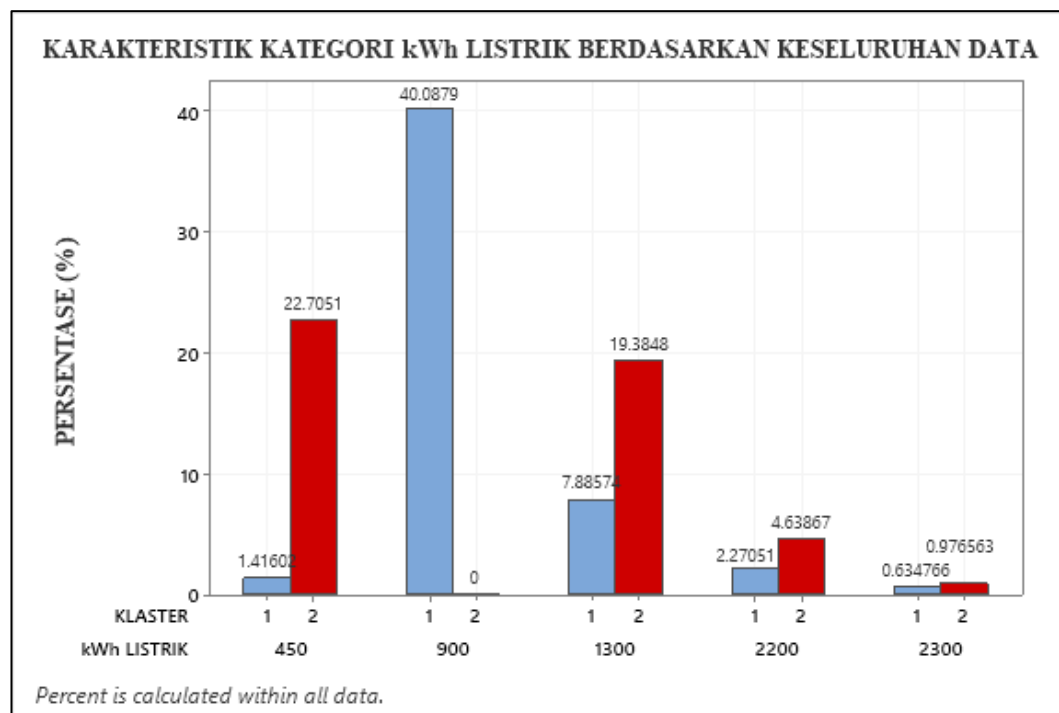


Gambar 10. Karakteristik Kategori SMTA Berdasarkan Anggota Klaster

Gambar 10 menampilkan gambaran persentase karakteristik kategori SMTA berdasarkan klasternya serta memberikan informasi berhubungan dengan perbandingan antara jumlah mahasiswa yang berasal dari beberapa jenis SMTA seperti SMAN, MAN, SMAS, SMKS, SMKN, MAS, dan lain-lain. Klaster 1 untuk mahasiswa yang berasal dari SMAN memiliki persentase 52,996%, MAN memiliki persentase 50,515%, SMAS berjumlah memiliki persentase 51,724%, SMKS memiliki persentase 57,265%, SMKN memiliki persentase 50,328%, MAS memiliki persentase 45,4545%, dan lain-lain memiliki persentase 49,101%. Selanjutnya pada klaster 2 untuk mahasiswa yang berasal dari SMAN memiliki persentase 47,004%, MAN memiliki persentase 49,484%, SMAS memiliki persentase 48,275%, SMKS memiliki persentase 42,735%, SMKN memiliki persentase 49,671%, MAS memiliki persentase 54,545%, dan lain-lain memiliki persentase 50,898%.

4.2.8.5 Karakteristik Kategori kWh Listrik

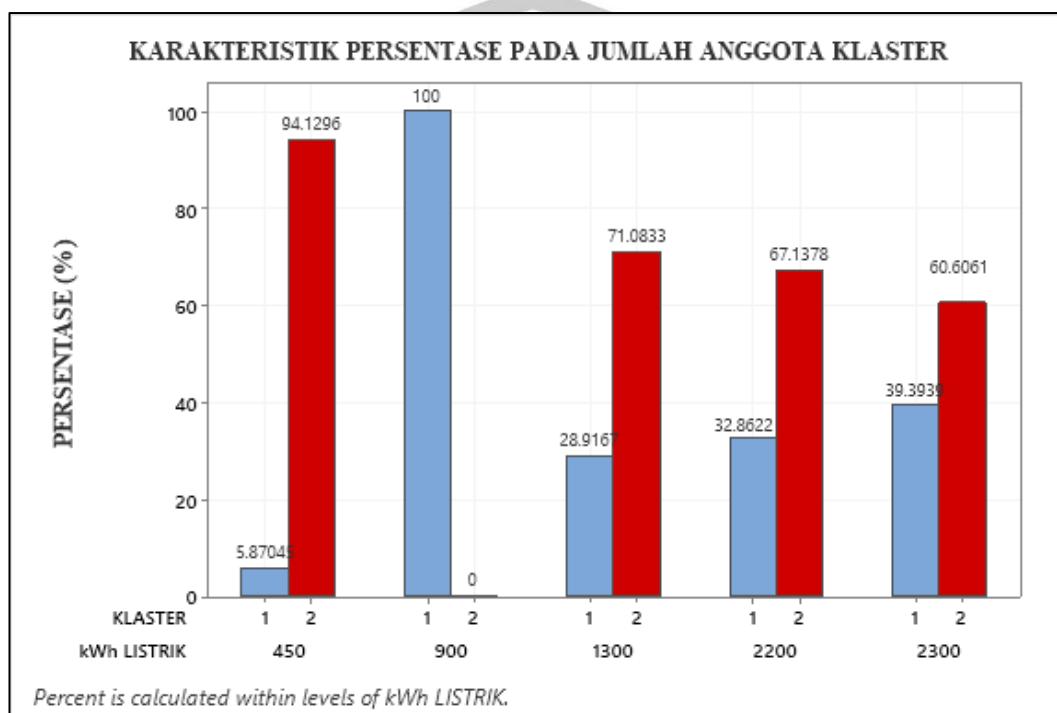
Perbandingan antara klaster 1 dengan klaster 2 dapat dilihat pada Gambar 11 dan Gambar 12 untuk karakteristik kategori kWh Listrik.



