

**CLUSTERING STATUS EKONOMI MAHASISWA
BARU MENGGUNAKAN METODE *K-MEANS***

(STUDI KASUS UNIVERSITAS ICHSAN GORONTALO)

Oleh

ANA FITRIANA

T3119077

SKRIPSI

Untuk memenuhi salah satu syarat ujian

Guna memperoleh gelar Sarjana



PROGRAM SARJANA

TEKNIK INFORMATIKA

UNIVERSITAS ICHSAN GORONTALO

GORONTALO

2023

PERSETUJUAN SKRIPSI

CLUSTERING STATUS EKONOMI MAHASISWA BARU MENGGUNAKAN METODE *K-MEANS* (STUDI KASUS UNIVERSITAS ICHSAN GORONTALO)

Oleh

ANA FITRIANA

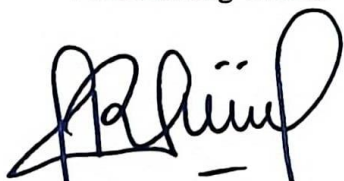
T3119077

SKRIPSI

Untuk memenuhi salah satu syarat ujian
guna untuk memperoleh gelar Sarjana
Program Studi Teknik Informatika,
ini telah disetujui oleh Tim Pembimbing.

Gorontalo, November 2023

Pembimbing Utama



Rezqiwati Ishak, M.Kom
NIDN: 0903087901

Pembimbing Pendamping



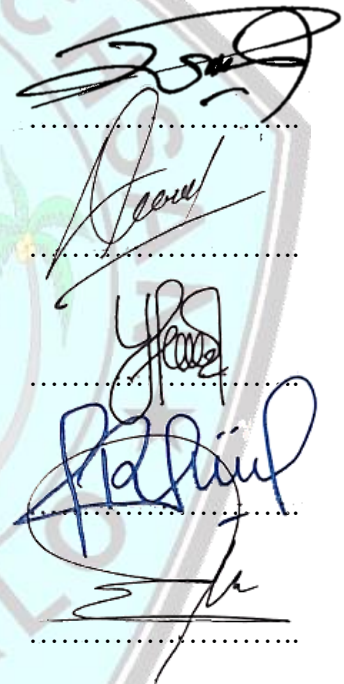
Sunarto Taliki, M.Kom
NIDN: 0906058301

PENGESAHAN SKRIPSI
CLUSTERING STATUS EKONOMI MAHASISWA
BARU MENGGUNAKAN METODE K-MEANS
(STUDI KASUS UNIVERSITAS ICHSAN GORONTALO)

Oleh
ANA FITRIANA
T3119077

Diperiksa oleh Panitia Ujian Strata Satu (S1)
Universitas Ichsan Gorontalo

1. Ketua Penguji
Irvan Abraham Salihi, M.Kom
2. Anggota
Husdi, M.Kom
3. Anggota
Yulianty Lasena, M.Kom
4. Anggota
Rezqiwati Ishak, M.Kom
5. Anggota
Sunarto Taliki, M.Kom



Mengetahui,

Dekan Fakultas Ilmu Komputer



Irvan Abraham Salihi, M.Kom
NIDN: 0928028101

Ketua Program Studi



Sudirman S. Panna, M.Kom
NIDN: 0924038205

PERNYATAAN SKRIPSI

Dengan ini saya menyatakan bahwa:

1. Karya tulis (Skripsi) saya ini adalah asli dan belum pernah diajukan untuk mendapatkan gelar akademik (Sarjana) baik di Universitas Ichsan Gorontalo maupun di perguruan tinggi lainnya.
2. Karya tulis (Skripsi) saya ini adalah murni gagasan, rumusan, dan peniliti saya sendiri, tanpa bantuan pihak lain, kecuali arahan dari Tim Pembimbing.
3. Dalam karya tulis (Skripsi) saya ini tidak terdapat karya atau pendapat yang telah dipublikasikan orang lain, kecuali secara tertulis dicantumkan sebagai acuan/sitasi dalam naskah dan dicantumkan pula dalam daftar pustaka.
4. Pernyataan ini saya buat dengan sesungguhnya dan apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka saya bersedia menerima sanksi akademik berupa pencabutan gelar yang telah diperoleh karena karya tulis ini, serta sanksi lainnya sesuai dengan norma-norma yang berlaku di Universitas Ichsan Gorontalo.

Gorontalo, November 2023

Yang Membuat Pernyataan,

Materai

10000

Ana Fitriana

ABSTRACT

ANA FITRIANA. T3119077. THE ECONOMIC STATUS CLUSTERING OF NEW STUDENTS USING THE K-MEANS METHOD (A CASE STUDY AT ICHSAN GORONTALO UNIVERSITY)

Students' economic status is one of the factors influencing success in studies. Therefore, it is important to determine students' economic status to provide appropriate support. One way to determine the students' economic status is by clustering. Clustering is the process of grouping data into homogeneous groups. The K-Means method is one of the most popular clustering methods. This method's work divides data into K groups with the minimum possible distance between group members. This research aims to apply the K-Means method in classifying students' economic status. The K-Means method forms clusters based on the economic status characteristics. The data used in this research are student data of 2022 with a total of 812 students. The data consists of attributes such as parents' occupation, income, and number of dependents. The analysis results show that there are four optimal clusters produced. Each cluster reflects a different level of economic status. The clustering analysis results of each cluster are namely, Cluster 1 with a low economic status level, Cluster 2 with a high economic status level, Cluster 3 with an average economic status level, and Cluster 4 with a medium economic status level. Apart from that, this research also considers evaluating the performance of the clustering model using internal evaluation metrics such as the Davies-Bouldin Index (DBI) to validate the clustering results with a DBI value of 0.80.

Keywords: *Clustering, Student Economic Status, K-Means, Elbow Method, Davies-Bouldin Index*

ABSTRAK

ANA FITRIANA. T3119077. CLUSTERING STATUS EKONOMI MAHASISWA BARU MENGGUNAKAN METODE K-MEANS (STUDI KASUS UNIVERSITAS ICHSAN GORONTALO)

Status ekonomi mahasiswa merupakan salah satu faktor yang dapat mempengaruhi keberhasilan mereka dalam studi. Oleh karena itu, penting untuk mengetahui status ekonomi mahasiswa agar dapat memberikan dukungan yang tepat. Salah satu cara untuk mengetahui status ekonomi mahasiswa adalah dengan menggunakan clustering. *Clustering* adalah proses pengelompokan data ke dalam kelompok-kelompok yang homogen. Metode *K-Means* adalah salah satu metode *clustering* yang paling populer. Metode ini bekerja dengan membagi data ke dalam *K* kelompok dengan jarak antar anggota kelompok yang seminimal mungkin. Penelitian ini bertujuan untuk menerapkan metode *K-Means* dalam mengelompokkan status ekonomi mahasiswa. Metode *K-Means* digunakan untuk membentuk *cluster* berdasarkan karakteristik status ekonomi. Data yang digunakan dalam penelitian ini adalah data mahasiswa pada tahun 2022 dengan jumlah 812 mahasiswa. Data tersebut terdiri dari atribut-atribut seperti pekerjaan orang tua, penghasilan orang tua, dan jumlah tanggungan orang tua. Hasil analisis menunjukkan bahwa terdapat 4 *cluster* optimal yang dihasilkan. Setiap *cluster* mencerminkan tingkat status ekonomi yang berbeda. Hasil analisa *clustering* untuk masing-masing *cluster* yaitu *cluster* 1 merupakan tingkat status ekonomi yang rendah, *cluster* 2 kelompok mahasiswa yang tingkat ekonomi tinggi, *cluster* 3 kelompok mahasiswa yang tingkat status ekonomi sedang, dan *cluster* 4 merupakan kelompok mahasiswa tingkat status ekonomi menengah. Selain itu, penelitian ini juga mempertimbangkan evaluasi kinerja model *clustering* menggunakan metrik evaluasi internal seperti *Davies-Bouldin Index* (DBI) untuk memvalidasi hasil *clustering* dengan nilai DBI sebesar 0.80.

Kata kunci: *Clustering*, Status Ekonomi Mahasiswa, *K-Means*, Metode *Elbow*, *Davies-Bouldin Index*

KATA PENGANTAR



Puji syukur penulis panjatkan kehadirat Allah Subhanahu Wa Ta'ala, karena hanya dengan izin dan kuasa-Nya penulisan dapat menyelesaikan penyusunan skripsi ini dengan judul: **“Clustering Status Ekonomi Mahasiswa Baru Menggunakan Metode K-Means”**, sebagai salah satu syarat Ujian Akhir guna memperoleh gelar Sarjana Komputer pada Program Studi Teknik Informatika Fakultas Ilmu Komputer Universitas Ichsan Gorontalo.

Penulis menyadari sepenuhnya bahwa skripsi ini tidak mungkin terwujud tanpa bantuan dan dorongan dari berbagai pihak, baik bantuan moril maupun materil. Untuk itu, dengan segala keikhlasan dan kerendahan hati, penulis banyak mengucapkan banyak terima kasih dan penghargaan setinggi-tingginya kepada:

1. Ibu Dr. Dra. Juriko Abdussamad, M.Si, selaku Ketua Yayasan Pengembangan Ilmu Pengetahuan dan Teknologi (YPIPT) Ichsan Gorontalo;
2. Bapak Dr. Abdul Gaffar La Tjokke, M.Si, selaku Rektor Universitas Ichsan Gorontalo;
3. Bapak Irvan Abraham Salihi, S.Kom, M.Kom, selaku Dekan Fakultas Ilmu Komputer Universitas Ichsan Gorontalo;
4. Bapak Sudirman Melangi, M.Kom, selaku Wakil Dekan I Bidang Akademik Fakultas Ilmu Komputer;
5. Ibu Irma Surya Kumala, M.Kom, selaku Wakil Dekan II bidang Administrasi Umum dan Keuangan Fakultas Ilmu Komputer;
6. Bapak Sudirman S. Panna, M.Kom, selaku Ketua Jurusan Teknik Informatika Fakultas ilmu Komputer Universitas Ichsan Gorontalo;
7. Ibu Rezqiwati Ishak, M.Kom, selaku Pembimbing I yang telah banyak membimbing penulis selama ini;
8. Bapak Sunarto Taliki, M.Kom, selaku Pembimbing II yang telah banyak membimbing penulis selama ini;
9. Bapak dan Ibu Dosen Universitas Ichsan Gorontalo yang sudah mendidik dan

mengajarkan berbagai ilmu disiplin kepada penulis;

10. Terima kasih kepada kedua Orang Tua saya yang selalu memberikan dukungan, membiayai hidup dan semangat hidup bagi penulis;
11. Terima kasih kepada Nenek dan Kakek saya yang sudah merawat, membesarkan dan menyayangi saya sepenuh hati;
12. Terima kasih kepada seluruh keluarga saya yang selalu memberikan dukungan;
13. Terima kasih kepada seluruh teman-teman yang telah membantu, memberikan dukungan dan semangat bagi penulis;
14. Terima kasih kepada bias-bias saya BTS, TXT, Enhypen, Watermelon Sugar dan masih banyak lagi, yang selalu menemani dan menghibur penulis saat pengerjaan skripsi ini;
15. Terima kasih kepada semua pihak yang membantu dalam menyelesaikan skripsi ini yang tak sempat penulis sebutkan satu-persatu.

Semoga Allah SWT, melimpahkan balasan atas jasa-jasa mereka kepada kami. Penulis menyadari sepenuhnya bahwa apa yang telah dicapai ini masih jauh dari kesempurnaan dan masih banyak terdapat kekurangan. Oleh karena itu, penulis sangat mengharapkan adanya kritik dan saran yang konstruktif. Akhirnya penulis berharap semoga hasil yang telah dicapai ini dapat bermanfaat bagi kita semua, Aamiin.

Gorontalo, November 2023

Penulis

DAFTAR ISI

PERSETUJUAN SKRIPSI	ii
PENGESAHAN SKRIPSI	iii
PERNYATAAN SKRIPSI	iv
ABSTRACT	v
ABSTRAK	vi
KATA PENGANTAR	vii
DAFTAR ISI	ix
DAFTAR TABEL	xii
DAFTAR GAMBAR	xiii
BAB I PENDAHULUAN	1
1.1 Latar Belakang.....	1
1.2 Identifikasi Masalah	3
1.3 Rumusan Masalah	3
1.4 Tujuan Penelitian.....	4
1.5 Manfaat Penelitian.....	4
BAB II LANDASAN TEORI	5
2.1 Tinjauan Studi	5
2.2 Tinjauan Pustaka	7
2.2.1 Status Ekonomi	7
2.2.2 Data Mining	8
2.2.3 Tahapan Data Mining.....	8
2.2.4 Clustering	10
2.2.5 Algoritma K-Means	10

2.2.6	Metode <i>Elbow</i>	11
2.2.7	<i>Davies-Bouldin Index</i> (DBI)	12
2.2.8	Contoh Penerapan Algoritma <i>K-Means</i>	12
2.2.9	Python	19
2.3	Perangkat Lunak.....	20
2.4	Kerangka Pikir.....	21
BAB III METODE PENELITIAN		22
3.1	Jenis, Metode, Subjek, Objek, Waktu dan Lokasi Penelitian.....	22
3.2	Pengumpulan Data.....	22
3.3	Pemodelan	23
BAB IV HASIL PENELITIAN		24
4.1	Hasil Pengumpulan Data	24
4.2	Tahapan <i>Clustering K-Means</i>	25
4.3	Hasil Pengembangan Model.....	29
4.3.1	Seleksi Data.....	29
4.3.2	Transformasi Data.....	30
4.3.3	Pengoptimalan.....	32
4.3.4	Pusat <i>Cluster</i> Dan Jumlah Data.....	36
BAB V PEMBAHASAN PENELITIAN		39
5.1	Pembahasan Kinerja Model.....	39
5.2	Pembahasan Hasil Pemodelan.....	41
5.2.1	Hasil Pemodelan Sebelum Optimalisasi	41
5.2.2	Hasil Pemodelan Sesudah Optimalisasi	42
BAB VI KESIMPULAN DAN SARAN.....		46
6.1	Kesimpulan.....	46

6.2	Saran.....	46
DAFTAR PUSTAKA		47
LAMPIRAN		

DAFTAR TABEL

Tabel 1.1 Dataset Mahasiswa	2
Tabel 2.1 Penelitian Tentang <i>Clustering</i> Menggunakan Metode <i>K-Means</i>	5
Tabel 2.2 Sampel Dataset Nilai Kalkulus.....	13
Tabel 2.3 Dataset Nilai Matematika Diskrit.....	13
Tabel 2.4 Sampel Hasil Penggabungan Dataset	14
Tabel 2.5 Sampel Hasil Nilai Huruf ke Angka	15
Tabel 2.6 <i>Centroid</i> Dan Jumlah Data	18
Tabel 2.7 <i>Sample</i> Hasil <i>Clustering</i>	18
Tabel 2.8 Perangkat Lunak Pendukung.....	20
Tabel 3.1 Variabel/Atribut Data	22
Tabel 4.1 Hasil Pengumpulan Dataset.....	24
Tabel 4.2 Dataset Hasil Olahan.....	25
Tabel 4.3 Konversi Nilai Pekerjaan Orang Tua	26
Tabel 4.4 Hasil Konversi Atribut Pekerjaan.....	26
Tabel 4.5 Dataset Hasil Olahan.....	27
Tabel 5.1 Hasil Uji <i>Cluster Davies Bouldin Index</i> (DBI)	40
Tabel 5.2 Hasil Uji <i>Cluster Davies Bouldin Index</i> (DBI)	41
Tabel 5.3 Hasil <i>Clustering</i> sesudah Optimalisasi.....	43
Tabel 5.4 Hasil <i>Clustering</i> Status Ekonomi Mahasiswa	44

DAFTAR GAMBAR

Gambar 2.1 Proses <i>Knowledge in Database</i> (KDD)	9
Gambar 2.2 Proses Algoritma K-Means.....	11
Gambar 2.3 Contoh Visualisasi Pemograman Metode <i>Elbow</i>	12
Gambar 2.4 Visualisasi Hasil Metode <i>Elbow</i>	16
Gambar 2.5 Visualisasi Hasil <i>Cluster</i>	17
Gambar 2.6 Informasi Hasil <i>Cluster</i>	17
Gambar 2.7 Bagan Kerangka Pikir.....	21
Gambar 4.1 Visualisasi Teknik <i>Elbow</i>	28
Gambar 5.1 Visualisasi Teknik <i>Elbow</i> sebelum di normalisasikan.....	39
Gambar 5.2 Visualisasi Teknik <i>Elbow</i> setelah di normalisasikan.....	40
Gambar 5.3 Visualisasi Hasil <i>Clustering</i> sebelum optimalisasi.....	41
Gambar 5.4 Pusat <i>Cluster</i> Sebelum Optimalisasi.....	42
Gambar 5.5 Visualisasi Hasil <i>Clustering</i> Sesudah Optimalisasi.....	43

BAB I

PENDAHULUAN

1.1 Latar Belakang

Status ekonomi mahasiswa merupakan salah satu faktor yang dapat mempengaruhi proses pembelajaran dan keberhasilan studi pada mahasiswa. Oleh karena itu, penting untuk memahami status ekonomi mahasiswa agar dapat memberikan dukungan yang tepat bagi mereka[1].

