

**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*

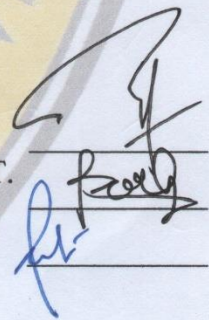
Telah berhasil dipertahankan di hadapan Dewan Penguji dan Diterima  
sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar  
Sarjana Teknik pada Jurusan Teknik Industri, Fakultas Teknik,  
Universitas Sultan Ageng Tirtayasa

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

	<b>Halaman</b>
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
<b>BAB I PENDAHULUAN.....</b>	<b>1</b>
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
<b>BAB II TINJAUAN PUSTAKA.....</b>	<b>7</b>
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
<b>BAB III METODE PENELITIAN.....</b>		<b>15</b>
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
<b>BAB IV HASIL PENELITIAN.....</b>		<b>23</b>
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</i> ANOVA.....	27
4.2.2.1	Menentukan Jenis Variabel Pengujian pada <i>One-Way</i> ANOVA.....	27
4.2.2.2	Hasil Pengujian Data <i>One-Way</i> ANOVA.....	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

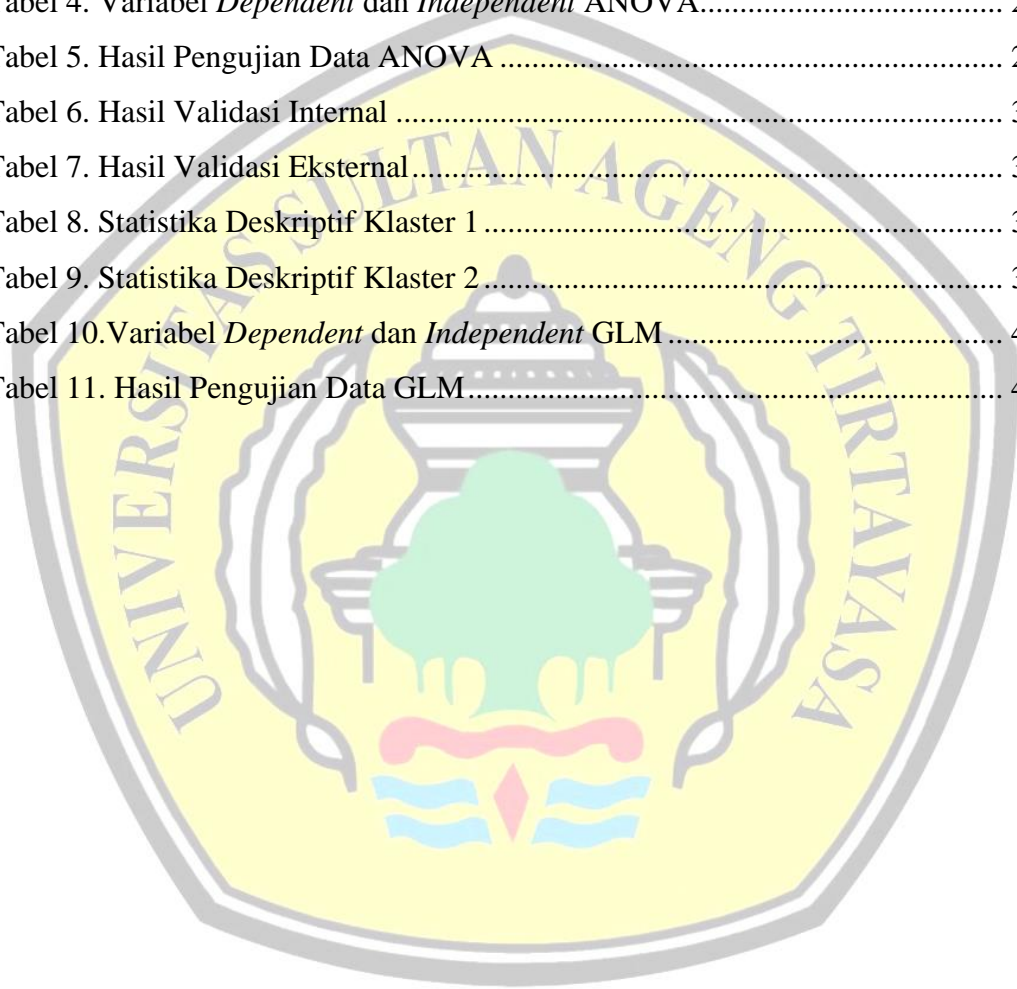
4.2.9.2	Hasil Pengujian Data <i>General Linear Model</i> (GLM).....	47
<b>BAB V ANALISIS DAN PEMBAHASAN.....</b>		<b>49</b>
5.1	Analisis Variabel-Variabel yang Signifikan terhadap Variabel IPK .....	49
5.2	Analisis Jumlah Klasterisasi Optimal dan Karakteristiknya.....	50
5.3	Rekomendasi Faktor-Faktor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