Gambar 11. Karakteristik Kategori kWh Listrik Berdasarkan Keseluruhan Data

Gambar 11 menampilkan gambaran persentase karakteristik kategori kWh listrik berdasarkan keseluruhan data serta memberikan informasi berhubungan dengan perbandingan antara jumlah mahasiswa yang menggunakan kWh listrik dari beberapa kategori kWh listrik antara lain seperti 450 watt, 900 watt, 1300 watt, 2200 watt, dan 2300 watt dengan jumlah keseluruhan populasi adalah 4096 mahasiswa. Klaster 1 memiliki jumlah data yaitu 2142 mahasiswa dengan persentase populasi 52,294%. Selanjutnya rincian penyebaran populasi pada klaster 1 yaitu, mahasiswa yang menggunakan kategori kWh listrik 450 watt berjumlah 58 mahasiswa dengan persentase 1,416%, kWh listrik 900 watt berjumlah 1642 mahasiswa dengan persentase 40,087%, kWh listrik 1300 watt berjumlah 323 mahasiswa dengan persentase 7,885%, kWh listrik 2200 watt berjumlah 93 mahasiswa dengan persentase 2,270%, dan kWh listrik 2300 watt berjumlah 26 mahasiswa dengan persentase 0,634%. Klaster 2 memiliki jumlah data yaitu 1954

mahasiswa dengan persentase populasi 47,705%. Selanjutnya rincian penyebaran populasi pada klaster 2 yaitu, mahasiswa yang menggunakan kategori kWh listrik 450 watt berjumlah 930 mahasiswa dengan persentase 22,705%, kWh listrik 900 watt berjumlah 0 mahasiswa dengan persentase 0%, kWh listrik 1300 watt berjumlah 794 mahasiswa dengan persentase 19,384%, kWh listrik 2200 watt berjumlah 190 mahasiswa dengan persentase 4,638%, dan kWh listrik 2300 watt berjumlah 40 mahasiswa dengan persentase 0,976%.



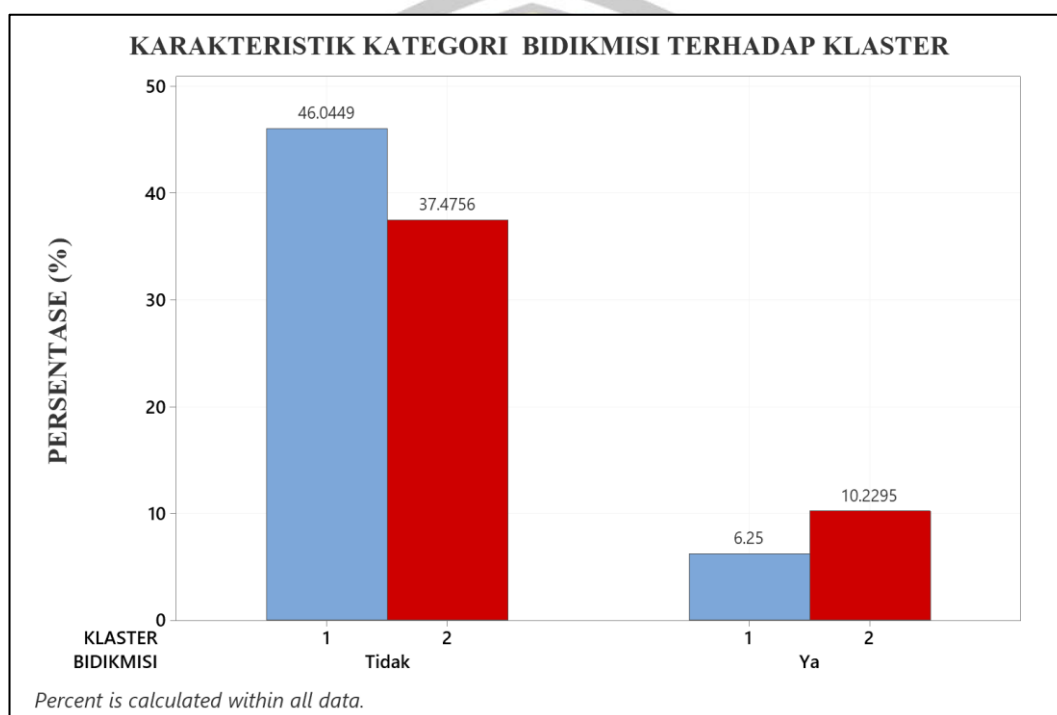
Gambar 12. Karakteristik Kategori kWh Listrik Berdasarkan Anggota Klaster

Gambar 12 menampilkan gambaran persentase karakteristik kategori kWh listrik berdasarkan klasternya serta memberikan informasi berhubungan dengan perbandingan antara jumlah mahasiswa yang menggunakan kWh listrik dari beberapa kategori kWh listrik antara lain seperti 450 watt, 900 watt, 1300 watt, 2200 watt, dan 2300 watt. Klaster 1 untuk mahasiswa yang menggunakan kategori kWh listrik 450 watt memiliki persentase 5,87045%, kWh listrik 900 watt berjumlah memiliki persentase 100%, kWh listrik 1300 watt memiliki persentase 28,916%, kWh listrik 2200 watt memiliki persentase 32,862%, dan kWh listrik 2300 watt memiliki persentase 39,393%. Selanjutnya pada klaster 2 untuk

mahasiswa yang menggunakan kategori kWh listrik 450 watt memiliki persentase 94,129%, kWh listrik 900 watt memiliki persentase 0%, kWh listrik 1300 watt memiliki persentase 71,083%, kWh listrik 2200 memiliki persentase 67,137%, dan kWh listrik 2300 watt memiliki persentase 60,606%.