Di era *digital* saat ini, data mengenai mahasiswa, termasuk informasi tentang status ekonomi mereka, mudah diakses dan dikelola. Penggunaan metode analisis data seperti clustering dapat membantu mengidentifikasi pola dalam data ini, memungkinkan pembuat kebijakan dan lembaga pendidikan untuk menyusun strategi yang tepat untuk mendukung mahasiswa dengan latar belakang ekonomi yang berbeda[2].

Salah satu metode *clustering* yang umum digunakan adalah algoritma K-means. Metode ini memungkinkan pengelompokan data berdasarkan kesamaan karakteristik[3], yang dalam konteks ini adalah status ekonomi mahasiswa. Dengan mengelompokkan mahasiswa ke dalam kelompok-kelompok berdasarkan status ekonomi mereka.

Penggunaan metode *K-Means* pada penelitian ini sebab *K-Means* merupakan proses pengelompokan data (*cluster*) berasal dari kumpulan data yang tidak diketahui kelompok-kelompok atau kelas-kelasnya serta proses memilih data-data termasuk ke pada *cluster* yang mana. *Clustering* adalah proses pembagian data kedalam kelas atau *cluster* berdasarkan tingkat kesamaannya[3]. Hasil ini bisa digunakan untuk memberi saran pertimbangan dalam menentukan Status Ekonomi Mahasiswa Baru di Universitas Ichsan Gorontalo.

Universitas Ichsan Gorontalo memiliki 13 program studi yang dimana memiliki cukup banyak mahasiswa didalamnya. Pada penelitian ini penulis ingin mengelompokan status ekonomi mahasiswa baru khususnya di tahun 2022 dengan menggunakan metode *K-Means* agar bisa terkelompok sesuai dengan

kategori masing-masing. Adapun data mahasiswa sebagai berikut:

Tabel 1.1 Dataset Mahasiswa

No	Nama	Pekerjaan Orang Tua	Penghasilan Orang Tua	Tanggungan Orang Tua
1	NURAFNI	PETANI	Rp. 550.000	3
2	CITRA SALEH	PETANI	Rp. 600.000	4
3	SAHRUN MOODUTO	WIRASWASTA	Rp. 2.250.000	4
4	BELA SAFIRA BONDE	PEGAWAI	Rp. 3.400.000	5
5	MOH. ADIT PURNAWANSYAH DJALILU	PETANI	Rp. 530.000	2
6	MEN WASIANGGE	WIRASWASTA	Rp. 4.800.000	5
7	NUR NITA M. LATIF	PEGAWAI	Rp. 3.200.000	5
8	SYAHRUL RAMADHAN KALUKU	PEGAWAI	Rp. 2.750.000	5
9	I KOMANG CRISFANDU	PEGAWAI	Rp. 3.500.000	5
10	NADIA PUTRI OKTAVIANI THOMASONG	PETANI	Rp. 1.500.000	4
...	...	...	...	...
812	SRI ARYAWATI LAILA MAYULU	PNS	Rp. 4.750.000	4

Sumber: BAAK UNISAN Gorontalo, 2022

Dari tabel di atas dapat kita lihat ada berbagai macam pekerjaan hingga penghasilan orang tua serta tanggungan orang tua dari masing-masing mahasiswa. Masalah dari penelitian ini pastinya dengan banyaknya data yang ada akan sulit dalam mengetahui status ekonomi mahasiswa. Oleh karena itu dengan adanya *clustering* status ekonomi mahasiswa dengan menggunakan metode *K-Means* bisa mempermudah untuk mengetahui kategori atau kelas-kelas dari masing-masing mahasiswa. Tujuan dari penelitian ini adalah untuk mengetahui analisa metode *K-Means* dalam mengcluster status ekonomi mahasiswa. Diharapkan dengan adanya pengelompokkan status ekonomi mahasiswa ini dapat memiliki kategori atau kelasnya masing-masing.

Pada beberapa penelitian sebelumnya yang dilakukan oleh Rezqiwati Ishak dan Amiruddin dengan judul “*Clustering* Tingkat Pemahaman Dasar Mahasiswa Pada Pra-Kuliah Probabilitas Statistika Dengan Metode *K-Means*”, dimana hasil dari penelitian ini mendapat hasil sebanyak 3 *cluster*, dimana *cluster* 1 artinya kumpulan Mahasiswa tersebut digolongkan kategori Baik karena dari taraf pemahamannya, *Cluster* 2 artinya kumpulan mahasiswa yang kategorinya Sangat Baik dan *Cluster* 3 ialah kumpulan mahasiswa yang kategorinya Kurang. Dari dataset yang ada dengan jumlah 110 Mahasiswa, masing-masing ditemukan untuk *cluster* 1 jumlahnya 30 Mahasiswa, *cluster* 2 jumlahnya 58 Mahasiswa dan *cluster* 3 jumlahnya 22 Mahasiswa. Mahasiswa mendapatkan klasterisasi dengan nilai yang mendekati karakteristik menjadi lebih efektif dan mendapatkan hasil *clustering* yang optimal[4]. Hal ini menunjukkan bila metode *K-Means* dapat digunakan dalam proses pengclusteran Status Ekonomi Mahasiswa Baru di Universitas Ichsan Gorontalo.

Berdasarkan hal tersebut diatas, peneliti bermaksud melakukan penelitian dengan judul “***Clustering* Status Ekonomi Mahasiswa Baru Dengan Metode *K-Means***” Studi kasus Universitas Ichsan Gorontalo.

1.2 Identifikasi Masalah

Berdasarkan uraian latar belakang diatas maka dapat diidentifikasi permasalahan yaitu sulitnya melakukan *Clustering* Status Ekonomi Mahasiswa Baru berdasarkan pekerjaan orang tua, penghasilan orang tua, dan tanggungan orang tua.

1.3 Rumusan Masalah

Dari uraian latar belakang masalah di atas, dapat dirumuskan masalah pokok yang berkaitan yaitu:

1. Bagaimana hasil penerapan metode *K-Means* untuk *Clustering* Status Ekonomi Mahasiswa Baru di Universitas Ichsan Gorontalo?

2. Bagaimana hasil analisa *Clustering* Status Ekonomi Mahasiswa Baru di Universitas Ichsan Gorontalo?

1.4 Tujuan Penelitian

Berdasarkan rumusan masalah yang telah diuraikan, maka tujuan penelitian ini sebagai berikut:

1. Menerapkan hasil metode *K-Means* untuk *Clustering* Status Ekonomi Mahasiswa Baru di Universitas Ichsan Gorontalo.
2. Menganalisis hasil *Clustering* Status Ekonomi Mahasiswa Baru di Universitas Ichsan Gorontalo.

1.5 Manfaat Penelitian

Penelitian ini dimohonkan membawa faedah secara teoritis dan praktis dan praktis kepada pihak terkait, sebagai pertimbangan, masukkan dan pedoman serta evaluasi.

1. Secara Teoritis, penelitian ini diharapkan dapat mengembangkan ilmu pengetahuan sehingga dapat meningkatkan metode yang digunakan.
2. Secara Praktis, penelitian ini diharapkan dapat mengembangkan ilmu pengetahuan bagi instansi sebagai bahan informasi agar dapat mengoptimalkan status ekonomi mahasiswa.

BAB II LANDASAN TEORI

2.1 Tinjauan Studi

Beberapa penelitian tentang *Clustering* dan penerapan metode K-Means sebagai berikut:

Tabel 2.1 Penelitian Tentang *Clustering* Menggunakan Metode *K-Means*

Peneliti	Judul	Hasil
Rezqiwati Ishak, Amiruddin (2022)[4].	<i>Clustering</i> Tentang Tingkat Pemahaman Dasar Mahasiswa Pada Pra-Perkuliahan Probabilitas Statistika Dengan Metode <i>K-Means</i>	Penelitian ini memiliki tujuan untuk menyelesaikan pengelompokkan atau <i>clustering</i> tingkat pemahaman dasar mahasiswa berdasarkan nilai mata kuliah prasyarat yaitu Kalkulus dan Matematika Diskrit. Berdasarkan hasil penelitan diperoleh hasil sejumlah 3 <i>cluster</i> , dimana <i>cluster</i> 1 ialah kelompok Mahasiswa yang kualitas pemahamannya dikategorikan Baik, cluster 2 ialah kelompok yang kategorinya sangat baik dan cluster 3 ialah yang kategorinya kurang. Dari

Peneliti	Judul	Hasil
		110 Mahasiswa yang menjadi sebagai dataset, masing-masing ditemukan untuk <i>cluster</i> 1 sejumlah 30 Mahasiswa, <i>cluster</i> 2 sejumlah 58 Mahasiswa dan <i>cluster</i> 3 sejumlah 22 Mahasiswa[4].
Andy Febrianto, Sentot Achmadi, Agung Panji Sasmito (2021)[5].	Penerapan Metode <i>K-Means</i> Untuk Clustering Pengunjung Perpustakaan IT Malang	Hasil dari penelitian ini yaitu aplikasi <i>website</i> , yang mempunyai fitur yaitu sistem pada websitenya menginformasikan pengunjung dan peminjam, <i>user</i> bisa memasukkan data pengunjung seperti <i>member id</i> , <i>visitor name</i> , <i>membership type</i> , <i>institution</i> serta data peminjam seperti <i>member id</i> , <i>member name</i> , <i>item code</i> , <i>title</i> , <i>loan status</i> , menurut pengujian fungsional sistem mengenai fitur yaitu seluruhnya berhasil dan berjalan sangat baik, menurut pengujian <i>user</i> hasil presentasi diketahui responden 50% sangat baik, 48% serta 2% kurang baik[5].

Peneliti	Judul	Hasil
Achmad Solichin, Khansa Khairunnisa, (2020)[6].	Klasterisasi Persebaran Virus Corona (Covid 19) Di DKI Jakarta Menggunakan Metode <i>K-Means</i>	Pada penelitian ini dilakukan klasterisasi persebaran virus Corona di DKI Jakarta dalam menggunakan metode <i>data mining</i> . Pengelompokan dilakukan berdasarkan kriteria jumlah ODP, PDP, kasus Positif, pasien sembuh dan pasien meninggal. Di penelitian ini, dalam melakukan klasterisasi data dipakai metode <i>K-Means</i> dan metode penaksiran jarak <i>Euclidean</i> . Penelitian ini menghasilkan prototipe aplikasi pengelompokan data penyebaran pasien Covid-19. Berdasarkan pengujian, total klaster yang direkomendasikan ialah 9 klaster[6]

2.2 Tinjauan Pustaka

2.2.1 Status Ekonomi

Status ekonomi merupakan kedudukan atau posisi seseorang pada masyarakat, dimana status ekonomi ini artinya latar belakang ekonomi seseorang yang didasari oleh beberapa faktor pada kepentingan manusia dikehidupannya, yaitu dengan taraf pendidikan, pekerjaan atau mata pencaharian, penghasilan/pendapatan, serta sosial[7].

Pengaruh status ekonomi orang tua sangat dibutuhkan untuk memenuhi kebutuhan anaknya. Manusia menjadi makhluk sosial yang memiliki kesanggupan untuk memungkinkan dia diterima dalam suatu lingkungan dengan individu lain. Sebab itulah dengan suatu kepentingan tertentu orang akan menyalurkan setiap potensinya[7].

2.2.2 Data Mining

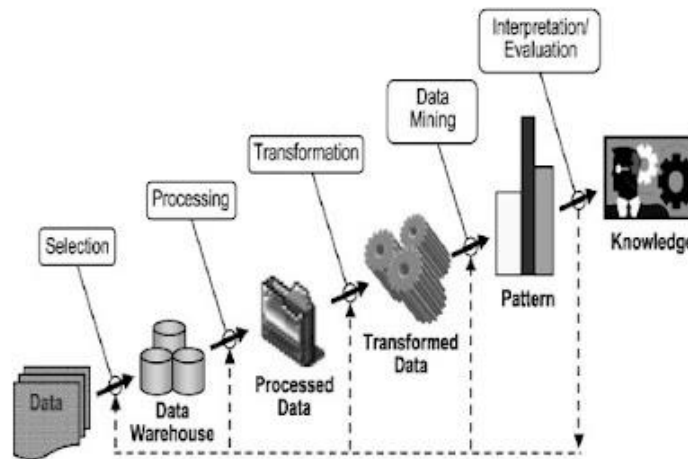
Data mining artinya proses mengumpulkan dan mengolah data yang bertujuan untuk mengekstrak informasi penting pada data. Proses tersebut bisa dilakukan dengan bantuan perhitungan statistika, matematika, ataupun teknologi *Artificial Intelligence* (AI). *Data mining* juga dilakukan untuk memiliki beberapa tujuan tertentu, seperti sebagai sarana menjelaskan, sebagai sarana konfirmasi serta sebagai sarana eksplorasi[8], [9].

Data mining memiliki banyak fungsi salah satunya yang paling utama adalah sebagai fungsi *descriptive* dan fungsi *predictive*. Fungsi *descriptive* (deskripsi) dalam *data mining* merupakan sebuah fungsi dalam memahami lebih jauh tentang data yang diamati, dengan begitu maka dapat ditemukan pola tertentu yang bersembunyi didalam sebuah data, sehingga karakteristik pola dan nilai suatu data dapat diketahui. Sedangkan fungsi *predictive* (prediksi) adalah suatu fungsi yang dimana sebuah cara untuk mendapatkan bentuk dari sebuah data. Bentuk tersebut didapatkan dari variabel yang ada di data. Jika sudah ditemukan pola tersebut, maka dapat digunakan untuk memprediksi variabel lain yang belum diketahui nilai serta jenis-jenisnya. Oleh karena itu, fungsi ini memudahkan dan menguntungkan dalam prediksi yang lebih tepat[9], [10].

2.2.3 Tahapan Data Mining

Dalam tahapan proses *data mining* dimulai dari menyaring data, data asal ke data target, tahapan *preprocessing* dalam memperbaiki kualitas data, transformasi, *data mining* dan tahapan interpretasi dan evaluasi yang menghasilkan

output berbentuk pengetahuan baru yang dibutuhkan untuk memberikan bantuan yang lebih baik[11]. Secara detail dijelaskan sebagai berikut:



Gambar 2.1 Proses *Knowledge in Database* (KDD)

1. Data Selection

Memilih atau menyeleksi data dari sekumpulan data operasional harus dilakukan sebelum tahap pemeriksaan informasi ketika KDD (*Knowledge Discovery in Database*) dimulai. Data yang terseleksi dipakai untuk proses *data mining*, disimpan pada suatu *file*, disisihkan dari basis data operasional.

2. Pre-processing / cleaning

Sebelum proses *data mining*, maka wajib dikerjakan *cleaning* untuk data yang sebagai penekanan KDD. Proses *cleaning* melingkupi penghapusan duplikat data, menyelidiki data yang tidak konsisten, dan membenahi kesalahan di data.

3. Transformation

Coding merupakan perubahan data yang dipilih, akibatnya data itu cocok dalam proses *data mining*. Proses coding pada KDD adalah proses yang inovatif serta tergantung jenis atau bentuk informasi yang akan dicari pada basis data.

4. Data Mining

Data mining merupakan proses pencarian pola atau informasi memikat di data yang terpilih dengan memakai gaya atau metode tertentu. Teknik, metode, atau

algoritma pada *data mining* beragam. Metode atau algoritma yang sesuai bergantung pada tujuan serta proses KDD secara keseluruhan.

5. *Interpretation / evaluation*

Menghasilkan bentuk informasi dari proses *data mining* pastinya harus menampilkan pada bentuk yang praktis dipahami oleh sisi yang penting. Tahapan ini ialah proses KDD yang diklaim dengan *interpretation*. Tahapan ini melingkupi investigasi dimana bentuk atau informasi berlawanan dengan fakta atau hipotesis sebelumnya.

2.2.4 Clustering

Clustering merupakan pengelompokkan data yang memiliki sifat yang mirip. Data di *clustering* belum memiliki kelas-kelas, sehingga konsep *clustering* ini bisa mengoptimalkan dan meminimalkan kesamaan antar kelas. Kemiripan yang menjadi dasar dalam mengelompokkan data ini tidak bersifat *universal* sehingga ukuran-ukuran persamaannya harus dijabarkan lebih dahulu oleh peneliti. Jadi, *clustering* merupakan proses pengelompokkan data ke dalam beberapa *cluster* atau kelompok sehingga data pada sebuah *cluster* mempunyai taraf kemiripan yang maksimum dan data antar *cluster* yang tidak selaras mempunyai kemiripan yang minimum[11].