## 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

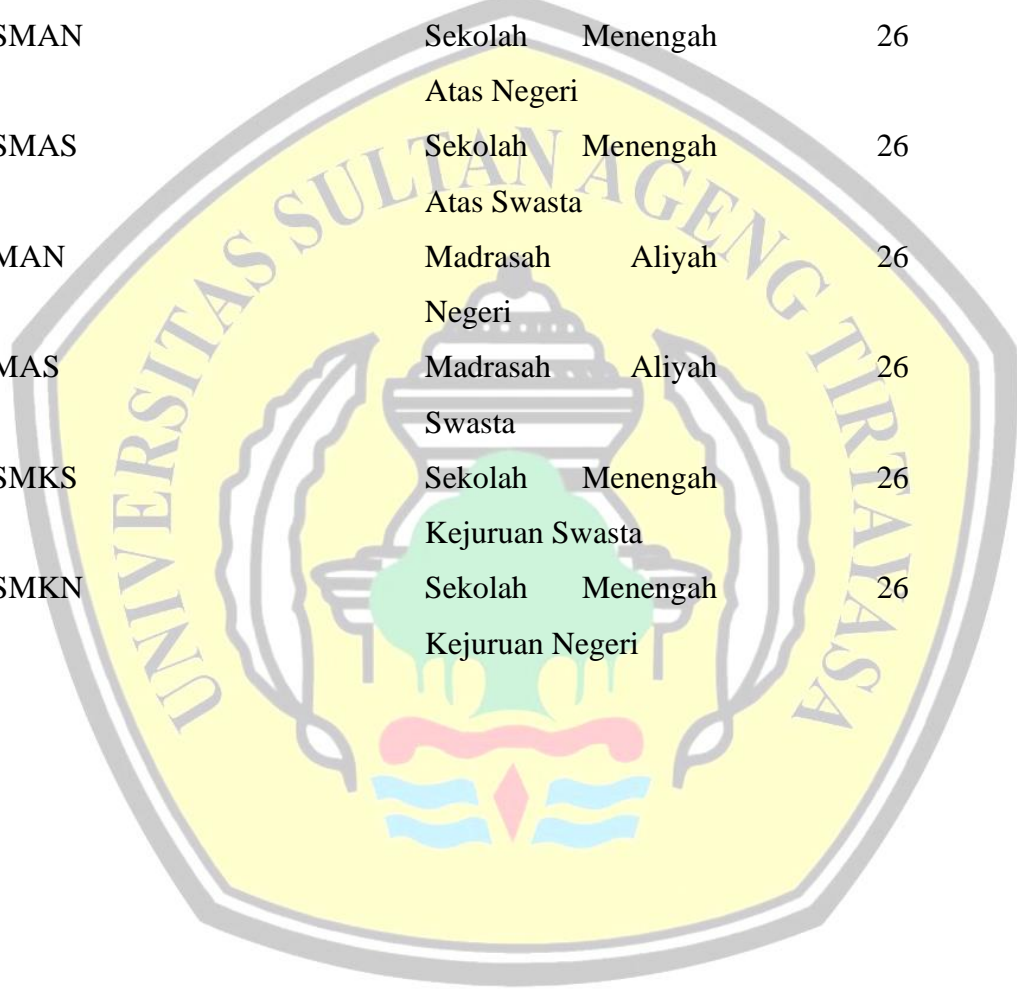
## DAFTAR ARTI LAMBANG, SINGKATAN, DAN ISTILAH

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.