4.2.8.6 Karakteristik Kategori Bidikmisi

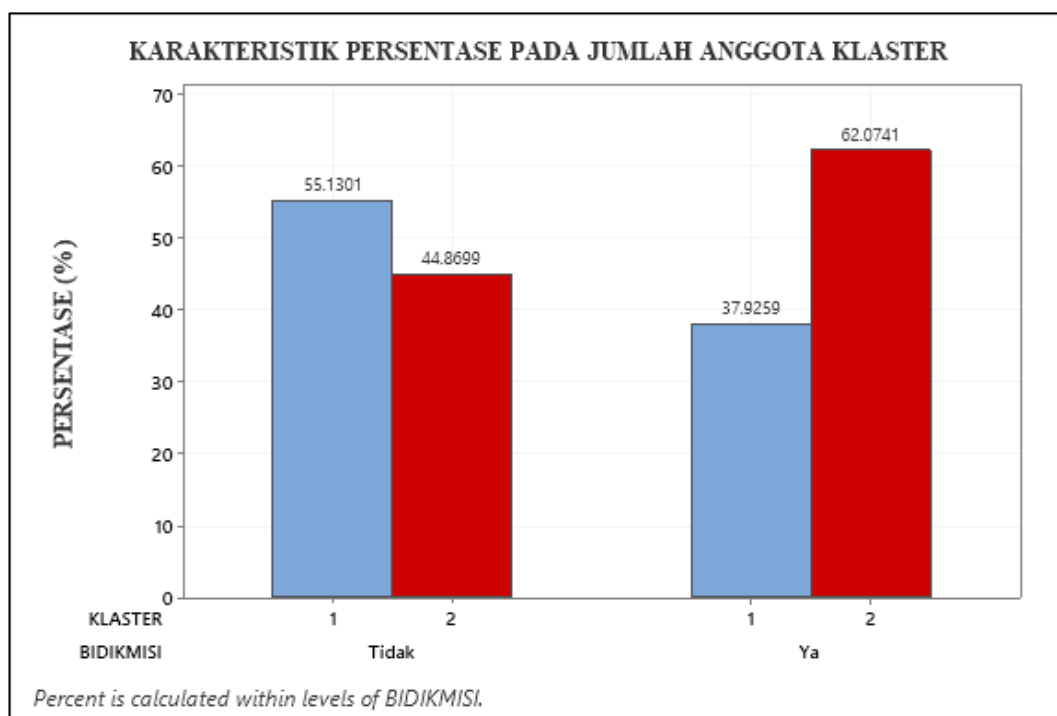
Perbandingan antara klaster 1 dengan klaster 2 dapat dilihat pada Gambar 13 dan Gambar 14 untuk karakteristik kategori bidikmisi.



Gambar 13. Karakteristik Kategori Bidikmisi Berdasarkan Keseluruhan Data

Gambar 13 menampilkan gambaran persentase karakteristik kategori bidikmisi berdasarkan keseluruhan data serta memberikan informasi berhubungan dengan perbandingan antara jumlah mahasiswa yang memiliki bidikmisi dan yang tidak memiliki bidikmisi dengan jumlah keseluruhan populasi adalah 4096 mahasiswa. Klaster 1 memiliki jumlah data yaitu 2142 mahasiswa dengan persentase populasi 52,294%. Selanjutnya rincian penyebaran populasi pada klaster 1 yaitu, mahasiswa yang memiliki bidikmisi berjumlah 256 mahasiswa dengan persentase 6,25% sedangkan yang tidak memiliki bidikmisi berjumlah 256 mahasiswa dengan persentase 46,044%. Klaster 2 memiliki jumlah data yaitu 1954

mahasiswa dengan persentase populasi 47,705%. Selanjutnya rincian penyebaran populasi pada klaster 2 yaitu, mahasiswa yang memiliki bidikmisi berjumlah 419 mahasiswa dengan persentase 10,229% sedangkan yang tidak memiliki bidikmisi berjumlah 1535 mahasiswa dengan persentase 37,475%.



Gambar 14 Karakteristik Kategori Bidikmisi Berdasarkan Anggota Klaster

Gambar 14 menampilkan gambaran persentase karakteristik kategori bidikmisi berdasarkan klasternya serta memberikan informasi berhubungan dengan perbandingan antara jumlah mahasiswa yang memiliki bidikmisi dan yang tidak memiliki bidikmisi. Klaster 1 untuk mahasiswa yang memiliki bidikmisi memiliki persentase 37,925% sedangkan yang tidak memiliki bidikmisi memiliki persentase 55,13%. Selanjutnya pada klaster 2 untuk mahasiswa yang memiliki bidikmisi memiliki persentase 62,074% sedangkan yang tidak memiliki bidikmisi memiliki persentase 44,869%.

4.2.9 Menentukan Variabel Signifikan dengan *General Linear Model (GLM)*

Pengujian dengan menggunakan metode *general linear model* dilakukan sebagai berikut.

4.2.9.1 Menentukan Jenis Variabel Pengujian pada *General Linear Model* (GLM)

hasil pengolahan dengan menggunakan *K-Prototype*, dari hasil pengolahan data tersebut dilakukan penentuan variabel untuk mengetahui hubungan variabel *independent* dengan variabel *dependent*.

Tabel 10. Variabel *Dependent* dan *Independent* GLM

Variabel	Jenis Variabel
Klaster	<i>Dependent</i>
Jenis Kelamin	<i>Independent</i>
Jalur Masuk	<i>Independent</i>
Kategori Provinsi	<i>Independent</i>
Kategori SMTA	<i>Independent</i>
kWh Listrik	<i>Independent</i>
Bidikmisi	<i>Independent</i>
IPK	<i>Independent</i>

Hasil pengujian dengan metode GLM menggunakan 8 variabel. Metode GLM membagi jenis variabelnya menjadi 2 (dua) jenis yaitu, *dependent* dan *independent*. Jenis variabel *dependent* merupakan variabel yang dipengaruhi oleh jenis variabel *independent*. Penelitian ini menggunakan 1 (satu) variabel *dependent* yaitu variabel klaster dan 7 (tujuh) variabel *independent* yaitu jenis kelamin, jalur masuk, kategori provinsi, kategori SMTA, kWh listrik, Bidikmisi, dan IPK. Tabel 10 memberikan informasi mengenai variabel serta jenis variabel yang digunakan pada penelitian ini.

4.2.9.2 Hasil Pengujian Data *General Linear Model* (GLM)

Hasil pengujian GLM mengetahui hubungan yang signifikan secara statistik dapat dilihat pada Tabel 11.

Tabel 11. Hasil Pengujian Data GLM

Variabel	DF	Nilai <i>F-Value</i>	Nilai <i>P-Value</i>
Jenis Kelamin	1	0,01	0,907
Jalur Masuk	2	1,19	0,304
Kategori Provinsi	1	0,71	0,401
Kategori SMTA	6	2,18	0,042
kWh Listrik	4	3348,76	0,000
Bidikmisi	1	5,78	0,016
IPK	239	13,53	0

Penelitian ini memberikan hasil hubungan yang signifikan secara statistik terhadap klaster mahasiswa. Berdasarkan Tabel 11 memberikan informasi mengenai hubungan variabel klaster dengan variabel-variabel jenis kelamin, jalur

masuk, kategori provinsi, kategori SMTA, kWh listrik, bidikmisi, dan IPK. Penelitian ini mendapatkan variabel-variabel yang signifikan antara lain adalah kategori SMTA, kWh listrik, bidikmisi dan IPK.



BAB V

ANALISIS DAN PEMBAHASAN

5.1 Analisis Variabel-Variabel yang Signifikan terhadap Variabel IPK

Pengumpulan data berasal dari populasi mahasiswa UNTIRTA angkatan 2021/2022 berdasarkan data IPK dan data biodata mahasiswa didapatkan 4096 mahasiswa yang memiliki kelengkapan data yang telah memenuhi persyaratan. Data tersebut didapatkan dari basis data UNTIRTA. Data dari mahasiswa berisikan variabel-variabel jenis kelamin, jalur masuk, kategori provinsi, kategori pulau, pekerjaan ayah, tahun lulus SMTA, kategori SMTA, kWh listrik, dan bidikmisi. Variabel-variabel pada penelitian ini dapat dilihat pada Lampiran 1.