2.2.5 Algoritma K-Means

K-Means clustering merupakan salah satu algoritma yang menganalisa data atau metode *data mining* yang melakukan proses pemodelan *unsupervised learning* serta memakai metode yang mengelompokkan data di berbagai partisi. *K-Means* tergolong dalam *Partitioning Clustering*, dimana setiap data wajib masuk ke dalam *cluster* tertentu serta pada tahapan proses memungkinkan setiap data yang masuk pada *cluster* tertentu, pada tahapan berikutnya berpindah ke *cluster* yang lain. Secara umum tahapan algoritma *K-Means* sebagai berikut[11].

1. Tentukan k buah *cluster*
2. Pilih sejumlah k buah objek secara acak yang akan dijadikan sebagai titik

centroid cluster

3. Tentukan k buah *centroid* (titik tengah)
4. Kelompokkan obyek ke *centroid cluster* terdekat berdasarkan *Euclidian distance*:

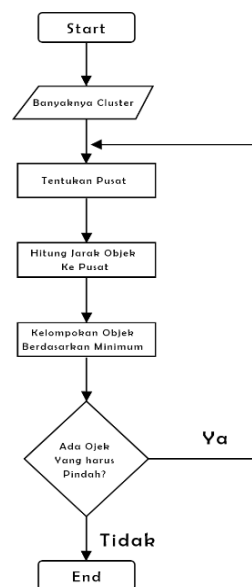
$$d_{ik} = \sqrt{\sum_{i=1}^n (C_{ij} - C_{kj})^2}$$

Keterangan:

C_{ij} = pusat *cluster*

C_{kj} = data

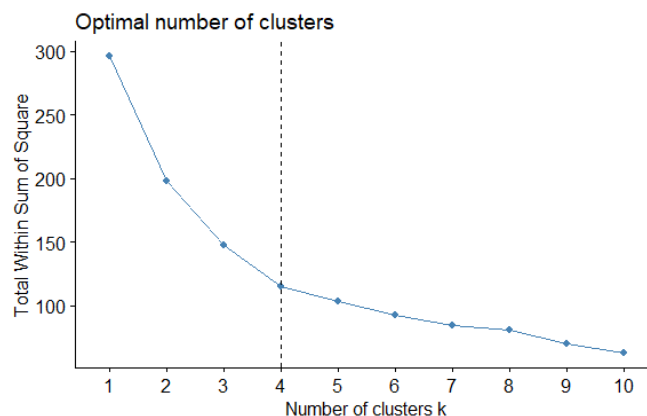
5. Hitung kembali titik *centroid*
6. Ulangi langkah 3-5 sehingga nilai titik *centroid* tidak lagi berubah.



Gambar 2.2 Proses Algoritma K-Means

2.2.6 Metode *Elbow*

Metode *Elbow* merupakan metode yang tak jarang digunakan untuk menentukan jumlah *cluster* yang digunakan pada *K-Means Clustering*. Tujuannya adalah untuk menghitung WCSS (*Whitinn-Cluster Sum of Square*) seminimum mungkin menggunakan jumlah *cluster* yang kecil agar bisa dilakukan interpretasi data. Hal ini dapat dilihat dari *graphic* dibawah ini, dimana jumlah $K=4$ adalah jumlah yang optimal[12], [13].



Gambar 2.3 Contoh Visualisasi Pemograman Metode *Elbow*

2.2.7 *Davies-Bouldin Index (DBI)*

Davies-Bouldin Index (DBI) adalah sebuah metrik yang digunakan untuk mengukur kualitas kluster dalam analisis cluster atau clustering. Tujuannya adalah untuk menentukan sejauh mana setiap *cluster* berbeda satu sama lain. DBI dapat membantu dalam memilih jumlah *cluster* yang optimal dalam suatu data, karena nilai DBI lebih rendah menunjukkan kluster yang lebih homogen dan berdiameter kecil. Lebih spesifik, DBI mengukur rata-rata dari rasio antar titik dalam kluster dengan jarak antar kluster. Semakin rendah nilai DBI, semakin baik kualitas klasternya[14], [15].

2.2.8 Contoh Penerapan Algoritma *K-Means*

Berikut ini adalah contoh penerapan algoritma *K-Means* yang diambil dari peneliti sebelumnya oleh Rezqiwati Ishak dan Amiruddin dengan judul "*Clustering Tingkat Pemahaman Dasar Mahasiswa Pada Pra-Perkuliahan Probabilitas Statistika dengan Metode K-Means*". Adapun *tools* yang digunakan dalam proses *clustering* ini adalah *tools Jupyter Notebook* dengan bahasa pemograman *Python* menggunakan beberapa *library* seperti *Numpy*, *Pandas*, *Matplotlib* dan *Scikit Learn*[4].

1. Tahapan *Selection*

Pada tahap ini dilakukan seleksi dataset yang akan digunakan yaitu nilai mahasiswa pada mata kuliah Kalkulus pada tabel berikut:

Tabel 2.2 Sampel Dataset Nilai Kalkulus

No	NIM	Nilai
1	T3119101	D
2	T3120002	A
3	T3120003	B
4	T3120004	A
5	T3120005	A
6	T3120006	B
7	T3120007	A
8	T3120008	A
9	T3120009	A
10	T3120010	A

Tabel 2.3 Dataset Nilai Matematika Diskrit

No	NIM	Nilai
1	T3114097	B
2	T3116126	D
3	T3116345	D
4	T3117006	B
5	T3117057	B
6	T317069	D
7	T3117149	B
8	T3117168	C
9	T3118003	D
10	T3120010	A

2. Tahap *Preprocessing*

Selanjutnya di tahapan ini dilakukan proses menghapus data yang sama, melengkapi data yang kosong dan menggabungkan data nilai Kalkulus dan nilai Matematika Diskrit, proses penggabungan data ini langsung diproses di pemograman *Phyton*, hasil akhir penggabungan data seperti pada tabel berikut[4]:

Tabel 2.4 Sampel Hasil Penggabungan Dataset

No	NIM	Kalkulus	M. Diskrit
1	T3119101	A	A
2	T3120002	B	C
3	T3120003	A	D
4	T3120004	A	A
5	T3120005	B	D
6	T3120006	A	A
7	T3120007	A	B
8	T3120008	A	B
9	T3120009	A	A
10	T3120010	B	A

3. Tahap *Transformation*

Tahapan ini melakukan konversi data nilai ialah nilai mata kuliah yang mana bentuk huruf dikonversi ke bentuk angka yaitu nilai A=4, B=3, C=2, D=1, keadaan ini dilakukan agar bisa menggunakan pemodelan metode *K-Means*. Tahapan konversi ini juga langsung dilakukan di dalam *Python* dengan hasil akhir seperti pada tabel berikut ini[4]:

Tabel 2.5 Sampel Hasil Nilai Huruf ke Angka

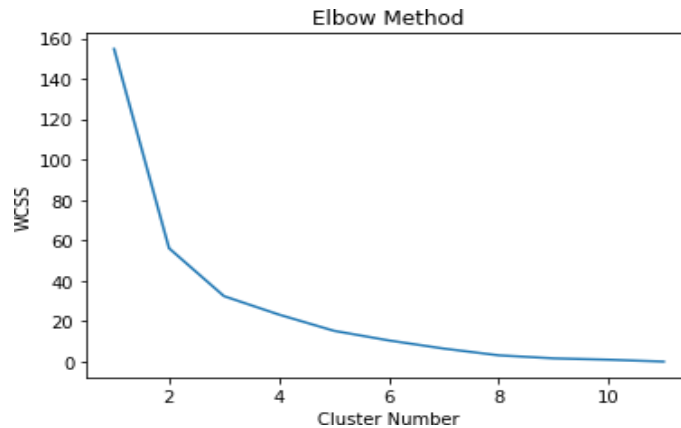
No	NIM	Nilai	M. Diskrit
1	T3119101	4	4
2	T3120002	3	2
3	T3120003	4	1
4	T3120004	4	4
5	T3120005	3	1
6	T3120006	4	4
7	T3120007	4	3
8	T3120008	4	3
9	T3120009	4	4
10	T3120010	3	4

4. Tahap *Data Mining*

Tahap berikutnya melakukan pengembangan model *K-Means* dalam melakukan *clustering*. Perintah dalam proses pemodelan metode *K-Means* ini memerlukan *library* dari *Scikit Learn*.

5. Tahap *Interpretation/Evaluation*

Tahapan terakhir melakukan evaluasi pada jumlah *cluster* yang optimum dengan memakai metode *elbow*, selanjutnya hasil metode tersebut dibuatkan dalam pola visualisasi dengan membutuhkan *Library Matplotlib*, setelah dijalankan perintahnya, maka dihasilkan visualisasi pada gambar berikut ini:



Gambar 2.4 Visualisasi Hasil Metode *Elbow*

Pada gambar di atas, dapat dilihat bahwa pada penelitian ini jumlah *cluster* yang optimum adalah sebanyak 3 *cluster*. Dengan menggunakan metode *Elbow* ini, jumlah *cluster* yang optimum bisa langsung diketahui tanpa harus menguji jumlah *cluster* secara berulang. Setelah didapatkan jumlah *cluster* yang optimum, maka jumlah *cluster* tersebut dimasukkan kedalam fungsi *K-Means* yang tersedia pada *library Scikit Learn* dengan tahapan sebagai berikut[4]:

1. Baca dataset yang sudah diproses:

```
X = dfData.iloc[:,[1,2]].values
```

2. Panggil Fungsi *K-Means*:

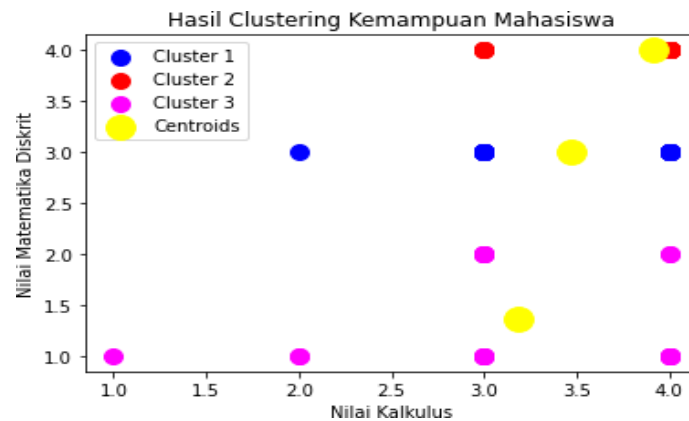
```
k-means = KMeans(n_clusters = 3, init = 'k-means++')
```

3. Terapkan hasil pemodelan terhadap dataset:

```
y_kmeans = kmeans.fit_predict(X)
```

4. Lakukan visualisasi hasil *clustering* atau pemodelan

Berikut visualisasi hasil *clustering* berdasarkan pemodelan *K-Means* yang sudah dilakukan :



Gambar 2.5 Visualisasi Hasil *Cluster*

Sesuai gambar di atas, dapat ditentukan titik *Centroid* masing-masing *cluster* dengan melihat posisi koordinat pusat *cluster* yang diberikan warna kuning atau dapat juga diketahui secara langsung dengan menggunakan perintah “(*kmeans.cluster_center_*)” sehingga didapatkan hasil sebagai berikut:

```
[[3.46666667 3.
 3.9137931 4.
 3.18181818 1.36363636]]
```

Gambar 2.6 Informasi Hasil *Cluster*

Pada jumlah anggota masing-masing *cluster* dapat diketahui dengan menggunakan perintah “*dfData['Kluster'].value_counts()*”. Berdasarkan kedua perintah di atas, dapat dimodel hasilnya dalam bentuk tabel seperti pada tabel berikut ini:

Tabel 2.6 *Centroid* Dan Jumlah Data

Cluster	Centroid		Jumlah Mahasiswa
	Kalkulus	Matematika Diskrit	
1	3,46	3	30
2	3,91	4	58
3	3,18	1,36	22

Berdasarkan tabel di atas, dengan melihat titik *Centroid* setiap *cluster*, maka bisa dianalisa atau dikategorikan bahwa *cluster* 1 adalah kategori tingkat pemahaman Mahasiswa kategori baik, *cluster* 2 adalah kategori sangat baik dan *cluster* 3 adalah kelompok Mahasiswa yang perlu diberikan materi tambahan sebelum mengikuti perkuliahan Mata Kuliah Probabilitas Statistika sebanyak 22 Mahasiswa, supaya proses perkuliahan berjalan dengan lancar dan apa yang menjadi tujuan capaian pembelajaran bisa tercapai. Hasil pelabelan *clustering* pada setiap Mahasiswa berdasarkan *file Excel* yang sudah di *download* dari hasil pemodelan di aplikasi *Python* bisa dilihat tabel di bawah ini[4]:

Tabel 2.7 *Sample Hasil Clustering*

NIM	Kalkulus	M_Diskrit	Kluster
T3120002	4	4	2
T3120003	3	2	3
T3120004	4	1	3
T3120005	4	4	2
T3120006	3	1	3
T3120007	4	4	2

NIM	Kalkulus	M_Diskrit	Kluster
T3120008	4	3	1
T3120009	4	3	1
T3120010	4	4	2
T3120011	3	4	1

Berdasarkan informasi tabel di atas, maka dengan praktis dapat diketahui setiap Mahasiswa masuk ke *cluster* berapa. Sehingga jika ingin dicari atau diketahui Mahasiswa yang masuk *cluster* 3, maka cukup dilakukan *filter* untuk *cluster* tersebut[4].

2.2.9 Python

Python merupakan bahasa pemrograman *interpretatif multifungsi* menggunakan teori perancangan yang fokusnya pada tingkat keterbacaan kode. *Python* di tuntut sebagai bahasa yang menggabungkan kapabilitas., kemampuan, dengan sintaksis kode yang sangat jelas, dan dilengkapi dengan fungsionalitas pustaka standar yang besar serta *komprehensif* [16].

Python mendukung banyak model pemrograman, utamanya namun tidak dibatasi pada program berorientasi objek, pemrograman *imperatif*, serta pemrograman fungsional. Salah satu *fitur* yang tersedia pada *python* merupakan bahasa pemrograman aktif yang dilengkapi dengan manajemen memori otomatis. Sama halnya dengan dengan bahasa pemrograman lainnya, *python* umumnya digunakan sebagai bahasa *script* walau pada praktiknya penggunaan bahasa ini sangat luas meliputi kondisi pemanfaatan yang umumnya tidak dilakukan dengan menggunakan bahasa *script*. *Python* dapat dipergunakan untuk beragam keperluan pengembangan perangkat lunak dan dapat berjalan di beragam *platform* sistem operasi[16].

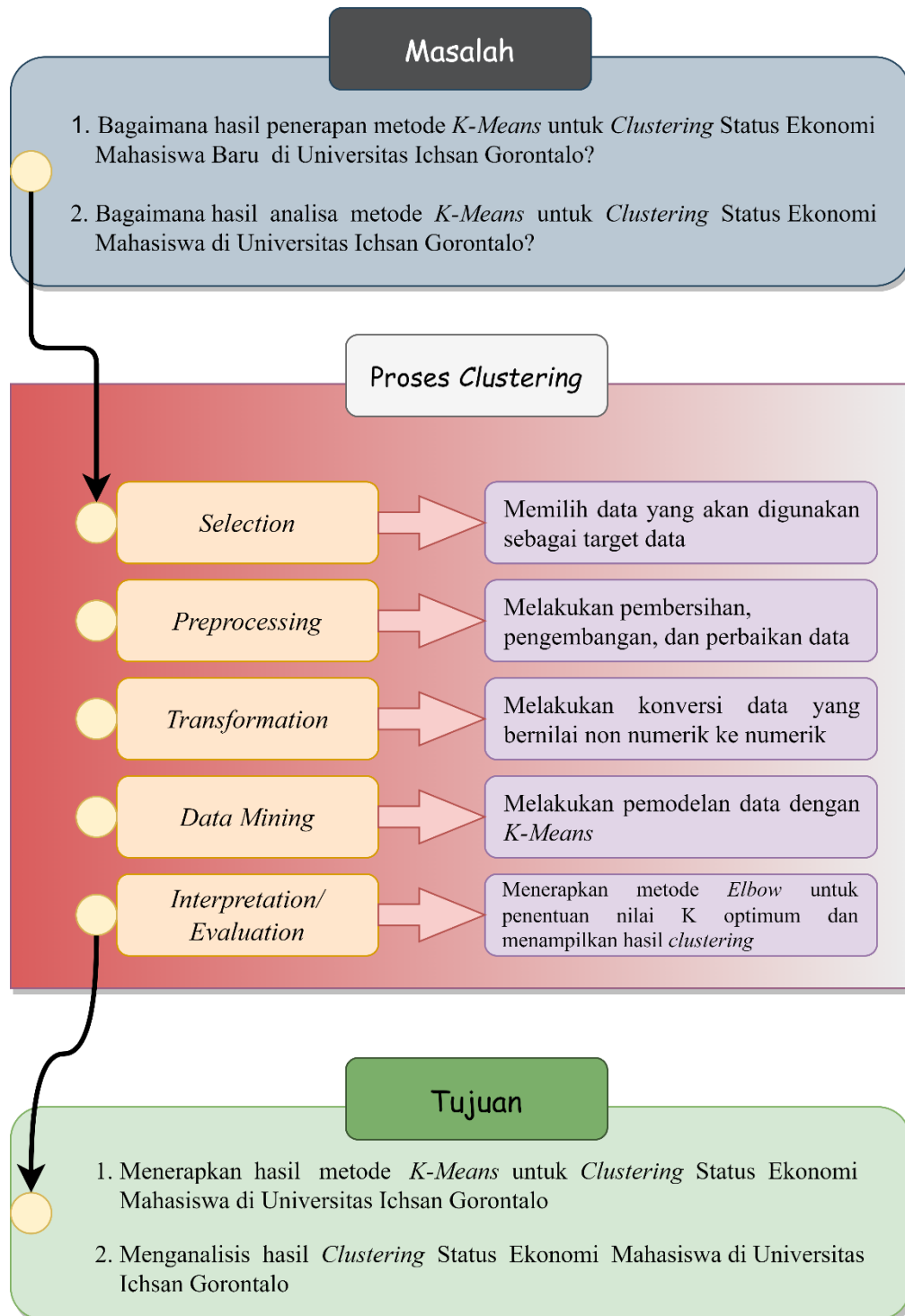
2.3 Perangkat Lunak

Perangkat lunak yang digunakan dalam membuat *clustering* ini yaitu, seperti pada tabel di bawah ini:

Tabel 2.8 Perangkat Lunak Pendukung

No	Perangkat Lunak	Keterangan
1.	<i>Python</i>	<i>Python</i> adalah bahasa pemrograman yang digunakan untuk membuat aplikasi, perintah komputer, dan melakukan analisa data. selain itu juga bisa digunakan untuk membuat program apa saja dan menyelesaikan berbagai masalah.
2.	<i>Google Colab</i>	Untuk digunakan menjalankan kode <i>Python</i> tanpa perlu melakukan proses instalasi.