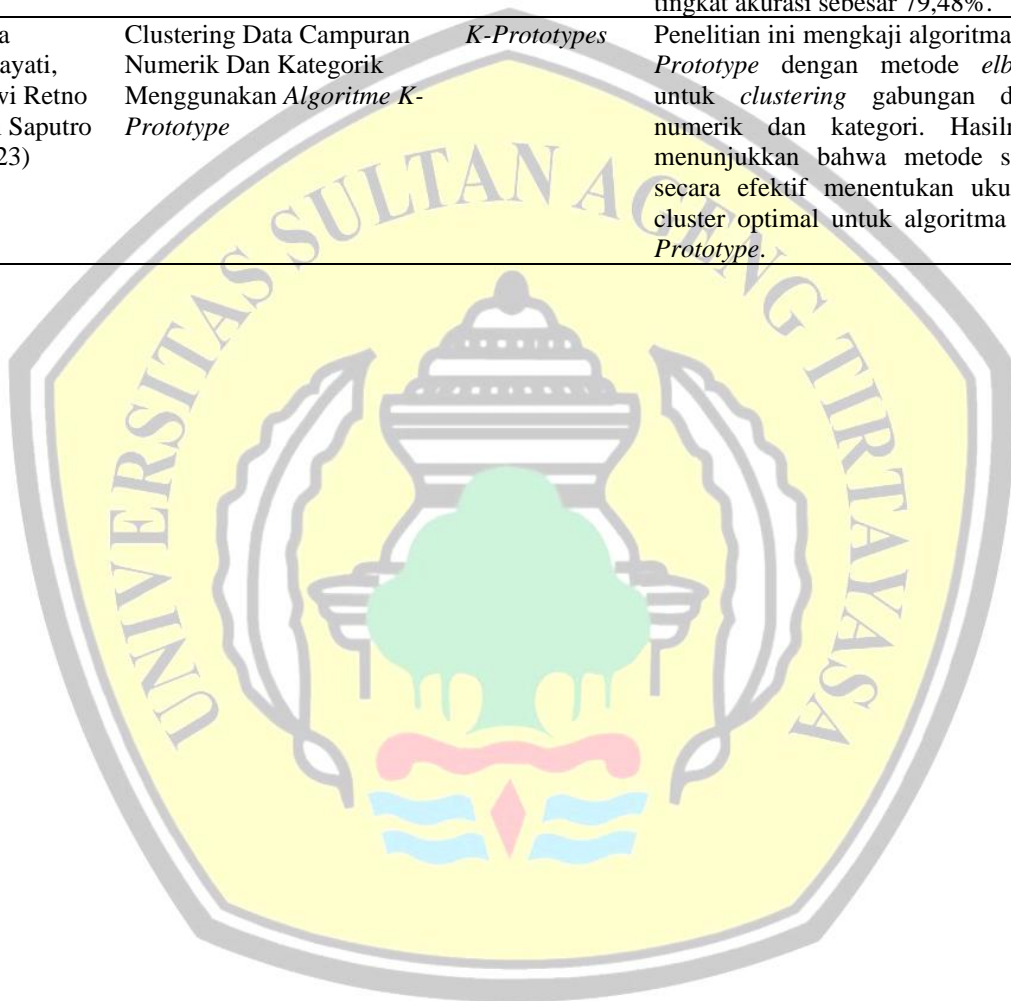
#### **1.6 Penelitian Terdahulu**

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 Sulastrri, 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 <i>K-Means</i> 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> .



## 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.