Data 4096 mahasiswa akan diuji dengan metode *one-way* ANOVA untuk mendapatkan variabel-variabel signifikan yang memiliki hubungan pengaruh yang sangat signifikan. Penelitian ini memiliki asumsi bahwa variabel IPK adalah variabel *dependent* yang terpengaruh oleh variabel lain oleh sebab itu perlu mencari hubungan yang mempengaruhi variabel IPK. Variabel-variabel yang sekiranya memiliki pengaruh terhadap variabel IPK seperti jenis kelamin, jalur masuk, kategori provinsi, kategori pulau, pekerjaan ayah, tahun lulus SMTA, kategori SMTA, kWh listrik, dan bidikmisi yang di mana variabel-variabel tersebut merupakan variabel *independent* pada penelitian ini. Pembagian jenis variabel pada penelitian ini dapat dilihat pada Tabel 4.

Pengujian *one-way* ANOVA dilakukan untuk mencari variabel yang signifikan. Variabel yang signifikan ditentukan jika hasil pengujian menyatakan bahwa menolak H_0 dan gagal menolak H_1 , landasan yang mendasari menerima H_0 atau H_1 ketika nilai *p-value* hasil pengujian lebih kecil dibandingkan taraf signifikansi menolak H_0 dan gagal menolak H_1 sedangkan jika hasil pengujian nilai *p-value* hasil pengujian lebih besar dibandingkan taraf signifikansi menyatakan menolak H_1 dan menerima H_0 (Fitrayuda, et al., 2020). Perlu diketahui bahwa H_0 menunjukkan bahwa hipotesis pada variabel *independent* tidak memiliki pengaruh

untuk variabel *dependent* sedangkan bahwa H_1 menunjukkan bahwa hipotesis pada variabel *independent* memiliki pengaruh untuk variabel *dependent*.

Hasil pengujian pada penelitian ini dengan melihat hipotesis dari perbedaan dari variabel *independent* tidak berpengaruh pada nilai IPK mahasiswa (H_0) dan variabel *independent* berpengaruh pada nilai IPK mahasiswa (H_1) serta penelitian ini juga menggunakan taraf signifikansi 0,05. Variabel-variabel yang menolak H_0 dan gagal menolak H_1 antara lain adalah jenis kelamin dengan nilai *p-value* 0, jalur masuk dengan nilai *p-value* 0, kategori provinsi dengan nilai *p-value* 0, kategori SMTA dengan nilai *p-value* 0,13, kWh listrik dengan nilai *p-value* 0,003, dan bidikmisi dengan nilai *p-value* 0. Variabel jenis kelamin, jalur masuk, kategori provinsi, kategori SMTA, kWh listrik, dan bidikmisi memiliki nilai *p-value* kurang dari taraf signifikansi yang digunakan pada penelitian ini. Hasil dari pengujian *one way ANOVA* dapat dilihat pada Tabel 5. Jadi variabel jenis kelamin, jalur masuk, kategori provinsi, kategori SMTA, kWh listrik, dan bidikmisi merupakan variabel-variabel signifikan yang berpengaruh terhadap nilai IPK mahasiswa UNTIRTA terkhususnya angkatan 2021/2022.

5.2 Analisis Jumlah Klasterisasi Optimal dan Karakteristiknya

Analisis jumlah klaster untuk mengetahui klaster optimal diperlukan validasi klaster. Validasi klaster adalah komponen yang penting untuk diperhitungkan dalam analisis klaster untuk mengetahui klaster yang terbaik (optimal) (Doi, et al., 2023). Analisis jumlah klaster memerlukan metode klasterisasi yang dapat digunakan untuk menghasilkan pengelompokan yang sesuai dengan data yang digunakan, oleh sebab itu dalam analisis jumlah klaster memerlukan validasi klaster. Setiap klaster yang dibuat memiliki ukuran yang berbeda untuk setiap karakteristiknya, seperti indeks validasi klaster (Brock, et al., 2008). Hasil klaster divalidasi menggunakan dua kriteria indeks validasi klaster yaitu, validasi internal dan validasi eksternal. Validasi internal melakukan evaluasi klaster dengan menggunakan informasi internal yang terkandung dalam data penelitian, sedangkan validasi eksternal melakukan perbandingan hasil analisis klaster dengan hasil yang didapatkan secara eksternal (Nerurkar, et al., 2019).

Analisis jumlah kluster optimal untuk validasi internal dapat dilihat pada Tabel 6. Penelitian dalam melakukan validasi internal menggunakan rata-rata *sillhoutte* dan *indeks dunn*. Rata-rata *sillhoutte* digunakan untuk menghitung derajat kepercayaan dalam proses klusterisasi pada suatu penelitian dengan kluster yang terbentuk dikatakan baik jika nilai rata-rata *sillhoutte* mendekati 1 sedangkan jika rata-rata *sillhoutte* mendekati -1 kluster yang terbentuk dikatakan kurang baik (Irwansyah & Faisal, 2015). *Indeks dunn* digunakan untuk mengukur rasio jarak terkecil antara observasi pada kluster yang berbeda dengan jarak terbesar pada setiap data kluster (Irwansyah & Faisal, 2015).

Analisis jumlah kluster optimal untuk validasi eksternal dapat dilihat pada Tabel 7. Penelitian dalam melakukan validasi eksternal menggunakan nilai *entropy*. Nilai *entropy* digunakan sebagai parameter untuk mengukur tingkat keseragaman (heterogenitas) dari kumpulan data. Nilai *entropy* yang semakin besar maka menunjukkan tingkat keragaman suatu kumpulan data semakin besar (Adinda, 2022). Jika nilai *entropy* semakin besar maka kluster yang terbentuk kurang baik karena menunjukkan data pada kluster tersebut berdistribusi beragam sedangkan jika nilai *entropy* semakin rendah menunjukkan data pada kluster tersebut memiliki distribusi data yang berpusat sehingga dapat dikatakan bahwa data kluster yang baik adalah data kluster yang memiliki nilai *entropy* yang rendah.