2.4 Kerangka Pikir



Gambar 2.7 Bagan Kerangka Pikir

BAB III

METODE PENELITIAN

3.1 Jenis, Metode, Subjek, Objek, Waktu dan Lokasi Penelitian

Dilihat dari tingkat penerapannya penelitian ini adalah penelitian terapan sedangkan jenis informasi yang diolah pada penelitian ini adalah penelitian kuantitatif. Dilihat dari perlakuan data, maka penelitian ini merupakan penelitian konfirmatori. Dengan demikian jenis penelitian ini merupakan penelitian deskriptif.

Subjek penelitian ini adalah klasterisasi mahasiswa pada objek status ekonomi mahasiswa baru menggunakan metode *K-Means*. Penelitian ini dimulai bulan September 2022 yang berlokasi di Universitas Ichsan Gorontalo.

3.2 Pengumpulan Data

Adapun jenis pengumpulan data ini yaitu data primer dan sekunder. Data primer merupakan data yang dikumpulkan langsung di lapangan. Sedangkan untuk data sekunder adalah data yang diperoleh dari penelitian sebelumnya seperti jurnal yang membahas tentang data mining serta membahas *clustering* dengan menggunakan metode *K-Means*. Baik dari internet maupun Perpustakaan. Adapun variabel/atribut dengan tipe datanya masing-masing yang ditunjukkan dari tabel berikut:

Tabel 3.1 Variabel/Atribut Data

No	Variabel/Atribut	Type	Keterangan
1	Pekerjaan Orang Tua	Varchar	Variabel Input
2	Penghasilan Orang Tua	Integer	Variabel Input
3	Tanggungan Orang Tua	Integer	Variabel Input

3.3 Pemodelan

Prosedur atau langkah-langkah pada proses *clustering* menggunakan metode *K-Means* untuk memudahkan dalam melakukan tahapan *clustering* yaitu:

1. Tahap *Selection* : pada tahap ini dilakukan seleksi terhadap dataset yang dikumpulkan dari data jumlah mahasiswa baru di Universitas Ichsan Gorontalo untuk di jadikan sebagai target data.
2. Tahap *Preprocessing* : selanjutnya ditahapan ini melakukan proses penghapusan data yang sama, pemeriksaan data yang tidak konsisten dan melengkapi data yang belum ada.
3. Tahap *Transformation* : pada tahap selanjutnya dilaksanakan perubahan atau konversi data yaitu bentuk huruf dikonversi ke bentuk angka agar bisa digunakan untuk pemodelan *K-Means*.
4. Tahap *Data Mining* : tahap selanjutnya dilaksanakan pengembangan model *K-Means* agar menghasilkan *cluster* yang akan dibentuk.
5. Tahap *Interpretation/Evaluation* : kemudian pada tahapan ini dilaksanakan penentuan beberapa hasil *cluster* yang optimal saat menggunakan metode *Elbow*. Hasil nilai K optimum dari metode *Elbow* yang kemudian diterapkan ke dalam metode *K-Means* kembali. Hasil dari *clustering* tersebut di visualisasikan jumlah *cluster* beserta member *clusternya*, maka dari itu mudah di pahami arti dari masing-masing kelompok *cluster* tersebut yang kemudian dijadikan sebagai pengetahuan atau *clustering* status ekonomi mahasiswa.

BAB IV HASIL PENELITIAN

4.1 Hasil Pengumpulan Data

Tabel 4.1 Hasil Pengumpulan Dataset

No	NIM	Nama	Tempat Lahir	Kelamin	Tanggal Lahir
1	E1122001	NURAFNI	POHUWATO	P	10/03/2005
2	E1122002	CITRA SALEH	TERNATE	P	11/10/2004
3	E1122003	SAHRUN MOODUTO	IRIAN JAYA	L	04/11/2001
4	E1122004	BELA SAFIRA BONDE	TORAUT	P	28/04/2003
5	E1122005	MOH. ADIT PURNAWANSYAH DJALILU	SUWAWA	L	16/04/2003
6	E1122006	MEN WASIANGGE	MBUA	L	14/07/2001
7	E1122007	NUR NITA M. LATIF	TELAGA	P	01/11/1989
8	E1122008	SYAHRUL RAMADHAN KALUKU	GORONTALO	L	09/12/2001
9	E1122009	I KOMANG CRISFANDU	BONGO IV	L	15/05/2003
10	E1122010	NADIA PUTRI OKTAVIANI THOMASONG	GORONTALO	P	05/10/2004
...	...	...	...	...	...
812	E1122022	DELIA TIARA PUTERI TAMUU	BUOL	L	07/08/2001

No	Prodi	Pekerjaan Orang Tua	Penghasilan Orang Tua	Tanggungan
1	S1 AKUNTANSI	PETANI	Rp550.000	3 ORANG
2	S1 AKUNTANSI	PETANI	Rp600.000	4 ORANG
3	S1 AKUNTANSI	WIRASWASTA	Rp2.250.000	4 ORANG
4	S1 AKUNTANSI	PEGAWAI	Rp3.400.000	5 ORANG
5	S1 AKUNTANSI	PETANI	Rp530.000	2 ORANG
6	S1 AKUNTANSI	WIRASWASTA	Rp4.800.000	5 ORANG
8	S1 AKUNTANSI	PEGAWAI	Rp3.200.000	5 ORANG
9	S1 AKUNTANSI	PEGAWAI	Rp2.750.000	5 ORANG
10	S1 AKUNTANSI	PEGAWAI	Rp3.500.000	5 ORANG
...	...	...	...	...
812	S1 AGROTEKNOLOGI	NELAYAN	Rp1.500.000	3 ORANG

4.2 Tahapan *Clustering K-Means*

1. Pengumpulan Dataset

Pada tahap ini dikumpulkan data untuk dapat diolah pada algoritma *K-Means*, dimana ada beberapa data yang dihapus karena data tersebut kosong sehingga hasil data yang dapat diolah tersisa 765 data dari 812 dataset. Berikut adalah hasil pengumpulan dataset yang sudah diolah pada Tabel 4.2 berikut:

Tabel 4.2 Dataset Hasil Olahan

No	Nama	Pekerjaan	Penghasilan	Tanggung
1	NURAFNI	PETANI	Rp550.000	3
2	CITRA SALEH	PETANI	Rp600.000	4
3	SAHRUN MOODUTO	WIRASWASTA	Rp2.250.000	4
4	BELA SAFIRA BONDE	PEGAWAI	Rp3.400.000	5
5	MOH. ADIT PURNAWANSYAH DJALILU	PETANI	Rp530.000	2
6	MEN WASIANGGE	WIRASWASTA	Rp4.800.000	5
7	NUR NITA M. LATIF	PEGAWAI	Rp3.200.000	5
8	SYAHRUL RAMADHAN KALUKU	PEGAWAI	Rp2.750.000	5
9	I KOMANG CRISFANDU	PEGAWAI	Rp3.500.000	5
10	NADIA PUTRI OKTAVIANI THOMASONG	PETANI	Rp1.500.000	4
11	APRILIA NGADI	PEGAWAI	Rp2.500.000	3
12	SRI ARYAWATI LAILA MAYULU	PNS	Rp4.750.000	4
13	YOLANDA RAHMAN	PNS	Rp2.500.000	2
14	ANGGUN DHEA PUTRI AMRAIN	WIRASWASTA	Rp4.500.000	3
15	DJOKO RAMADHON	PNS	Rp3.200.000	4
16	HERLINA MOHA	PETANI	Rp750.000	5
17	ASTRIA FATMAH YUNUS	PNS	Rp3.000.000	2
18	FARHAN HULUKATI	PNS	Rp3.500.000	4
...	...	...	...	...
765	ALDI	NELAYAN	Rp1.500.000	3

2. Tahapan Transformasi

Pada tahap transformasi ini dilakukan konversi atribut yang ada dari nilai non numerik ke numerik agar bisa diolah yang dapat dilihat pada Tabel 4.4 berikut:

Tabel 4.3 Konversi Nilai Pekerjaan Orang Tua

No	Pekerjaan	Jumlah
1	Petani	207
2	Wiraswasta	192
3	Pegawai	130
4	PNS	125
5	Wirausaha	51
6	Nelayan	34
7	Lainnya	26

Setelah didapatkan jumlah banyaknya dari masing-masing pekerjaan yang ada, maka didapatkan hasil dari konversi pada atribut pekerjaan yang dapat dilihat pada Tabel 4.5 sebagai berikut:

Tabel 4.4 Hasil Konversi Atribut Pekerjaan

No	Pekerjaan	Hasil Konversi
1	Petani	1
2	Nelayan	2
3	Lainnya	3
4	Pegawai	4
5	PNS	5
6	Wiraswasta	6
7	Wirausaha	7

Setelah dilakukan konversi data seperti di atas jadi tahap selanjutnya dilakukan pemodelan *clustering* dengan menggunakan metode *K-Means*.

3. Tahapan *Selection*

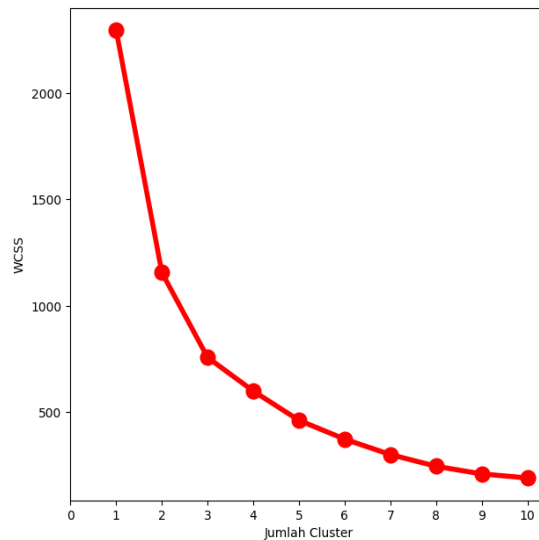
Pada tahap ini akan dilakukan seleksi data dari dataset yang sudah dikumpulkan. Dari dataset yang ada hanya digunakan pekerjaan orang tua, penghasilan orang tua, dan tanggungan orang tua sebagai atribut yang akan digunakan.

Tabel 4.5 Dataset Hasil Olahan

No	Pekerjaan	Penghasilan	Tanggungan
1	PETANI	550.000	3
2	PETANI	600.000	4
3	WIRASWASTA	2.250.000	4
4	PEGAWAI	3.400.000	5
5	PETANI	530.000	2
6	WIRASWASTA	4.800.000	5
7	PEGAWAI	3.200.000	5
8	PEGAWAI	2.750.000	5
9	PEGAWAI	3.500.000	5
10	PETANI	1.500.000	4
11	PEGAWAI	2.500.000	3
12	PNS	4.750.000	4
13	PNS	2.500.000	2
14	WIRASWASTA	4.500.000	3
15	PNS	3.200.000	4
16	PETANI	750.000	5
17	PNS	3.000.000	2
18	PNS	3.500.000	4
...	...	...	...
765	NELAYAN	Rp1.500.000	3

4. Tahapan Pemilihan Nilai K Optimum

Sebelum dilakukan proses pemodelan dengan algoritma *K-Means*, terlebih dahulu dilakukan perhitungan dengan menentukan nilai K atau jumlah *cluster* yang optimum dengan menggunakan metode *Elbow*.



Gambar 4.1 Visualisasi Teknik *Elbow*

5. Proses *Clustering* dengan *K-Means*

Setelah di tentukan atribut dan nilai K (jumlah *cluster*) yang akan digunakan, maka tahap selanjutnya yaitu melakukan pemodelan dengan algoritma *K-Means*.

6. Evaluasi Hasil *Cluster*

Hasil *clustering* dengan menggunakan algoritma *K-Means* dapat dievaluasi hasil cluster. Adapun teknik yang dilakukan dalam mengevaluasi adalah dengan menggunakan *Davies-Bouldin Index*.

7. Visualisasi dan Analisis Hasil *Clustering*

Setelah didapatkan hasil pengukuran *clustering*, maka selanjutnya dibuatkan visualisasi dalam bentuk grafik 3 dimensi.

4.3 Hasil Pengembangan Model

4.3.1 Seleksi Data

Pengembangan model *K-Means* menggunakan bahasa pemrograman *python* dengan memakai *Tools Notebook Colaboratory*.

1. Import Library Python

Berikut Library *Python* yang perlu di import agar dalam proses *k-means* bisa dijalankan.

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn.cluster import KMeans
import seaborn as sns
import plotly.graph_objs as go
```

2. Berikutnya pengaktifan *google drive* dengan perintah sebagai berikut.