Penelitian ini menggunakan hipotesis kluster 2-6, oleh sebab itu perlu diketahui kluster yang optimal. Kluster yang optimal pada penelitian ini didapatkan 2 kluster. Penentuan 2 kluster ini dilihat dari validasi internal untuk nilai rata-rata *sillhoutte* dan *indeks dunn* pada kluster 2 memiliki nilai hasil yang mendekati 1 dibandingkan nilai yang dihasilkan oleh 3-6 kluster. Kluster 2 memiliki nilai rata-rata *sillhoutte* sebesar 0,267 dan nilai *indeks dunn* sebesar 0,000005. Penentuan 2 kluster ini juga ditentukan oleh validasi eksternal dengan menggunakan nilai *entropy*. Nilai *entropy* untuk pada 2 kluster memiliki nilai 0,692 yang lebih rendah dibandingkan kluster 3-6. Kluster optimal yang dipilih sesuai validasi internal dan validasi eksternal adalah pengelompokan 2 kluster.

Hasil penentuan 2 kluster mendapatkan karakteristik untuk 4096 mahasiswa yang dikelompok menjadi kelompok kluster 1 dan kelompok kluster 2.

Pengelompokan klaster ini membagi 2142 mahasiswa UNTIRTA berada pada kelompok klaster 1 sedangkan 1954 mahasiswa UNTIRTA berada pada kelompok klaster 2. Karakteristik hasil penelitian ini adalah nilai IPK, jenis kelamin, jalur masuk, kategori provinsi, kategori SMTA, kWh listrik dan bidikmisi.

Karakteristik mahasiswa pada pengelompokan klaster 1 berdasarkan variabel IPK diketahui bahwa minimal mahasiswa pada pengelompokan klaster 1 adalah 0,28 dengan maksimal IPK pada klaster 1 adalah 3,98 serta memiliki nilai media 3,590 dan rata-rata 3,422 dengan standar deviasi 0,340. Karakteristik mahasiswa pada pengelompokan klaster 2 berdasarkan variabel IPK diketahui bahwa minimal mahasiswa pada pengelompokan klaster 1 adalah 2,770 dengan maksimal IPK pada klaster 1 adalah 3,99 serta memiliki nilai median 3,74 dan rata-rata 3,701 dengan standar deviasi 0,031. Maka klaster 1 memiliki distribusi data yang lebih beragam daripada klaster 2 dilihat dari nilai standar deviasi yang lebih besar. Nilai rata-rata dan media pada klaster 2 lebih tinggi dibandingkan nilai klaster 1. Klaster 2 merupakan klaster yang memiliki peluang sukses dalam menjalankan studinya berdasarkan nilai IPK mahasiswa.

5.3 Rekomendasi Faktor-Faktor yang Penting untuk Dipertimbangkan dalam Seleksi Penerimaan Jalur Masuk SNBP

Rekomendasi yang dapat diberikan berdasarkan hasil pengolahan data dengan menggunakan *general linear model* (GLM) ditampilkan pada Tabel 11. *General linear model* digunakan untuk mendeskripsikan struktur dari variabel prediktor, sedangkan fungsi penghubung secara khusus mendeskripsikan hubungan antara model regresi dengan nilai ekspektasi dari variabel respon (Santi, et al., 2021). Penelitian menggunakan GLM untuk memberikan faktor-faktor yang penting untuk dipertimbangkan dalam seleksi penerimaan jalur masuk SNBP.

Penelitian ini berusaha mencari pengaruh yang signifikan terhadap hasil klaster yang telah dibuat. Hasil dari pengujian dengan metode GLM menggunakan variabel klaster sebagai variabel *dependent* sedangkan IPK, jenis kelamin, jalur masuk, kategori provinsi, kategori SMTA, kWh listrik dan bidikmisi digunakan sebagai variabel *independent*. Pengujian ini menggunakan taraf signifikansi 0,05 didapatkan hasil bahwa variabel yang memiliki pengaruh signifikan adalah kategori

SMTA (p -value= 0,042), kWh listrik dengan nilai sebesar (p -value= 0), bidikmisi dengan nilai sebesar (p -value= 0,016), dan IPK dengan nilai sebesar (p -value= 0). Hasil penelitian ini menunjukkan bahwa hal yang perlu diperhatikan dalam rekomendasi faktor-faktor yang sangat penting dalam penerimaan jalur masuk SNBP adalah kategori SMTA, kWh listrik, dan bidikmisi sedangkan IPK digunakan sebagai representasi kompetensi mahasiswa.

Faktor-faktor yang direkomendasikan untuk mendapatkan calon mahasiswa yang memiliki potensi pada studinya dilihat dari beberapa karakteristik yaitu, kategori SMTA, kWh listrik, dan bidikmisi. Faktor-faktor yang direkomendasikan berdasarkan karakteristik pada klaster 2 seperti pada kategori SMTA adalah penambahan kuota untuk calon mahasiswa yang berasal dari MAS. Faktor yang direkomendasikan selanjutnya adalah calon mahasiswa yang memiliki kWh Listrik terkecuali 900 watt, hal ini diharapkan untuk memprioritaskan penerimaan calon mahasiswa yang berasal dari kategori lainnya seperti seperti 450 watt, 1300 watt, 2200 watt, dan 2300 watt. Faktor yang direkomendasikan untuk calon mahasiswa yang memiliki bidikmisi adalah calon mahasiswa yang mendaftar yang memiliki Kartu Indonesia Pintar (KIP) dan memiliki prestasi.

BAB V I

KESIMPULAN DAN SARAN

6.1 Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan dalam menentukan kesimpulan berdasarkan tujuan penelitian adalah sebagai berikut.

1. Hasil penelitian memberikan hasil mengenai variabel-variabel yang signifikan terhadap variabel IPK yaitu, jenis kelamin, jalur masuk, kategori provinsi, kategori SMTA, kWh listrik, dan bidikmisi.
2. Hasil penelitian ini mendapatkan pengelompokan mahasiswa dengan 2 klaster optimal. Hasil dari pengelompokan klaster tersebut didapatkan bahwa karakteristik penelitian ini adalah nilai IPK, jenis kelamin, jalur masuk, kategori provinsi, kategori SMTA, kWh listrik dan bidikmisi. Karakteristik yang menjadi pembeda signifikan antara karakteristik klaster 1 dan klaster 2 adalah pada klaster 1 distribusi data beragam sedangkan pada klaster 2 distribusi data berpusat.
3. Hasil penelitian ini menunjukkan bahwa hal yang perlu diperhatikan dalam rekomendasi faktor-faktor yang sangat penting dalam penerimaan jalur masuk SNBP adalah kategori SMTA, kWh listrik, dan bidikmisi.

6.2 Saran

Berikut merupakan saran yang dapat peneliti sampaikan yang bertujuan untuk meningkatkan kualitas penelitian di masa yang akan datang.

1. Penelitian selanjutnya dapat dilakukan untuk klasterisasi setiap fakultas atau program studi sehingga dapat diketahui karakteristik klaster masing-masing fakultas dan jurusan yang ada di Universitas Sultan Ageng Tirtayasa.
2. Penelitian dapat dilanjutkan setiap angkatan untuk mengidentifikasi perbedaan pada setiap angkatan agar dapat diketahui parameter yang selalu muncul pada setiap angkatan.

3. Penelitian selanjutnya dapat dilakukan klasterisasi berdasarkan asal SMTA untuk menentukan asal sekolah dari calon mahasiswa yang berpeluang sukses menjalankan masa studinya dengan menggunakan metode *k-prototype* atau metode klasterisasi lainnya.



DAFTAR PUSTAKA

- Adinda, P. R., 2022. Pohon Keputusan C4. 5 Algoritma Untuk Klasifikasi Program Bantuan Belajar. *Jurnal Portal Data*, Volume 2, pp. 1-13.
- Andini, R. & Trianasari, N., 2021. Pengaruh *Brand Ambassador* dan *Brand Image* terhadap Proses Keputusan Pembelian pada *E-commerce* Tokopedia (Studi Kasus pada *Boy Group* BTS dan *Girl Group* Blackpink sebagai Brand Ambassador). s.l., *eProceedings of Management*.
- Aranda, J. & Natasya, W. A. G., 2016. Penerapan Metode *K-Means Cluster Analysis* pada Sistem Pendukung Keputusan Pemilihan Konsentrasi untuk Mahasiswa International Class STMIK AMIKOM Yogyakarta. *Semnasteknomedia Online*, 4(1), pp. 4-2.
- Azkiyah, M., 2017. Pengelompokan Kabupaten/Kota di Nusa Tenggara Timur Berdasarkan Pelayanan Kesehatan Ibu dan Anak Menggunakan *K-Means* dan *Fuzzy C-Means Cluster*, Surabaya: Institut Teknologi Sepuluh Nopember.
- Bolshakova, N. & Azuaje, F., 2003. *Improving Expression Data Mining Through Cluster Validation*. s.l., IEEE, pp. 19-22.
- Brock, G., Pihur, V., Datta, S. & Datta, S., 2008. *clValid: An R Package for Cluster Validation*. *Journal of Statistical Software*, Volume 25, pp. 1-22.
- Doi, M. D., Rusgiyono, A. & Wuryandari, T., 2023. Analisis *K-Medoids* dengan Validasi Indeks pada IPM Daerah 3T di Indonesia. *Jurnal Gaussian*, Volume 12, pp. 178-188.
- Edurank, 2023. *List Of 100 Best Universities In Indonesia*. [Online] Available at: <https://edurank.org/geo/id/> [Diakses 4 Desember 2023].
- Fitrayuda, A., Fajrin, J. & Anshari, B., 2020. Analisis Sifat Meknais Komposit Polyester Sisal Menggunakan Metode ANOVA. *Bina Ilmiah*, Volume 14, pp. 2817-2824.
- Hadjon, R. P., 2022. Analisis Webometric untuk Rekomendasi Peningkatan Peringkat Perguruan Tinggi pada Repositori Universitas Citra Bangsa.

- TIMOR CERDAS-Jurnal Teknologi Informasi, Manajemen Informasi dan Rekayasa Sistem Cerdas*, 1(1), pp. 1-8.
- Hamdanah, F. H. & Fitriana, D., 2021. Analisis Performansi Algoritma *Linear Regression* dengan *Generalized Linear Model* untuk Prediksi Penjualan pada Usaha Mikro, Kecil, dan Menengah. *Jurnal Nasional Pendidikan Teknik Informatika: JANAPATI*, Volume 10, pp. 23-32.
- Irwansyah, E. & Faisal, M., 2015. *Advanced Clustering: Teori dan Aplikasi*. s.l.:Deepublish.
- Jain, K. A. & Dubes, R. C., 1988. *Algorithms for Clustering Data. First Edition*. New Jersey: Prentice-Hall, Inc.
- J., Caraka, R. E. & Herliansyah, R., 2018. Aplikasi *Generalized Linear Model* pada R. Edisi Pertama. Yogyakarta: Innosain.
- Nerurkar, P., Pavate, A., Shah, M. & Samuel, J., 2019. *Performance of Internal Cluster Validations Measures for Evolutionary Clustering*. s.l., Springer Singapore.
- Palupi, R. & Prasetya, A. E., 2022. Pengaruh Implementasi *Content Management System* terhadap Kecepatan Kinerja Menggunakan *One Way Anova*. *Jurnal Ilmiah Informatika*, Volume 10, pp. 74-79.
- Qadrini, L., 2018. *Ensemble Fuzzy, K-Prototypes & Density Peaks Clustering Mixed* pada Pengelompokan Data Pelamar Bidikmisi Sejawat-Timur Tahun 2016-2017, Surabaya: Institut Teknologi Sepuluh Nopember.
- Santi, V. M., Wiyono, A. & Sudarwanto, 2021. Pemodelan Jumlah Kasus Malaria di Indonesia Menggunakan *Generalized Linear Model*. *Jurnal Statistika dan Aplikasinya*, 5(1), pp. 112-120.
- Steinbach, M., Karypis, G. & Kumar, V., 2000. *A Comparison of Document Clustering Techniques*. [Online] Available at: <https://conservancy.umn.edu/bitstream/handle/11299/215421/00-034.pdf?sequence=1&isAllowed=y>
- Subhan, A., Faqih, A. & Irawan, B., 2022. *Clustering Item Fast Moving dan Slow Moving* pada Produk Unilever Menggunakan Algoritma *K-Prototype*. *JATI (Jurnal Mahasiswa Teknik Informatika)*, pp. 629-634.

- Sudarsono, B. G., Leo, M. I., Santoso, A. & Hendrawan, F., 2021. Analisis Data Mining Data Netflix Menggunakan Aplikasi Rapid Miner. *Journal of Business and Audit Information Systems*, 4(1), pp. 13-21.
- Sulastri, S., Usman, L. & Syafitri, U. D., 2021. *K-Prototypes Algorithm for Clustering Schools Based on The Student Admission Data in IPB University. Indonesian Journal of Statistics and Its Applications*, pp. 228-242.
- Tan, P. N., Steinbach, M. & Kumar, V., 2005. *Introduction to Data Mining*. Boston: Addison-Wesley.
- Virtusena, V., Johar, A. & Wijanarko, A., 2021. Pengelompokan Potensi Kelulusan Mahasiswa Fakultas Teknik Unib Menggunakan *Algoritme K-Means* (Studi Kasus: Fakultas Teknik Universitas Bengkulu). *Rekursif: Jurnal Informatika*, pp. 206-225.
- Waruwu, M., 2023. Pendekatan Penelitian Pendidikan: Metode Penelitian Kualitatif, Metode Penelitian Kuantitatif dan Metode Penelitian Kombinasi (*Mixed Method*). *Jurnal Pendidikan Tambusai*, pp. 2896-2910.
- Wijayanti, R. & Saputro, D. R. S., 2023. *Clustering Data Campuran Numerik dan Kategorik*. s.l., s.n., pp. 702-706.