```
from google.colab import drive
drive.mount('/content/drive')
```

```
#Baca Dataset
data = pd.read_excel('/content/drive/MyDrive/File Skripsi/dataset skripsi.xlsx')
data.head()
```

Setelah perintah tersebut dijalankan maka muncul hasil sebagai berikut:

	No	Pekerjaan	Penghasilan	Tanggungan
0	1	PETANI	550000	3
1	2	PETANI	600000	4
2	3	WIRASWASTA	2250000	4
3	4	PEGAWAI	3400000	5
4	5	PETANI	530000	2

4. Konversi Dataset *Excel* ke CSV, dataset dalam bentuk format file *excel* akan di konversi dalam bentuk CSV (*COMMA Saparated Values*) agar dapat digunakan pada pemodelan data dengan menggunakan perintah sebagai berikut:

```
#Konversi dataset Excel ke CSV
data.to_csv("/content/drive/MyDrive/Folder Skripsi/dataset asli.csv", index=False)
data.info()
```

Setelah dijalankan maka hasil yang diperoleh sebagai berikut:

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 765 entries, 0 to 764
Data columns (total 4 columns):
#   Column          Non-Null Count  Dtype
---  -
0    No              765 non-null   int64
1    Pekerjaan       765 non-null   object
2    Penghasilan     765 non-null   int64
3    Tanggungan     765 non-null   int64
dtypes: int64(3), object(1)
memory usage: 24.0+ KB
```

4.3.2 Transformasi Data

1. Menghitung banyaknya jumlah data pada atribut, sebelum dilakukan konversi dengan *coding* berikut ini.

```
data['Pekerjaan'].value_counts()
```

Berikut adalah hasil dari potongan *coding* diatas.

```
PETANI      207
WIRASWASTA  192
PEGAWAI     130
PNS         125
WIRAUSAHA   51
NELAYAN     34
LAINNYA     26
```

2. Selanjutnya mengkonversi semua atribut yang bertipe string atau non numerik ke numerik dimana harus dilakukan konversi nilai sehingga dapat digunakan dalam pemodelan K-Means. Berikut adalah perintah untuk konversi data atribut “Pekerjaan”:

Berikut hasil konversi nilai atribut:

```
# Konversi Data String ke Angka
data['Pekerjaan'].replace(['PETANI', 'NELAYAN', 'LAINNYA', 'PEGAWAI', 'PNS',
                          'WIRASWASTA', 'WIRAUSAHA'], [1,2,3,4,5,6,7],
                          inplace=True)
data.head()
```

Berikut adalah hasil potongan *coding* di atas:

	No	Pekerjaan	Penghasilan	Tanggungan
0	1	1	550000	3
1	2	1	600000	4
2	3	6	2250000	4
3	4	4	3400000	5
4	5	1	530000	2

3. Sebelum dilakukan pemodelan *K-Means*, kemudian menentukan kolom dataset

```
# Memilih/menentukan kolom dataset
x1 = data[['Pekerjaan', 'Penghasilan', 'Tanggungan']].values
x1
```

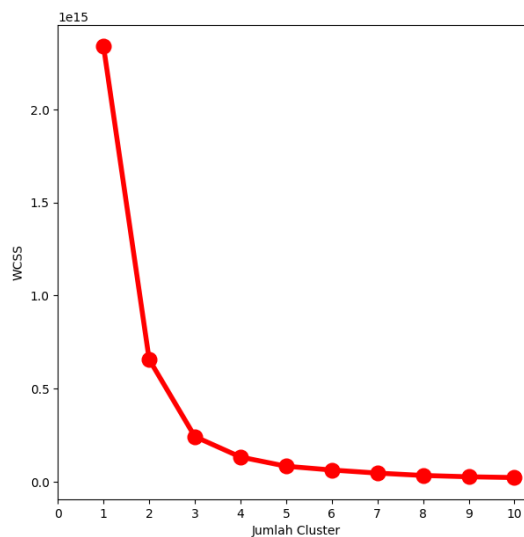
Hasil dari potongan *coding* diatas adalah sebagai berikut:

```
array([[ 1, 550000, 3],
       [ 1, 600000, 4],
       [ 6, 2250000, 4],
       ...,
       [ 6, 4500000, 3],
       [ 5, 3500000, 3],
       [ 2, 1500000, 3]])
```

4. Menentukan jumlah *cluster* yang optimum dengan teknik *Elbow* sebelum optimalisasi sebagai berikut:

```
# Menentukan jumlah cluster yang optimum dengan teknik Elbow
x = x1
WCSS = []
for i in range(1,11):
    model = KMeans(n_clusters = i, init = "k-means++", max_iter = 300,
                   n_init = 10, random_state = 0)
    model.fit(x)
    WCSS.append(model.inertia_)
fig = plt.figure(figsize = (7,7))
plt.plot(range(1,11),WCSS, linewidth=4, markersize=12,marker='o',color = 'red')
plt.xticks(np.arange(11))
plt.xlabel("Jumlah Cluster")
plt.ylabel("WCSS")
plt.show()
```

Berikut adalah hasil perintah dari *coding* di atas:



4.3.3 Pengoptimalan

1. Normalisasi data, guna dari menormalisasikan data yaitu agar data yang akan digunakan pada pemodelan dapat berkumpul sesuai pusat *cluster* masing-masing karena sebelumnya data yang tidak normalisasikan menyebar ke pusat *cluster* lain. Berikut adalah potongan *coding* untuk normalisasi data:

```
# Normalization
from sklearn.preprocessing import StandardScaler
scaler = StandardScaler()
data_scaled = scaler.fit_transform(x1)
```

```
data_scaled
```

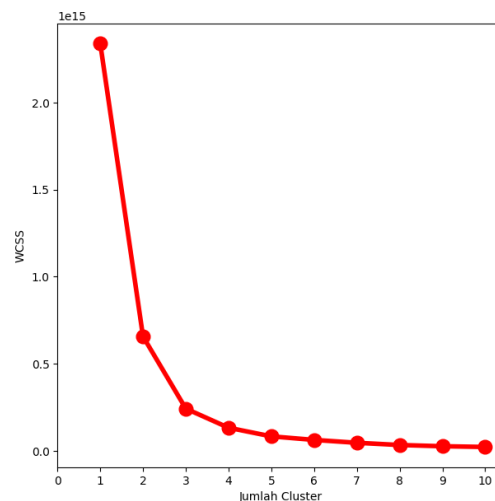
Berikut adalah hasil dari potongan *coding* di atas:

```
array([[ -1.39564704,  -1.37121944,  -0.01068589],
       [ -1.39564704,  -1.34261642,   1.1571292 ],
       [  0.98541894,  -0.39871702,   1.1571292 ],
       ...,
       [  0.98541894,   0.88841854,  -0.01068589],
       [  0.50920574,   0.31635829,  -0.01068589],
       [-0.91943384,  -0.8277622 ,  -0.01068589]])
```

2. kemudian setelah optimalisasi dilakukan jumlah *cluster* yang optimum dengan teknik *Elbow* akan seperti sebagai berikut:

```
# Menentukan jumlah cluster yang optimum dengan teknik Elbow
x = data_scaled
WCSS = []
for i in range(1,11):
    model = KMeans(n_clusters = i, init = "k-means++", max_iter = 300,
                   n_init = 10, random_state = 0)
    model.fit(x)
    WCSS.append(model.inertia_)
fig = plt.figure(figsize = (7,7))
plt.plot(range(1,11),WCSS, linewidth=4, markersize=12,marker='o',color = 'red')
plt.xticks(np.arange(11))
plt.xlabel("Jumlah Cluster")
plt.ylabel("WCSS")
plt.show()
```

berikut adalah hasil perintah di atas:



3. Pemodelan *K-Means*

Setelah didapatkan jumlah *cluster* yang optimum maka tahap selanjutnya adalah menerapkan dalam pemodelan *K-Means* seperti pada potongan *coding* dibawah ini:

```
# Pemodelan KMeans dengan jumlah cluster yang optimum
model = KMeans(n_clusters = 3, init = "k-means++", max_iter = 300, n_init = 10,
               random_state = 0)
y_clusters = model.fit_predict(x1)
```

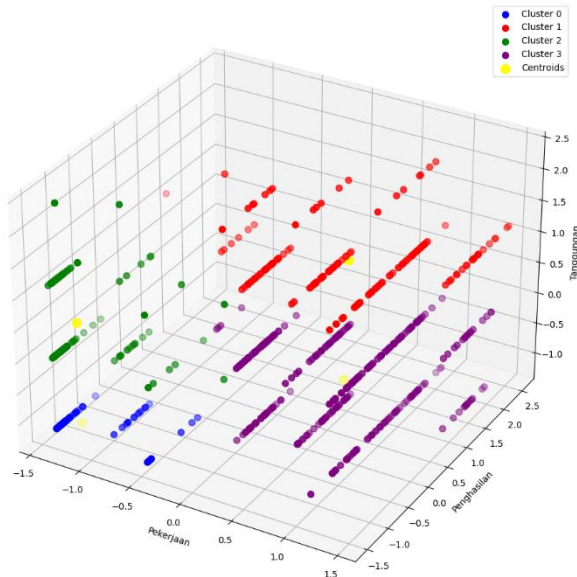
Setelah dilakukan Pemodelan *K-Means*, maka selanjutnya dibuatkan visualisasi 3d disetiap *cluster* sebagai berikut:

```
# Visualisasi hasil cluster dengan 3d scatterplot menggunakan matplotlib

fig = plt.figure(figsize = (12,12))
ax = fig.add_subplot(111, projection='3d')
ax.scatter(x[y_clusters == 0,0],x[y_clusters == 0,1],x[y_clusters == 0,2],
           s = 50 , color = 'blue', label = "Cluster 0")
ax.scatter(x[y_clusters == 1,0],x[y_clusters == 1,1],x[y_clusters == 1,2],
           s = 50 , color = 'red', label = "Cluster 1")
ax.scatter(x[y_clusters == 2,0],x[y_clusters == 2,1],x[y_clusters == 2,2],
           s = 50 , color = 'green', label = "Cluster 2")
ax.scatter(x[y_clusters == 3,0],x[y_clusters == 3,1],x[y_clusters == 3,2],
           s = 50 , color = 'purple', label = "Cluster 3")

ax.scatter(model.cluster_centers[:,0],model.cluster_centers[:,1],
           model.cluster_centers[:,2], s = 100, c = "yellow",
           label = "Centroids")
ax.set_xlabel('Pekerjaan')
ax.set_ylabel('Penghasilan')
ax.set_zlabel('Tanggungan')
ax.legend()
plt.show()
```

Berikut adalah hasil dari perintah di atas:



Berdasarkan gambar di atas dapat dilihat setiap *cluster* diwakili titik berwarna yaitu *cluster* 0 warna biru, *cluster* 1 warna merah, *cluster* 2 warna hijau, *cluster* 3 warna ungu, dan warna kuning adalah pusat *cluster*.

4. Menghitung skor *Davies-Bouldin Index*

Setelah dilakukan visualisasi dan mendapatkan hasil cluster yang optimal selanjutnya menghitung skor *Davies-Bouldin Index*

```
from sklearn.metrics import davies_bouldin_score

# Melakukan clustering dengan K-Means
kmeans = KMeans(n_clusters=4, random_state=42)
kmeans.fit(x)

# Menghitung skor Davies-Bouldin Index
score = davies_bouldin_score(x, kmeans.labels_)
print("Davies-Bouldin Index:", score)
```

Berikut adalah sebagai berikut:

Davies-Bouldin Index: 0.8002096210333453

4.3.4 Pusat *Cluster* Dan Jumlah Data

1. Pada Cetak pusat *cluster* ini dapat diketahui nilai dari masing-masing atribut, berikut adalah potongan *coding* untuk mencetak pusat *cluster*:

```
#Cetak Pusat Cluster
print(model.cluster_centers_)
```

Hasil output potongan *coding* diatas adalah sebagai berikut:

```
[[-1.25958613 -1.13717091 -1.17850098]
 [ 0.63565784  0.64922948  1.29568353]
 [-1.26985487 -1.14825266  0.44468855]
 [ 0.64300583  0.54109155 -0.55754794]]
```

2. Kemudian menampilkan hasil *cluster* dapat dilihat pada potongan *coding* berikut ini:

```
# Menampilkan hasil kluster
print(model.labels_)
# Menambahkan kolom "kluster" dalam data frame Data Nilai
data["Kluster"] = model.labels_
data
```

Berikut adalah hasil output dari potongan *coding* di atas:

```
[2 2 1 1 0 1 1 1 1 2 3 1 3 3 1 2 3 1 3 1 0 3 1 3 2 2 2 3 2 3 1 1 1 1 1 3 3
1 3 1 1 2 0 2 2 0 2 3 2 3 2 0 0 2 1 2 3 3 1 1 2 2 2 2 3 1 0 2 0 1 1 2 2
0 3 2 2 2 0 0 2 2 2 3 3 2 1 3 1 2 2 3 1 3 1 0 3 3 1 2 2 2 3 3 3 1 2 3 1 0
3 3 3 0 0 2 3 1 2 3 3 0 1 3 0 3 3 2 2 1 1 2 3 1 3 1 3 0 3 3 3 3 1 3 3 0 3
1 3 3 2 1 1 3 3 3 3 3 1 2 3 1 2 3 0 0 0 1 0 3 3 3 2 3 3 3 0 0 2 3 1 3 1 2
0 3 2 3 3 3 3 3 2 3 3 0 0 3 2 1 3 3 3 3 2 2 0 3 3 0 2 3 2 2 2 1 1 3 2 3
0 3 2 2 0 3 0 2 1 3 2 2 2 2 3 3 3 2 2 1 3 3 3 3 0 0 3 3 2 3 2 3 2 0 2 2 0
2 3 3 2 3 0 1 3 0 3 3 2 3 3 1 2 3 3 3 3 2 0 2 2 3 3 3 1 3 0 2 0 0 1 3 0
2 0 3 1 3 2 0 3 1 3 3 0 3 1 2 2 3 3 3 1 2 0 3 2 0 1 1 3 3 2 0 0 1 3 2 1
0 2 2 1 3 0 1 1 3 3 3 0 3 1 1 3 3 1 3 0 3 1 3 3 3 3 3 3 2 2 3 2 1 2 3 2
2 0 3 0 2 2 0 2 0 2 3 2 1 1 1 3 3 0 3 1 1 1 3 1 3 2 1 3 3 2 3 3 1 0 0 1
1 3 3 3 3 3 1 3 3 3 2 1 3 3 3 1 3 0 3 3 1 1 1 3 1 3 0 2 1 1 0 1 3 3 3 1 3
1 3 3 3 1 1 3 3 0 3 1 1 2 2 1 3 2 3 1 3 1 1 3 1 1 3 3 1 1 2 2 2 0 0 2 2
3 3 2 3 3 2 2 0 1 3 2 1 1 3 3 1 2 3 1 3 3 1 3 2 3 1 3 3 2 3 3 3 3 3 0 3 3
1 1 0 3 1 0 0 2 1 3 2 3 3 1 3 1 2 3 3 3 1 3 1 2 3 3 1 1 0 3 3 3 1 3 3 1 2
1 3 3 3 3 3 1 3 0 3 3 3 3 3 2 3 3 3 3 1 0 3 1 1 1 0 3 2 0 2 3 3 1 1 3 3
3 3 2 1 3 0 1 3 3 3 2 2 3 3 3 3 3 3 3 1 1 0 1 1 3 1 3 1 1 0 3 2 2 1 2 2 3
3 1 3 3 3 2 2 3 3 3 2 0 2 0 1 3 3 3 1 3 2 3 1 3 3 2 2 1 3 3 3 0 3 3 3 3
1 3 3 3 3 2 2 3 3 0 0 2 0 3 1 1 0 3 1 3 0 2 3 3 1 2 1 1 3 3 3 1 2 1 2 3 1
3 2 0 2 1 1 1 1 1 1 3 3 1 3 3 1 3 1 2 3 3 2 0 1 3 3 0 3 3 3 0 3 3 0 0 3
0 3 3 1 2 1 2 3 3 3 1 1 1 3 3 3 3 0 1 1 1 3 3 3 2]
```

	No	Pekerjaan	Penghasilan	Tanggungan	Kluster	Ket
0	1	1	550000	3	2	C3
1	2	1	600000	4	2	C3
2	3	6	2250000	4	1	C2
3	4	4	3400000	5	1	C2
4	5	1	530000	2	0	C1
...	...	...	...	...	...	...
760	761	7	5750000	4	1	C2
761	762	4	2500000	2	3	C4
762	763	6	4500000	3	3	C4
763	764	5	3500000	3	3	C4
764	765	2	1500000	3	2	C3

3. Keterangan lebih lanjut untuk nilai *cluster* bisa dibuatkan kondisi dengan potongan *coding* sebagai berikut:

```
conditions = [
    (data['Kluster']==0),
    (data['Kluster']==1),
    (data['Kluster']==2),
    (data['Kluster']==3)]
choices = ['C1','C2','C3','C4']
data['Ket'] = np.select(conditions, choices)
data
```

Berikut adalah hasil *output* dari perintah diatas:

	No	Pekerjaan	Penghasilan	Tanggungan	Kluster
0	1	1	550000	3	2
1	2	1	600000	4	2
2	3	6	2250000	4	1
3	4	4	3400000	5	1
4	5	1	530000	2	0
...	...	...	...	...	...
760	761	7	5750000	4	1
761	762	4	2500000	2	3
762	763	6	4500000	3	3
763	764	5	3500000	3	3
764	765	2	1500000	3	2

4. kemudian menghitung jumlah data per *cluster* yang dapat dilihat pada *coding* dibawah ini:

```
#Hitung Jumlah Data Per Cluster
data['Kluster'].value_counts()
```

Berikut adalah hasil dari potongan *coding* di atas:

```
3    331
1    177
2    159
0     98
```

5. Setelah semua perintah sudah dijalankan langkah terakhir yaitu menyimpan hasil *cluser* ke *Excel*.

```
#Simpan Hasil Cluster ke Excel
data.to_excel("/content/drive/MyDrive/Folder Skripsi/hasil cluster.xlsx",
              index=False)
```

BAB V

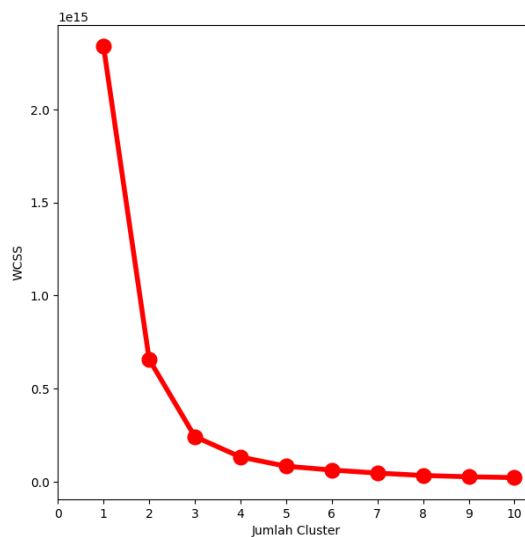
PEMBAHASAN PENELITIAN

5.1 Pembahasan Kinerja Model

Dalam pemodelan *K-Means* dalam pengelompokkan status mahasiswa dilakukan uji coba pemodelan sebanyak dua kali yaitu sebelum dilakukan optimalisasi dan setelah dilakukan optimalisasi.

1. Kinerja model sebelum optimalisasi

Pada pengujian pertama ini dalam penentuan jumlah *cluster* yang optimum menggunakan metode *Elbow* sebagaimana hasilnya dapat dilihat pada Gambar 5.1 dibawah ini:



Gambar 5.1 Visualisasi Teknik *Elbow* sebelum di normalisasikan

Berdasarkan Gambar 5.1 di atas menunjukkan bahwa jumlah *cluster* yang optimum yaitu sebanyak 3 *cluster*. Kemudian hasil *Elbow* di atas kembali diuji menggunakan dengan DBI (*Davies-Bouldin Index*) untuk mengukur kualitas *cluster*, adapun perhitungan DBI dapat dilihat pada Tabel 5.1 sebagai berikut:

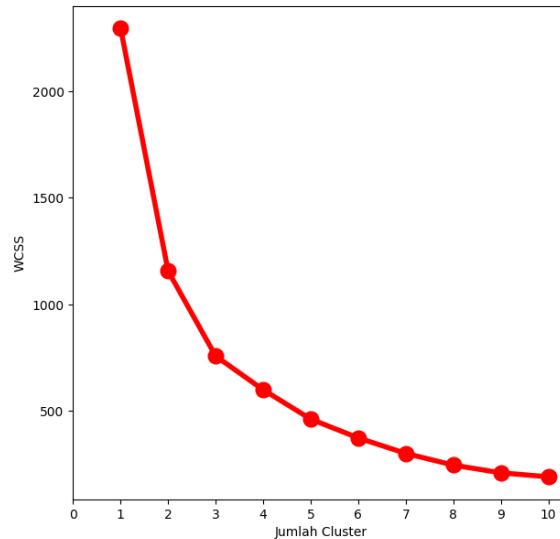
Tabel 5.1 Hasil Uji *Cluster Davies Bouldin Index* (DBI)

Jumlah <i>Cluster</i>	<i>Davies-Bouldin Index</i> (DBI)
2	0.5307
3	0.4627
4	0.4839
5	0.4628

Berdasarkan Tabel 5.1 di atas dapat dilihat bahwa jumlah *cluster* yang memiliki nilai DBI yang paling kualitas yang baik pada jumlah *cluster* 3 dengan nilai DBI sebesar 0.4627, dimana nilai DBI ini semakin mendekati nilai 0 semakin bagus.

2. Kinerja model setelah optimalisasi

Kemudian pada pengujian kedua ini dalam penentuan jumlah *cluster* yang optimum menggunakan metode *Elbow* sebagaimana hasilnya dapat dilihat pada Gambar 5.2 dibawah ini:

**Gambar 5.2** Visualisasi Teknik *Elbow* setelah di normalisasikan

Berdasarkan Gambar 5.2 di atas menunjukkan bahwa jumlah *cluster* yang optimum yaitu sebanyak 4 *cluster*. Kemudian hasil *Elbow* di atas kembali diuji menggunakan dengan DBI (*Davies-Bouldin Index*) untuk mengukur kualitas *cluster*, adapun perhitungan DBI dapat dilihat pada Tabel 5.2 sebagai berikut:

Tabel 5.2 Hasil Uji *Cluster Davies Bouldin Index (DBI)*

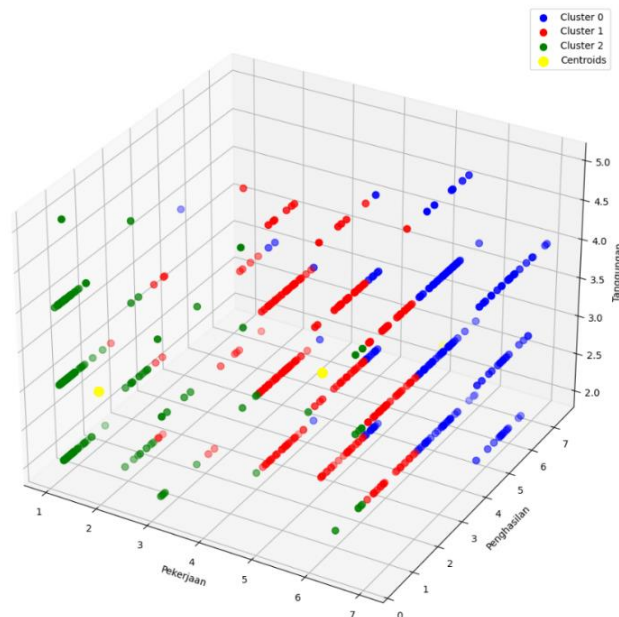
Jumlah <i>Cluster</i>	<i>Davies-Bouldin Index (DBI)</i>
2	0.8205
3	0.8974
4	0.8002
5	0.8891

Berdasarkan Tabel 5.2 di atas dapat dilihat bahwa jumlah *cluster* yang memiliki nilai DBI yang kualitas yang paling baik pada jumlah *cluster* 4 dengan nilai DBI sebesar 0.8002, dimana nilai DBI ini semakin mendekati nilai 0 semakin bagus.

5.2 Pembahasan Hasil Pemodelan

5.2.1 Hasil Pemodelan Sebelum Optimalisasi

Setelah dilakukan pemodelan berdasarkan jumlah cluster yang optimal sesuai dengan Gambar 5.1 di atas yaitu sebanyak 3 *cluster*, kemudian dilakukan visualisasi hasil *clustering* sebagaimana dapat dilihat pada Gambar 5.3 sebagai berikut:

**Gambar 5.3** Visualisasi Hasil *Clustering* sebelum optimalisasi

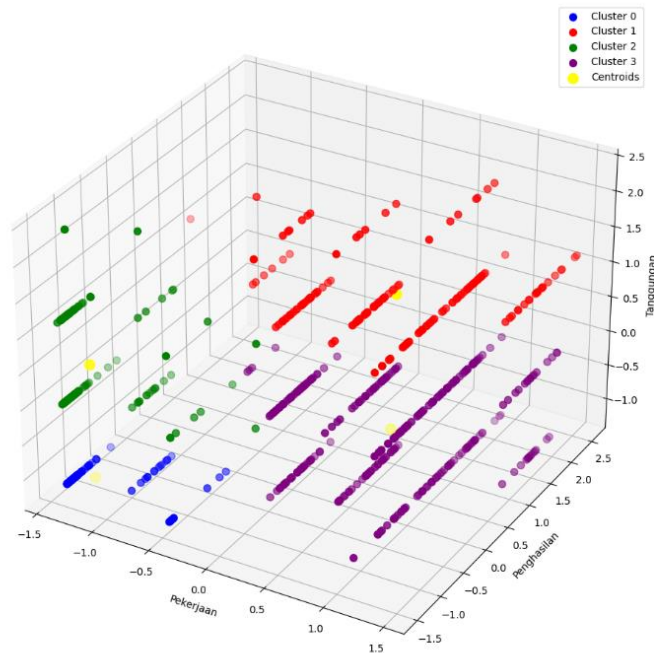
Berdasarkan Gambar 5.3 di atas dapat dilihat bahwa penyebaran pengelompokan data belum maksimal karena masih ada anggota *cluster* yang jauh dari pusat clusternya, sehingga hal ini masih perlu dilakukan optimalisasi atribut yaitu nilai atribut penghasilan perlu dilakukan normalisasi agar nilai semua atribut memiliki jarak yang sama. Adapun pembagian jumlah untuk setiap *cluster* dapat dilihat pada Tabel 5.3 dibawah ini:

Gambar 5.4 Pusat *Cluster* Sebelum Optimalisasi

<i>Cluster</i>	Titik <i>Centroid</i>			Jumlah Data
	Pekerjaan	Penghasilan	Tanggungan	
1	6.03	5.26	3.09	191
2	4.68	3.21	3.08	313
3	1.50	9.41	2.87	261

5.2.2 Hasil Pemodelan Sesudah Optimalisasi

Setelah dilakukan optimalisasi atribut dengan melakukan normalisasi atribut yang ada sehingga memiliki jarak yang sama. Setelah itu di hitung kembali jumlah *cluster* yang optimum dengan menggunakan teknik *Elbow*, adapun hasilnya dapat dilihat pada Gambar 5.2 di atas kemudian dilakukan pemodelan *K-Means* kembali dengan jumlah *cluster* optimum = 4, kemudian dilakukan visualisasi hasil *clustering* sebagaimana ditampilkan pada Gambar 5.3 berikut:



Gambar 5.5 Visualisasi Hasil *Clustering* Sesudah Optimalisasi

Berdasarkan Gambar 5.4 di atas dapat dilihat bahwa penyebaran pengelompokan data sudah lebih maksimal dimana disetiap anggota *cluster* sudah mendekati pusat *cluster* masing-masing, adapun hasil pembagian kelompok untuk masing-masing *cluster* dapat dilihat Tabel 5.4 berikut:

Tabel 5.3 Hasil *Clustering* sesudah Optimalisasi

<i>Cluster</i>	Titik <i>Centroid</i>			Jumlah Data
	Pekerjaan	Penghasilan	Tanggungan	
1	-1.26	-1.14	-1.18	98
2	0.64	0.65	1.19	159
3	-1.27	-1.15	0.44	177
4	0.64	0.54	-0.56	331

Dari tabel di atas dapat disimpulkan bahwa:

Cluster 1 adalah kelompok mahasiswa dengan status ekonomi yang paling rendah karena penghasilan yang kisaran sekitar $\leq Rp. 1,500.000$, pekerjaan rata-rata petani dan nelayan, serta tanggungan sebanyak 2 orang.

Cluster 2 adalah kelompok mahasiswa yang tingkat status ekonomi tinggi dengan penghasilan orang tua diatas Rp. 3.500.000, pekerjaan rata-rata orang tua adalah pegawai, PNS, dan wirausaha, serta jumlah tanggungan keluarga yaitu sebanyak 4 hingan 5 orang.

Cluster 3 adalah kelompok mahasiswa yang tingkat status ekonomi sedang degan rata-rata penghasilan orang tua yaitu mulai dari Rp. 500.000 – Rp. 2.500.000, pekerjaan rata-rata orang tua yaitu petani , nelayan dan lainnya, serta tanggungan keluarga sebanyak 2 orang.

Cluster 4 adalah kelompok mahasiswa dengan tingkat status ekonomi yang menengah karena pengahsilan pekerjaan orang tua yaitu mulai dari Rp. 2.000.000 – Rp. 5.000.000, pekerjaan orang tua rata-rata adalah sebagai PNS, wiraswasta, dan wiraswasta, serta tanggungan dalam keluarga 3 orang.

Pengelompokkan data status ekonomi mahasiswa berdasarkan hasil *clustering* dapat dilihat pada Tabel 5.5 berikut ini:

Tabel 5.4 Hasil *Clustering* Status Ekonomi Mahasiswa

No	Nama	Pekerjaan	Penghasilan	Tanggungan	Cluster	Ket	Keterangan
1	NURAFNI	1	550000	3	2	C3	Ekonomi Sedang
2	CITRA SALEH	1	600000	4	2	C3	Ekonomi Sedang
3	SAHRUN MOODUTO	6	2250000	4	1	C2	Ekonomi Tinggi
4	BELA SAFIRA BONDE	4	3400000	5	1	C2	Ekonomi Tinggi
5	MOH. ADIT PURNAWANSYAH DJALILU	1	530000	2	0	C1	Ekonomi Rendah
6	MEN WASIANGGE	6	4800000	5	1	C2	Ekonomi Tinggi
7	SYAHRUL RAMADHAN KALUKU	4	3200000	5	1	C2	Ekonomi Tinggi
8	I KOMANG CRISFANDU	4	2750000	5	1	C2	Ekonomi Tinggi
9	NADIA PUTRI OKTAVIANI THOMASONG	4	3500000	5	1	C2	Ekonomi Tinggi
10	APRILIA NGADI	1	1500000	4	2	C3	Ekonomi Sedang
11	SRI ARYAWATI LAILA MAYULU	4	2500000	3	3	C4	Ekonomi Menengan

No	Nama	Pekerjaan	Penghasilan	Tanggungan	Cluster	Ket	Keterangan
12	YOLANDA RAHMAN	5	4750000	4	1	C2	Ekonomi Tinggi
13	ANGGUN DHEA PUTRI AMRAIN	5	2500000	2	3	C4	Ekonomi Menengan
14	DJOKO RAMADHON	6	4500000	3	3	C4	Ekonomi Menengan
15	FARHAN HULUKATI	5	3200000	4	1	C2	Ekonomi Tinggi
16	AL IKSAN INGGO	5	3200000	4	2	C3	Ekonomi Sedang
17	MOHAMAD LARISKI Y. TALIB	1	750000	5	3	C4	Ekonomi Menengah
18	MUNAWIR RIZKY KARTO	5	3000000	2	1	C2	Ekonomi Tinggi
19	DELIA TIARA PUTERI TAMUU	5	3500000	4	2	C3	Ekonomi Sedang
20	PUTRI INTAN FITRIANI USMAN	4	2800000	3	1	C2	Ekonomi Tinggi
21	BILA NUGRAHANINGTYAS POLAPA	6	2000000	4	2	C3	Ekonomi Sedang
22	RAZLY R. PAKAYA	1	1500000	2	0	C1	Ekonomi Rendah
23	FITRI HARIYANI SUHARTONO	5	3750000	2	1	C2	Ekonomi Tinggi
24	RUSTIANI	6	5500000	5	2	C3	Ekonomi Sedang
25	NURAFNI	7	4700000	3	1	C2	Ekonomi Tinggi
26	DENI SAVITRI SIGILON	1	600000	3	3	C4	Ekonomi Menengah
27	YOLANDA HANAPI	1	750000	5	3	C4	Ekonomi Menengah
28	JUWITA N. ABAS	3	1800000	4	3	C4	Ekonomi Menengah
29	SITI KARMILA LARUFE	6	4500000	3	1	C2	Ekonomi Tinggi
30	HASAN R. TONA	2	1200000	3	3	C4	Ekonomi Menengah
...	...	...	...	...	...	...	...
765	ALDI	2	1500000	3	2	C3	Ekonomi Sedang

BAB VI

KESIMPULAN DAN SARAN

6.1 Kesimpulan

Berdasarkan hasil dari penelitian yang sudah di uraikan di atas tentang clustering status ekonomi mahasiswa dengan metode K-Means, maka dapat disimpulkan sebagai berikut:

1. Hasil penerapan metode *K-Means* dalam Clustering status ekonomi mahasiswa baru di Universitas Ichsan Gorontalo didapatkan 4 *cluster* yang optimum dengan nilai DBI sebesar 0.80.
2. Hasil analisa *clustering* untuk masing-masing *cluster* yaitu *cluster* 1 merupakan tingkat status ekonomi yang rendah, *cluster* 2 kelompok mahasiswa yang tingkat ekonomi tinggi, *cluster* 3 kelompok mahasiswa yang tingkat status ekonomi sedang, dan *cluster* 4 merupakan kelompok mahasiswa tingkat status ekonomi menengah.

6.2 Saran

Berdasarkan kesimpulan di atas, peneliti dapat memberikan saran kepada penelitian selanjutnya, yaitu:

1. Penelitian selanjutnya bisa menggunakan manambahkan dataset lebih banyak lagi, atau bisa melakukan model fitur seleksi yang lain dan melakukan normalisasi yang lain.
2. Dapat dikembangkan dengan menggunakan metode *clustering* lain seperti *K-Modes*, *K-Prototype* dan yang lainnya. Sehingga bisa membandingkan mana yang lebih baik.

DAFTAR PUSTAKA

- [1] O. Jeverson Taluke, L. Lesawengen, and E. A. A. Suwu, “Pengaruh Status Sosial Ekonomi Orang Tua Terhadap Tingkat Keberhasilan Mahasiswa Di Desa Buo Kecamatan Lolada Kabupaten Halmahera Barat.”
- [2] R. Pristian and F. Astuti, “PENGARUH STATUS SOSIAL EKONOMI ORANG TUA, LITERASI EKONOMI DAN LIFE STYLE TERHADAP PERILAKU KONSUMSI MAHASISWA JURUSAN PENDIDIKAN EKONOMI IKIP PGRI BOJONEGORO,” 2016.
- [3] J. A. Sains, D. Komputer, R. Dwi Romadhona, and A. A. Rofiq, “Algoritma K-Means Clustering Sebagai Penentuan Beasiswa PIP pada SMPN 9 Kota Blitar,” *Jurnal Aplikasi Sains, Informasi, Elektronika, dan Komputer*, vol. 5, no. 1, 2023, doi: 10.26905/jasiek.v5i1.10162.
- [4] R. Ishak, “Clustering Tingkat Pemahaman Dasar Mahasiswa Pada Pra-Perkuliahan Probabilitas Statistika Dengan Metode K-Means,” vol. 4, 2022.
- [5] A. Febrianto, S. Achmadi, and A. P. Sasmito, “PENERAPAN METODE K-MEANS UNTUK CLUSTERING PENGUNJUNG PERPUSTAKAAN ITN MALANG,” 2021.
- [6] A. Solichin and K. Khairunnisa, “Klasterisasi Persebaran Virus Corona (Covid-19) Di DKI Jakarta Menggunakan Metode K-Means,” *Fountain of Informatics Journal*, vol. 5, no. 2, p. 52, Oct. 2020, doi: 10.21111/fij.v5i2.4905.
- [7] Psycho Logic Mania, “Pengertian Status Sosial Ekonomi .” Accessed: Nov. 14, 2022. [Online]. Available: <https://www.psychologicmania.com/2012/pengertian-status-sosial-ekonomi.html?m=>
- [8] R. Setiawan, “PENERAPAN DATA MINING MENGGUNAKAN ALGORITMA K-MEANS CLUSTERING UNTUK MENENTUKAN

STRATEGI PROMOSI MAHASISWA BARU (Studi Kasus : Politeknik LP3I Jakarta),” 2016.