LAMPIRAN

Lampiran 1. Data Mahasiswa UNTIRTA 2021/2022

No	NIM	Jenis Kelamin	Jalur Masuk	Kategori Provinsi	Kategori Pulau	Pekerjaan Ayah	Tahun Lulus SMTA	Kategori SMTA	kWh Listrik	Bidik misi	IPK
1	1111210001	P	SNM PTN	Prop. Banten	Pulau Jawa	Karyawan Swasta	2021	SMAN	1300	Tidak	3.98
2	1111210002	P	SNM PTN	Prop. Banten	Pulau Jawa	Karyawan Swasta	2021	SMAN	900	Tidak	3.76
3	1111210003	P	SNM PTN	Prop. Banten	Pulau Jawa	Karyawan Swasta	2021	SMAS	2200	Tidak	3.92
4	1111210005	P	SNM PTN	Prop. Banten	Pulau Jawa	Pedagang/Wiraswasta	2021	MAN	450	Ya	3.92
...
4096	8884210039	P	SMM PTN	Prop. Banten	Luar Pulau Jawa	Petani/Nelayan	2021	SMAN	450	Tidak	2.92

**Data tidak dilampirkan semua karena bersifat rahasia*

Lampiran 2. Hasil One-Way ANOVA pada Variabel Jenis Kelamin dengan Variabel IPK

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Jenis Kelamin	1	64.83	64.8331	330.68	0.000
Error	4094	802.67	0.1961		
Total	4095	867.50			

Lampiran 3. Hasil One-Way ANOVA pada Variabel Jalur Masuk dengan Variabel IPK

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Jalur Masuk	2	13.37	6.6873	32.05	0.000
Error	4093	854.13	0.2087		
Total	4095	867.50			

Lampiran 4. Hasil *One-Way* ANOVA pada Variabel Kategori Provinsi dengan Variabel IPK

Analysis of Variance						
Source	DF	Adj SS	Adj MS	F-Value	P-Value	
Kategori Provinsi	1	3.281	3.2812	15.54	0.000	
Error	4094	864.221	0.2111			
Total	4095	867.502				

Lampiran 5. Hasil *One-Way* ANOVA pada Variabel Kategori Pulau dengan Variabel IPK

Analysis of Variance						
Source	DF	Adj SS	Adj MS	F-Value	P-Value	
Kategori Pulau	1	0.080	0.07960	0.38	0.540	
Error	4094	867.422	0.21188			
Total	4095	867.502				

Lampiran 6. Hasil *One-Way* ANOVA pada Variabel Pekerjaan Ayah dengan Variabel IPK

Analysis of Variance						
Source	DF	Adj SS	Adj MS	F-Value	P-Value	
Pekerjaan Ayah	16	5.497	0.3436	1.63	0.054	
Error	4079	862.004	0.2113			
Total	4095	867.502				

Lampiran 7. Hasil *One-Way* ANOVA pada Variabel Tahun Lulus SMTA dengan Variabel IPK

Analysis of Variance						
Source	DF	Adj SS	Adj MS	F-Value	P-Value	
Tahun Lulus SMTA	2	0.429	0.2144	1.01	0.364	
Error	4093	867.073	0.2118			
Total	4095	867.502				

Lampiran 8. Hasil *One-Way* ANOVA pada Variabel kWh Listrik dengan Variabel IPK

Analysis of Variance					
Source	DF	Adj SS	Adj MS	F-Value	P-Value
Kwh Listrik	4	3.385	0.8463	4.01	0.003
Error	4091	864.117	0.2112		
Total	4095	867.502			

Lampiran 9. Hasil *One-Way* ANOVA pada Variabel Bidikmisi dengan Variabel IPK

Analysis of Variance					
Source	DF	Adj SS	Adj MS	F-Value	P-Value
Bidikmisi	1	5.927	5.9269	28.16	0.000
Error	4094	861.575	0.2104		
Total	4095	867.502			

Lampiran 10. Hasil *One-Way* ANOVA pada Variabel Kategori SMTA dengan Variabel IPK

Analysis of Variance					
Source	DF	Adj SS	Adj MS	F-Value	P-Value
Kategori SMTA	6	3.421	0.5702	2.70	0.013
Error	4089	864.081	0.2113		
Total	4095	867.502			

Lampiran 11. Syntax K-Prototype

```

#====input data====
data_awal<-read.csv("DATA VARIABEL SIGNIFIKAN ANOVA.csv", header = T, sep = ",")

# Simpan nama baris sebelum menghapus kolom "NIM"
nama_baris <- data_awal$NIM

# Menghapus kolom "NIM"
data_klaster <- subset(data_awal, select = -NIM)

# Konversi data menjadi data frame dan tetapkan nama baris
data_klaster <- as.data.frame(data_klaster)
rownames(data_klaster) <- nama_baris

# Menampilkan data dengan perubahan
print(data_klaster)

# Menentukan jenis variabel
data_klaster$Jenis_kelamin <- as.factor(data_klaster$Jenis_kelamin)
data_klaster$Jalur_Masuk <- as.factor(data_klaster$Jalur_Masuk)
data_klaster$Kategori_Provinsi <- as.factor(data_klaster$Kategori_Provinsi)
data_klaster$Kategori_SMTA <- as.factor(data_klaster$Kategori_SMTA)
data_klaster$KWH_Listrik <- as.factor(data_klaster$KWH_Listrik)
data_klaster$Bidikmisi <- as.factor(data_klaster$Bidikmisi)

# Cek tipe data
str(data_klaster)

#====ALGORITMA K-PROTOTYPES====
library(clustMixType)
library(data.table)
library(dplyr)

# Menentukan pemilihan alternatif
nilai.seed <- sample(1:9999,1)
set.seed(nilai.seed)