- [9] Rony Setiawan, “Apa Itu Data Mining Dan Bagaimana Metodenya?,” dicoding. Accessed: Nov. 14, 2022. [Online]. Available: <https://www.dicoding.com/blog/apa-itu-data-mining/>
- [10] Nagitec, “Apa Itu Data Mining.” Accessed: Nov. 14, 2022. [Online]. Available: <https://www.nagitec.com/apa-itu-data-mining>
- [11] Suyanto, *Data Mining Untuk Klasifikasi dan Klasterisasi*. Bandung: Informatika Bandung, 2017.
- [12] R. Ishak, “Volume 5 Nomor 2 Juli 2023 Penerapan Algoritma XGBoost untuk Seleksi Atribut pada K-Means dalam Clustering Penerima KIP Kuliah,” *Jambura Journal of Electrical and Electronics Engineering*, vol. 192.
- [13] skillplus.web.id, “Elbow Method.” Accessed: Nov. 15, 2022. [Online]. Available: <https://skillplus.web.id/elbow-method/>
- [14] J. Xiao, J. Lu, and X. Li, “Davies Bouldin Index based hierarchical initialization K-means,” *Intelligent Data Analysis*, vol. 21, no. 6, pp. 1327–1338, 2017, doi: 10.3233/IDA-163129.
- [15] B. Jumadi Dehotman Sitompul, O. Salim Sitompul, and P. Sihombing, “Enhancement Clustering Evaluation Result of Davies-Bouldin Index with Determining Initial Centroid of K-Means Algorithm,” in *Journal of Physics: Conference Series*, Institute of Physics Publishing, Jul. 2019. doi: 10.1088/1742-6596/1235/1/012015.
- [16] Nur Syahrudin Akbar K, “Jurnal Dasar Pemograman Python STMIK,” 2018.

LAMPIRAN

DATA YANG DIKUMPULKAN

No	NIM	Nama	Tempat Lahir	Kelamin	Tanggal Lahir
1.	E1122001	NURAFNI	POHUWATO	P	10/03/2005
2.	E1122002	CITRA SALEH	TERNATE	P	11/10/2004
3.	E1122003	SAHRUN MOODUTO	IRIAN JAYA	L	04/11/2001
4.	E1122004	BELA SAFIRA BONDE	TORAUT	P	28/04/2003
5.	E1122005	MOH. ADIT PURNAWANSYAH DJALILU	SUWAWA	L	16/04/2003
6.	E1122006	MEN WASIANGGE	MBUA	L	14/07/2001
7.	E1122007	NUR NITA M. LATIF	TELAGA	P	01/11/1989
8.	E1122008	SYAHRUL RAMADHAN KALUKU	GORONTALO	L	09/12/2001
9.	E1122009	I KOMANG CRISFANDU	BONGO IV	L	15/05/2003
10.	E1122010	NADIA PUTRI OKTAVIANI THOMASONG	GORONTALO	P	05/10/2004
11.	E1122011	APRILIA NGADI	GORONTALO	P	04/04/2004
12.	E1122012	SRI ARYAWATI LAILA MAYULU	TELAGA	P	22/12/2003
13.	E1122013	YOLANDA RAHMAN	GORONTALO	P	03/07/2002
14.	E1122014	ANGGUN DHEA PUTRI AMRAIN	GORONTALO	P	22/01/2004
15.	E1122015	DJOKO RAMADHON	PALU	L	12/01/1999
16.	E1122016	HERLINA MOHA	GORONTALO	P	05/04/1982
17.	E1122017	ASTRIA FATMAH YUNUS	GORONTALO	P	28/05/1989
18.	E1122018	FARHAN HULUKATI	GORONTALO	L	16/06/2003
19.	E1122019	AL IKSAN INGGO	TORAUT	L	29/12/2001
20.	E1122020	MOHAMAD LARISKI Y. TALIB	LOBUTO	L	29/04/2002
21.	E1122021	MUNAWIR RIZKY KARTO	MANADO	L	24/03/2002
22.	E1122022	DELIA TIARA PUTERI TAMUU	GORONTALO	P	13/09/2002
23.	E1122023	PUTRI INTAN FITRIANI USMAN	WONOSARI	P	03/04/2004
...	...	...	...	...	...
812	E1122025	RAZLY R. PAKAYA	MARISA	L	29/05/2001

No	Prodi	Pekerjaan Orang Tua	Penghasilan Orang Tua	Tanggungan
1	S1 AKUNTANSI	PETANI	Rp550.000	3 ORANG
2	S1 AKUNTANSI	PETANI	Rp600.000	4 ORANG
3	S1 AKUNTANSI	WIRASWASTA	Rp2.250.000	4 ORANG
4	S1 AKUNTANSI	PEGAWAI	Rp3.400.000	5 ORANG
5	S1 AKUNTANSI	PETANI	Rp530.000	2 ORANG
6	S1 AKUNTANSI	PEGAWAI	Rp4.800.000	5 ORANG
8	S1 AKUNTANSI	PEGAWAI	Rp3.200.000	5 ORANG
9	S1 AKUNTANSI	PEGAWAI	Rp2.750.000	5 ORANG
10	S1 AKUNTANSI	PETANI	Rp3.500.000	5 ORANG
11	S1 AKUNTANSI	PEGAWAI	Rp1.500.000	4 ORANG
12	S1 AKUNTANSI	PNS	Rp2.500.000	3 ORANG
13	S1 AKUNTANSI	PNS	Rp4.750.000	4 ORANG
14	S1 AKUNTANSI	WIRASWASTA	Rp2.500.000	2 ORANG
15	S1 AKUNTANSI	PNS	Rp4.500.000	3 ORANG
16	S1 AKUNTANSI	PETANI	Rp3.200.000	4 ORANG
17	S1 AKUNTANSI	PNS	Rp750.000	5 ORANG
18	S1 AKUNTANSI	PNS	Rp3.000.000	2 ORANG
19	S1 AKUNTANSI	PEGAWAI	Rp3.500.000	4 ORANG
20	S1 AKUNTANSI	WIRASWASTA	Rp2.800.000	3 ORANG
21	S1 AKUNTANSI	PETANI	Rp2.000.000	4 ORANG
22	S1 AKUNTANSI	PNS	Rp1.500.000	2 ORANG
23	S1 AKUNTANSI	WIRASWASTA	Rp3.750.000	2 ORANG
...	...	...	...	...
812	S1 AKUNTANSI	PEGAWAI	Rp4.700.000	3 ORANG

HASIL CLUSTER

No	Pekerjaan	Penghasilan	Tanggungan	Kluster	Ket	Keterangan
5	1	530000	2	0	C1	Ekonomi Rendah
21	1	1500000	2	0	C1	Ekonomi Rendah
43	1	520000	2	0	C1	Ekonomi Rendah
46	1	750000	2	0	C1	Ekonomi Rendah
52	1	650000	2	0	C1	Ekonomi Rendah
53	1	700000	2	0	C1	Ekonomi Rendah
68	1	650000	2	0	C1	Ekonomi Rendah
70	1	1270000	2	0	C1	Ekonomi Rendah
75	1	1500000	2	0	C1	Ekonomi Rendah
80	1	750000	2	0	C1	Ekonomi Rendah
81	1	870000	2	0	C1	Ekonomi Rendah
97	1	1100000	2	0	C1	Ekonomi Rendah
111	1	850000	2	0	C1	Ekonomi Rendah
115	1	650000	2	0	C1	Ekonomi Rendah
116	1	700000	2	0	C1	Ekonomi Rendah
123	1	800000	2	0	C1	Ekonomi Rendah
126	1	1200000	2	0	C1	Ekonomi Rendah
139	1	850000	2	0	C1	Ekonomi Rendah
147	1	625000	2	0	C1	Ekonomi Rendah
166	1	750000	2	0	C1	Ekonomi Rendah
167	1	500000	2	0	C1	Ekonomi Rendah
168	1	800000	2	0	C1	Ekonomi Rendah
170	1	1200000	2	0	C1	Ekonomi Rendah
178	1	800000	2	0	C1	Ekonomi Rendah
179	1	550000	2	0	C1	Ekonomi Rendah
186	3	450000	2	0	C1	Ekonomi Rendah
197	1	1000000	2	0	C1	Ekonomi Rendah
198	1	570000	2	0	C1	Ekonomi Rendah
208	2	2150000	2	0	C1	Ekonomi Rendah
211	2	1700000	2	0	C1	Ekonomi Rendah
223	1	750000	2	0	C1	Ekonomi Rendah
227	1	1500000	2	0	C1	Ekonomi Rendah
229	3	500000	2	0	C1	Ekonomi Rendah
247	3	2120000	2	0	C1	Ekonomi Rendah
248	1	780000	2	0	C1	Ekonomi Rendah
256	3	350000	2	0	C1	Ekonomi Rendah
259	3	1680000	2	0	C1	Ekonomi Rendah

No	Pekerjaan	Penghasilan	Tanggungan	Kluster	Ket	Keterangan
3	6	2250000	4	1	C2	Ekonomi Tinggi
4	4	3400000	5	1	C2	Ekonomi Tinggi
6	6	4800000	5	1	C2	Ekonomi Tinggi
7	4	3200000	5	1	C2	Ekonomi Tinggi
8	4	2750000	5	1	C2	Ekonomi Tinggi
9	4	3500000	5	1	C2	Ekonomi Tinggi
12	5	4750000	4	1	C2	Ekonomi Tinggi
15	5	3200000	4	1	C2	Ekonomi Tinggi
18	5	3500000	4	1	C2	Ekonomi Tinggi
20	6	2000000	4	1	C2	Ekonomi Tinggi
23	6	5500000	5	1	C2	Ekonomi Tinggi
31	4	2500000	4	1	C2	Ekonomi Tinggi
32	4	3750000	4	1	C2	Ekonomi Tinggi
33	6	4500000	5	1	C2	Ekonomi Tinggi
34	5	3800000	4	1	C2	Ekonomi Tinggi
35	7	5250000	4	1	C2	Ekonomi Tinggi
38	4	2750000	4	1	C2	Ekonomi Tinggi
40	5	3800000	4	1	C2	Ekonomi Tinggi
41	3	3400000	4	1	C2	Ekonomi Tinggi
55	5	4600000	5	1	C2	Ekonomi Tinggi
59	4	2700000	5	1	C2	Ekonomi Tinggi
60	5	3900000	4	1	C2	Ekonomi Tinggi
67	6	5550000	5	1	C2	Ekonomi Tinggi
71	5	3550000	4	1	C2	Ekonomi Tinggi
72	3	3800000	4	1	C2	Ekonomi Tinggi
88	5	4200000	5	1	C2	Ekonomi Tinggi
90	7	4500000	4	1	C2	Ekonomi Tinggi
94	6	5450000	4	1	C2	Ekonomi Tinggi
96	6	3750000	4	1	C2	Ekonomi Tinggi
100	7	5700000	4	1	C2	Ekonomi Tinggi
107	6	5500000	4	1	C2	Ekonomi Tinggi
110	6	3500000	4	1	C2	Ekonomi Tinggi
119	4	3000000	4	1	C2	Ekonomi Tinggi
124	4	3500000	4	1	C2	Ekonomi Tinggi
131	6	2780000	4	1	C2	Ekonomi Tinggi
132	4	3400000	4	1	C2	Ekonomi Tinggi
135	5	4500000	4	1	C2	Ekonomi Tinggi

No	Pekerjaan	Penghasilan	Tanggungan	Kluster	Ket	Keterangan
1	1	550000	3	2	C3	Ekonomi Sedang
2	1	600000	4	2	C3	Ekonomi Sedang
10	1	1500000	4	2	C3	Ekonomi Sedang
16	1	750000	5	2	C3	Ekonomi Sedang
25	1	600000	3	2	C3	Ekonomi Sedang
26	1	750000	5	2	C3	Ekonomi Sedang
27	3	1800000	4	2	C3	Ekonomi Sedang
29	2	1200000	3	2	C3	Ekonomi Sedang
42	1	800000	4	2	C3	Ekonomi Sedang
44	2	1500000	3	2	C3	Ekonomi Sedang
45	2	1750000	3	2	C3	Ekonomi Sedang
47	1	800000	3	2	C3	Ekonomi Sedang
49	2	1200000	3	2	C3	Ekonomi Sedang
51	2	1800000	3	2	C3	Ekonomi Sedang
54	4	1550000	4	2	C3	Ekonomi Sedang
56	1	1300000	3	2	C3	Ekonomi Sedang
61	1	1200000	4	2	C3	Ekonomi Sedang
62	1	500000	3	2	C3	Ekonomi Sedang
63	2	2110000	4	2	C3	Ekonomi Sedang
64	1	520000	4	2	C3	Ekonomi Sedang
65	1	1350000	3	2	C3	Ekonomi Sedang
69	1	800000	3	2	C3	Ekonomi Sedang
73	1	1500000	4	2	C3	Ekonomi Sedang
74	1	750000	3	2	C3	Ekonomi Sedang
77	1	1640000	3	2	C3	Ekonomi Sedang
78	2	1350000	5	2	C3	Ekonomi Sedang
79	1	600000	3	2	C3	Ekonomi Sedang
82	1	700000	4	2	C3	Ekonomi Sedang
83	1	500000	4	2	C3	Ekonomi Sedang
84	1	800000	4	2	C3	Ekonomi Sedang
87	1	850000	3	2	C3	Ekonomi Sedang
91	1	640000	3	2	C3	Ekonomi Sedang
92	1	580000	4	2	C3	Ekonomi Sedang
101	1	900000	3	2	C3	Ekonomi Sedang
102	1	750000	3	2	C3	Ekonomi Sedang
103	3	370000	4	2	C3	Ekonomi Sedang
108	3	450000	3	2	C3	Ekonomi Sedang
117	3	640000	3	2	C3	Ekonomi Sedang
120	1	1000000	3	2	C3	Ekonomi Sedang

No	Pekerjaan	Penghasilan	Tanggungan	Kluster	Ket	Keterangan
11	4	2500000	3	3	C4	Ekonomi Menengah
13	5	2500000	2	3	C4	Ekonomi Menengah
14	6	4500000	3	3	C4	Ekonomi Menengah
17	5	3000000	2	3	C4	Ekonomi Menengah
19	4	2800000	3	3	C4	Ekonomi Menengah
22	5	3750000	2	3	C4	Ekonomi Menengah
24	7	4700000	3	3	C4	Ekonomi Menengah
28	6	4500000	3	3	C4	Ekonomi Menengah
30	4	2800000	2	3	C4	Ekonomi Menengah
36	7	6500000	3	3	C4	Ekonomi Menengah
37	6	4250000	3	3	C4	Ekonomi Menengah
39	4	2800000	2	3	C4	Ekonomi Menengah
48	4	3500000	2	3	C4	Ekonomi Menengah
50	4	2500000	3	3	C4	Ekonomi Menengah
57	6	2500000	3	3	C4	Ekonomi Menengah
58	6	3600000	2	3	C4	Ekonomi Menengah
66	4	3200000	3	3	C4	Ekonomi Menengah
76	6	4500000	3	3	C4	Ekonomi Menengah
85	4	3500000	3	3	C4	Ekonomi Menengah
86	6	2250000	2	3	C4	Ekonomi Menengah
89	6	3350000	3	3	C4	Ekonomi Menengah
93	3	3240000	3	3	C4	Ekonomi Menengah
95	4	3100000	2	3	C4	Ekonomi Menengah
98	5	4300000	3	3	C4	Ekonomi Menengah
99	6	4670000	3	3	C4	Ekonomi Menengah
104	6	4500000	3	3	C4	Ekonomi Menengah
105	4	2500000	3	3	C4	Ekonomi Menengah
106	4	2600000	2	3	C4	Ekonomi Menengah
109	4	2350000	3	3	C4	Ekonomi Menengah
112	5	4300000	3	3	C4	Ekonomi Menengah
113	6	3450000	2	3	C4	Ekonomi Menengah
114	7	6700000	3	3	C4	Ekonomi Menengah
118	6	5540000	3	3	C4	Ekonomi Menengah
121	6	3500000	3	3	C4	Ekonomi Menengah
122	6	3750000	2	3	C4	Ekonomi Menengah
125	4	2500000	2	3	C4	Ekonomi Menengah
127	5	4700000	2	3	C4	Ekonomi Menengah
128	7	5600000	3	3	C4	Ekonomi Menengah
134	6	5300000	3	3	C4	Ekonomi Menengah

No	Nama	Pekerjaan	Penghasilan	Tanggungan	Cluster	Ket	Keterangan
1	NURAFNI	1	550000	3	2	C3	Ekonomi Sedang
2	CITRA SALEH	1	600000	4	2	C3	Ekonomi Sedang
3	SAHRUN MOODUTO	6	2250000	4	1	C2	Ekonomi Tinggi
4	BELA SAFIRA BONDE	4	3400000	5	1	C2	Ekonomi Tinggi
5	MOH. ADIT PURNAWANSYAH DJALILU	1	530000	2	0	C1	Ekonomi Rendah
6	MEN WASIANGGE	6	4800000	5	1	C2	Ekonomi Tinggi
7	SYAHRUL RAMADHAN KALUKU	4	3200000	5	1	C2	Ekonomi Tinggi
8	I KOMANG CRISFANDU	4	2750000	5	1	C2	Ekonomi Tinggi
9	NADIA PUTRI OKTAVIANI THOMASONG	4	3500000	5	1	C2	Ekonomi Tinggi
10	APRILIA NGADI	1	1500000	4	2	C3	Ekonomi Sedang
11	SRI ARYAWATI LAILA MAYULU	4	2500000	3	3	C4	Ekonomi Menengan
12	YOLANDA RAHMAN	5	4750000	4	1	C2	Ekonomi Tinggi
13	ANGGUN DHEA PUTRI AMRAIN	5	2500000	2	3	C4	Ekonomi Menengan
14	DJOKO RAMADHON	6	4500000	3	3	C4	Ekonomi Menengan
15	FARHAN HULUKATI	5	3200000	4	1	C2	Ekonomi Tinggi
16	AL IKSAN INGGO	1	750000	5	2	C3	Ekonomi Sedang
17	MOHAMAD LARISKI Y. TALIB	5	3000000	2	3	C4	Ekonomi Menengan
18	MUNAWIR RIZKY KARTO	5	3500000	4	1	C2	Ekonomi Tinggi
19	DELIA TIARA PUTERI TAMUU	4	2800000	3	3	C4	Ekonomi Menengan
20	PUTRI INTAN FITRIANI USMAN	6	2000000	4	1	C2	Ekonomi Tinggi
...	...	...	...	...	...	...	...
765	ALDI	2	1500000	3	2	C3	Ekonomi Sedang

KODE PROGRAM

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn.cluster import KMeans
import seaborn as sns
import plotly.graph_objs as go

from google.colab import drive
drive.mount('/content/drive')

#Baca Dataset
data = pd.read_excel('/content/drive/MyDrive/Folder
Skripsi/dataset asli.xlsx')
data.head()

#Konversi dataset Excel ke CSV
data.to_csv("/content/drive/MyDrive/Folder Skripsi/dataset
asli.csv", index=False)
data.info()

data.head()

data['Pekerjaan'].value_counts()

# Konversi Data String ke Angka
data['Pekerjaan'].replace(['PETANI', 'NELAYAN', 'LAINNYA', 'PE
GAWAI', 'PNS',
                        'WIRASWASTA', 'WIRAUSAHA'],
[1,2,3,4,5,6,7],
                        inplace=True)
data.head()

data.info()

#Simpan Hasil Konversi
data.to_excel("/content/drive/MyDrive/Folder
Skripsi/dataset angka.xlsx", index=False)

#Baca Dataset
data = pd.read_excel('/content/drive/MyDrive/Folder
Skripsi/dataset angka.xlsx')
```

```

# Memilih/menentukan kolom dataset
x1 = data[['Pekerjaan','Penghasilan','Tanggungan']].values
x1

# Menentukan jumlah cluster yang optimum dengan teknik
Elbow
x = x1
WCSS = []
for i in range(1,11):
    model = KMeans(n_clusters = i, init = "k-means++",
max_iter = 300,
                    n_init = 10, random_state = 0)
    model.fit(x)
    WCSS.append(model.inertia_)
fig = plt.figure(figsize = (7,7))
plt.plot(range(1,11),WCSS, linewidth=4,
markersize=12,marker='o',color = 'red')
plt.xticks(np.arange(11))
plt.xlabel("Jumlah Cluster")
plt.ylabel("WCSS")
plt.show()

# Pemodelan KMeans dengan jumlah cluster yang optimum
model = KMeans(n_clusters = 3, init = "k-means++", max_iter
= 300, n_init = 10,
                random_state = 0)
y_clusters = model.fit_predict(x1)

# Visualisasi hasil cluster dengan 3d scatterplot
menggunakan matplotlib

fig = plt.figure(figsize = (12,12))
ax = fig.add_subplot(111, projection='3d')
ax.scatter(x[y_clusters == 0,0],x[y_clusters ==
0,1],x[y_clusters == 0,2],
           s = 50 , color = 'blue', label = "Cluster 0")
ax.scatter(x[y_clusters == 1,0],x[y_clusters ==
1,1],x[y_clusters == 1,2],
           s = 50 , color = 'red', label = "Cluster 1")
ax.scatter(x[y_clusters == 2,0],x[y_clusters ==
2,1],x[y_clusters == 2,2],

```

```

        s = 50 , color = 'green', label = "Cluster 2")

ax.scatter(model.cluster_centers_[:,0],model.cluster_center
s_[:,1],
           model.cluster_centers_[:,2], s = 100, c =
"yellow",
           label = "Centroids")
ax.set_xlabel('Pekerjaan')
ax.set_ylabel('Penghasilan')
ax.set_zlabel('Tanggungan')
ax.legend()
plt.show()

```

```

from sklearn.metrics import davies_bouldin_score

# Melakukan clustering dengan K-Means
kmeans = KMeans(n_clusters=3, random_state=42)
kmeans.fit(x1)

# Menghitung skor Davies-Bouldin Index
score = davies_bouldin_score(x1, kmeans.labels_)
print("Davies-Bouldin Index:", score)

```

Pengoptimalan

```

# Normalization
from sklearn.preprocessing import StandardScaler
scaler = StandardScaler()
data_scaled = scaler.fit_transform(x1)

data_scaled

# Menentukan jumlah cluster yang optimum dengan teknik
Elbow
x = data_scaled
WCSS = []
for i in range(1,11):
    model = KMeans(n_clusters = i, init = "k-means++",
max_iter = 300,
                    n_init = 10, random_state = 0)
    model.fit(x)

```



```

        WCSS.append(model.inertia_)
fig = plt.figure(figsize = (7,7))
plt.plot(range(1,11),WCSS, linewidth=4,
markersize=12,marker='o',color = 'red')
plt.xticks(np.arange(11))
plt.xlabel("Jumlah Cluster")
plt.ylabel("WCSS")
plt.show()

# Pemodelan KMeans dengan jumlah cluster yang optimum
model = KMeans(n_clusters = 4, init = "k-means++", max_iter
= 300, n_init = 10, random_state = 0)
y_clusters = model.fit_predict(x)

# Visualisasi hasil cluster dengan 3d scatterplot
menggunakan matplotlib

fig = plt.figure(figsize = (12,12))
ax = fig.add_subplot(111, projection='3d')
ax.scatter(x[y_clusters == 0,0],x[y_clusters ==
0,1],x[y_clusters == 0,2], s = 50 , color = 'blue', label =
"Cluster 0")
ax.scatter(x[y_clusters == 1,0],x[y_clusters ==
1,1],x[y_clusters == 1,2], s = 50 , color = 'red', label =
"Cluster 1")
ax.scatter(x[y_clusters == 2,0],x[y_clusters ==
2,1],x[y_clusters == 2,2], s = 50 , color = 'green', label
= "Cluster 2")
ax.scatter(x[y_clusters == 3,0],x[y_clusters ==
3,1],x[y_clusters == 3,2], s = 50 , color = 'purple', label
= "Cluster 3")
#ax.scatter(x[y_clusters == 4,0],x[y_clusters ==
4,1],x[y_clusters == 4,2], s = 50 , color = 'purple', label
= "Cluster 4")
ax.scatter(model.cluster_centers[:,0],model.cluster_center
s[:,1],model.cluster_centers[:,2], s = 100, c = "yellow",
label = "Centroids")
ax.set_xlabel('Pekerjaan')
ax.set_ylabel('Penghasilan')
ax.set_zlabel('Tanggungan')
ax.legend()

```

```

plt.show()

from sklearn.metrics import davies_bouldin_score
# Melakukan clustering dengan K-Means
kmeans = KMeans(n_clusters=5, random_state=42)
kmeans.fit(x)

# Menghitung skor Davies-Bouldin Index
score = davies_bouldin_score(x, kmeans.labels_)
print("Davies-Bouldin Index:", score)

# Pemodelan KMeans dengan jumlah cluster yang optimum
model = KMeans(n_clusters = 4, init = "k-means++", max_iter
= 300, n_init = 10, random_state = 0)
y_clusters = model.fit_predict(x)

#Cetak Pusat Cluster
print(model.cluster_centers_)

# Menampilkan hasil kluster
print(model.labels_)
# Menambahkan kolom "kluster" dalam data frame Data Nilai
data["Kluster"] = model.labels_
data

conditions = [
    (data['Kluster']==0),
    (data['Kluster']==1),
    (data['Kluster']==2),
    (data['Kluster']==3)]
choices = ['C1', 'C2', 'C3', 'C4']
data['Ket'] = np.select(conditions, choices)
data

#Hitung Jumlah Data Per Cluster
data['Kluster'].value_counts()

#Simpan Hasil Cluster ke Excel
data.to_excel("/content/drive/MyDrive/Folder Skripsi/hasil
cluster.xlsx",
              index=False)

```



**KEMENTERIAN PENDIDIKAN, KEBUDAYAAN,
RISET DAN TEKNOLOGI
UNIVERSITAS ICHSAN
(UNISAN) GORONTALO**

SURAT KEPUTUSAN MENDIKNAS RI NO. 84/D/O/2001
JL. Achmad Nadjamuddin No. 17 Telp. (0435) 829975 Fax (0435) 829976 Gorontalo

SURAT KETERANGAN
NO : 167/UNISAN-G/SKT/XI/2023

Yang bertanda tangan dibawah ini :

Nama : Amiruddin, S.Kom., M.Kom, MCF
NIDN : 0910097601
Jabatan : Wakil Rektor Bidang Akademik, Tata Kelola & Sistem Informasi

Menyatakan,

Nama : Ana Fitriana
Nim : T3119077
Prodi : Teknik Informatika
Fakultas : Ilmu Komputer

Bahwa yang bersangkutan diberikan izin penelitian di Universitas Ichsan Gorontalo dengan judul penelitian **"CLUSTERING STATUS EKONOMI MAHASISWA BARU MENGGUNAKAN METODE K-MEANS"** Kepada calon peneliti diharapkan:

1. Mematuhi segala ketentuan dan peraturan yang berlaku di Universitas Ichsan Gorontalo
2. Data penelitian bersifat rahasia dan tidak di izinkan menyampaikan informasi kepada pihak-pihak lain yang tidak berkepentingan.

Demikian surat keterangan ini dibuat untuk dapat dipergunakan sebagaimana mestinya,

Gorontalo, 21 November 2023

An. Rektor

Wakil Rektor Bidang Akademik, Tata Kelola &
Sistem Informasi



Amiruddin, S.Kom., M.Kom, MCF
NIDN. 0910097601

Tembusan:

1. Rektor Universitas Ichsan Gorontalo
2. Wakil Rektor I Bidang Akademik, Tata Kelola & Sistem Informasi
3. Mahasiswa yang bersangkutan
4. Arsip



KEMENTERIAN PENDIDIKAN, KEBUDAYAAN, RISET, DAN TEKNOLOGI

UNIVERSITAS ICHSAN GORONTALO

FAKULTAS ILMU KOMPUTER

SURAT KEPUTUSAN MENDIKNAS RI NOMOR 84/D/O/2001

Jl. Achmad Najamuddin No. 17 Telp. (0435) 829975 Fax (0435) 829976 Gorontalo

SURAT REKOMENDASI BEBAS PLAGIASI

No. 485/FIKOM-UIG/R/XI/2023

Yang bertanda tangan di bawah ini :

Nama : Irvan Abraham Salihi, M.Kom
NIDN : 0928028101
Jabatan : Dekan Fakultas Ilmu Komputer

Dengan ini menerangkan bahwa :

Nama Mahasiswa : Ana Fitriana
NIM : T3119077
Program Studi : Teknik Informatika (S1)
Fakultas : Fakultas Ilmu Komputer
Judul Skripsi : Clustering Status Ekonomi Mahasiswa Baru
Menggunakan Metode K-Means

Sesuai hasil pengecekan tingkat kemiripan skripsi melalui aplikasi **Turnitin** untuk judul skripsi di atas diperoleh hasil *Similarity* sebesar **28%**, berdasarkan Peraturan Rektor No. 32 Tahun 2019 tentang Pendeteksian Plagiat pada Setiap Karya Ilmiah di Lingkungan Universitas Ichsan Gorontalo dan persyaratan pemberian surat rekomendasi verifikasi calon wisudawan dari LLDIKTI Wil. XVI, bahwa batas kemiripan skripsi maksimal 30%, untuk itu skripsi tersebut di atas dinyatakan **BEBAS PLAGIASI** dan layak untuk diujikan.

Demikian surat rekomendasi ini dibuat untuk digunakan sebagaimana mestinya.

Mengetahui
Dekan,

Irvan Abraham Salihi, M.Kom
NIDN. 0928028101

Gorontalo, 20 November 2023
Tim Verifikasi,

Zulfrianto Y. Lamasiqi, M.Kom
NIDN. 0914089101

Terlampir :
Hasil Pengecekan Turnitin

PAPER NAME

SKRIPSI_T3119077_ANA_FITRIANA.pdf
anafitriana959@gmail.co

AUTHOR

ANA FITRIANA

m

WORD COUNT 5714

Words

CHARACTER COUNT 33063

Characters

PAGE COUNT 44

Pages

FILE SIZE

951.0KB

SUBMISSION DATE

Nov 18, 2023 1:11 PM GMT+8

REPORT DATE

Nov 18, 2023 1:12 PM GMT+8

● 28% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

- 27% Internet database19% Publications
- Crossref databaseCrossref Posted Content
- 1% Submitted Works database
- database
- database

● Excluded from Similarity Report

- Bibliographic materialQuoted material
- Cited materialSmall Matches (Less than 10 words)
-



**KEMENTERIAN PENDIDIKAN, KEBUDAYAAN, RISET, DAN TEKNOLOGI
UNIVERSITAS ICHSAN GORONTALO**

FAKULTAS ILMU KOMPUTER

UPT. PERPUSTAKAAN FAKULTAS

SK. MENDIKNAS RI NO. 84/D/0/2001

Jl. Achmad Nadjamuddin No.17 Telp(0435) 829975 Fax. (0435) 829976 Gorontalo

SURAT KETERANGAN BEBAS PUSTAKA

No : 012/Perpustakaan-Fikom/XI/2023

Perpustakaan Fakultas Ilmu Komputer (FIKOM) Universitas Ichsan Gorontalo dengan ini menerangkan bahwa :

Nama Anggota : Ana Fitriana

No. Induk : T3119077

No. Anggota : M202374

Terhitung mulai hari, tanggal : Jumat, 10 November 2023, dinyatakan telah bebas pinjam buku dan koleksi perpustakaan lainnya.

Demikian keterangan ini di buat untuk di pergunakan sebagaimana mestinya.



Gorontalo, 10 November 2023

**Mengetahui,
Kepala Perpustakaan**

Apriyanto Alhamad, M.Kom

NIDN : 0924048601

RIWAYAT HIDUP PENELITI



Nama : Ana Fitriana
 NIM : T3119077
 Tempat, Tgl Lahir : Molosipat, 10 Januari 2000
 Pekerjaan : Mahasiswa
 Agama : Islam
 Jenis Kelamin : Perempuan
 Fakultas/Jurusan : Ilmu Komputer/Teknik
 Informatika
 Konsentrasi : Software Engineering
 Email : anafitriana959@gmail.com
 Alamat : Jl. Achmad Nadjamuddin

Riwayat Pendidikan:

1. SD Negeri Anutapura (2006 – 2012)
2. Madrasah Tsanawiyah Negeri 3 Parigi (2012 – 2015)
3. SMA Negeri 1 Bolano Lambunu (2015 – 2018)
4. Universitas Ichsan Gorontalo (2019 – Sekarang)