```

Lampiran 11. Syntax K-Prototype (Lanjutan)

```

# Menentukan pemilihan alternatif
nilai.seed <- sample(1:9999,1)
set.seed(nilai.seed)

# Menghitung jarak euclidean
jarak_euclidean <- dist(data_klaster, method = "euclidean")

# Menentukan klaster
#2 klaster
dua_klaster <- kproto(data_klaster, k=2)
#3 klaster
tiga_klaster <- kproto(data_klaster, k=3)
#4 klaster
empat_klaster <- kproto(data_klaster, k=4)
#5 klaster
lima_klaster <- kproto(data_klaster, k=5)
#6 klaster
enam_klaster <- kproto(data_klaster, k=6)

hasil_klaster <- cbind(data_awal,
                      "dua klaster" = dua_klaster$cluster,
                      "tiga klaster" = tiga_klaster$cluster,
                      "empat klaster" = empat_klaster$cluster,
                      "lima klaster" = lima_klaster$cluster,
                      "enam klaster" = enam_klaster$cluster)

# statistik setiap klaster
library(fpc)
stat_dua_klaster <- cluster.stats(jarak_euclidean, dua_klaster$cluster)
stat_tiga_klaster <- cluster.stats(jarak_euclidean, tiga_klaster$cluster)
stat_empat_klaster <- cluster.stats(jarak_euclidean, empat_klaster$cluster)
stat_lima_klaster <- cluster.stats(jarak_euclidean, lima_klaster$cluster)
stat_enam_klaster <- cluster.stats(jarak_euclidean, enam_klaster$cluster)

```

Lampiran 11. *Syntax K-Prototype (Lanjutan)*

```

====VISUALISASI====
library(factoextra)
library(cluster)
library(NbClust)

#2 klaster
sil_dua_klaster <- silhouette(dua_klaster$cluster, jarak_euclidean)
grafik_dua_klaster <- fviz_silhouette(sil_dua_klaster)

#3 klaster
sil_tiga_klaster <- silhouette(tiga_klaster$cluster, jarak_euclidean)
grafik_tiga_klaster <- fviz_silhouette(sil_tiga_klaster)

#4 klaster
sil_empat_klaster <- silhouette(empat_klaster$cluster, jarak_euclidean)
grafik_empat_klaster <- fviz_silhouette(sil_empat_klaster)

#5 klaster
sil_lima_klaster <- silhouette(lima_klaster$cluster, jarak_euclidean)
grafik_lima_klaster <- fviz_silhouette(sil_lima_klaster)

#6 klaster
sil_enam_klaster <- silhouette(enam_klaster$cluster, jarak_euclidean)
grafik_enam_klaster <- fviz_silhouette(sil_enam_klaster)

====Validasi Internal====

#average silhouette width
silhouette_avg <- cbind.data.frame("2 klaster" = stat_dua_klaster$avg.silwidth,
                                   "3 klaster" = stat_tiga_klaster$avg.silwidth,
                                   "4 klaster" = stat_empat_klaster$avg.silwidth,
                                   "5 klaster" = stat_lima_klaster$avg.silwidth,
                                   "6 klaster" = stat_enam_klaster$avg.silwidth)
rownames(silhouette_avg) <- "Rata-rata Koefisien Silhouette"

```

Lampiran 11. *Syntax K-Prototype (Lanjutan)*

```

#indeks Dunn
indeks_dunn <- cbind.data.frame("2 klaster" = stat_dua_klaster$dunn,
                                "3 klaster" = stat_tiga_klaster$dunn,
                                "4 klaster" = stat_empat_klaster$dunn,
                                "5 klaster" = stat_lima_klaster$dunn,
                                "6 klaster" = stat_enam_klaster$dunn)
rownames(indeks_dunn) <- "Indeks Dunn"

#resume validasi internal
validasi_internal <- rbind.data.frame(silhouette_avg, indeks_dunn)

====validasi eksternal====
entropy_klaster <- cbind.data.frame("2 klaster" = stat_dua_klaster$entropy,
                                    "3 klaster" = stat_tiga_klaster$entropy,
                                    "4 klaster" = stat_empat_klaster$entropy,
                                    "5 klaster" = stat_lima_klaster$entropy,
                                    "6 klaster" = stat_enam_klaster$entropy)

write.csv(validasi_internal, "validasi internal fix.csv")
write.csv(entropy_klaster, "validasi eksternal fix.csv")
write.csv(hasil_klaster, "hasil klaster fix.csv")

getwd()

```


Lampiran 12. Hasil Klaster

No	NIM	Klaster				
		2	3	4	5	6
1	1111210001	2	2	1	2	6
2	1111210002	1	3	4	2	6
3	1111210003	2	2	4	2	6
4	1111210005	2	2	3	4	6
...
4096	8884210039	2	2	2	3	5

Lampiran 13. Hasil *General Linear Model*

Analysis of Variance						
Source	DF	Adj SS	Adj MS	F-Value	P-Value	
Jenis Kelamin	1	0.00	0.001	0.01	0.907	
Jalur Masuk	2	0.12	0.058	1.19	0.304	
Kategori Provinsi	1	0.03	0.034	0.71	0.401	
Kategori SMTA	6	0.64	0.106	2.18	0.042	
Kwh Listrik	4	650.89	162.722	3348.76	0.000	
Bidikmisi	1	0.28	0.281	5.78	0.016	
IPK	239	157.18	0.658	13.53	0.000	
Error	3841	186.64	0.049			
Lack-of-Fit	2611	186.64	0.071		*	*
Pure Error	1230	0.00	0.000			
Total	4095	1021.84				

DAFTAR RIWAYAT HIDUP PENULIS

Nama : Findi Choerun Annisa
NIM : 3333200111
Tempat/Tanggal Lahir : Kab. Pandeglang / 26 September 2001
Jenis Kelamin : Perempuan
Agama : Islam
Alamat Email : findichoerunannisa@gmail.com
LinkedIn : www.linkedin.com/in/findica

Riwayat Pendidikan

Sekolah Dasar : SDN 3 PANDEGLANG
SLTP : SMPN 1 PANDEGLANG
SLTA : SMAN 1 PANDEGLANG

Riwayat Organisasi

1. Zetizen Banten (2018-2023) - Anggota
2. Bilik Urang (2021) - Divisi Multimedia
3. Mading & Karya Tulis Ilmiah (2016-2018) -Ketua
4. OSIS SMAN 1 Pandeglang (2017-2018) – Anggota Divisi Kewirausahaan
5. OSIS SMPN 1 Pandeglang (2013-2014) – Koordinator Divisi Jurnalistik
6. OSIS SMPN 1 Pandeglang (2012-2013) – Anggota Divisi Jurnalistik

Prestasi

1. Zetizen Face Literasi 2023
2. Juara III Lomba Karya Tulis “JURNALISTIK” Tingkat SLTA/MA Sederajat & Perguruan Tinggi Provinsi Banten Tahun 2017
3. Juara I Lomba Karya Tulis “JURNALISTIK” Tingkat SLTA/MA Sederajat & Perguruan Tinggi Kab. Pandeglang Tahun 2016