

***CLUSTERING* TINGKAT EKONOMI MAHASISWA  
CALON PENERIMA KARTU INDONESIA PINTAR  
(KIP) KULIAH DENGAN METODE *K-MEANS***

(Studi Kasus: Universitas Ichsan Gorontalo)

**Oleh**

**MAIMUN S. SOMPA**

**T3115070**

**SKRIPSI**

Untuk memenuhi salah satu syarat ujian  
guna memperoleh gelar Sarjana



**PROGRAM SARJANA  
FAKULTAS ILMU KOMPUTER  
UNIVERSITAS ICHSAN GORONTALO  
GORONTALO**

**2022**

**LEMBAR PERSETUJUAN SKRIPSI**  
**CLUSTERING TINGKAT EKONOMI MAHASISWA**  
**CALON PENERIMA KARTU INDONESIA PINTAR**  
**(KIP) KULIAH DENGAN METODE K-MEANS**  
( Studi Kasus : Universitas Ichsan Gorontalo )

Oleh

MAIMUN S. SOMPA

T3115070


**SKRIPSI**

Untuk memenuhi salah satu syarat ujian  
guna memperoleh gelar Sarjana Program Studi Teknik Informatika.

Ini telah disetujui oleh pembimbing

Gorontalo, Juni 2022

Pembimbing I

  
**Rezqiwati Ishak M. Kom**  
**NIDN: 0903087901**

Pembimbing II

  
**Yulianty Lasena, M. Kom**  
**NIDN: 0907078603**

 Dipindai dengan CamScanner

Dekan Fakultas Ilmu Komputer  
Studi

Ketua Program

**LEMBAR PENGESAHAN SKRIPSI**  
**CLUSTERING TINGKAT EKONOMI MAHASISWA**  
**CALON PENERIMA KARTU INDONESIA PINTAR**  
**(KIP) KULIAH DENGAN METODE K-MEANS**

OLEH  
MAIMUN S. SOMPA

T3115070

Di Periksa oleh Panitia Ujian Strata Satu (S1)  
Universitas Ichsan Gorontalo  
Gorontalo, Juni 2022

1. Ketua Penguji  
Amiruddin, M.Kom
2. Anggota Penguji  
Sudirman S. Panna, M.Kom
3. Anggota Penguji  
Roys Pakaya, M.Kom
4. Anggota Penguji  
Rezqiwati Ishak, M.Kom
5. Anggota Penguji  
Yulianty Lasena, M.Kom











Mengetahui :

  
Dekan Fakultas Ilmu Komputer  
**Lorry Karim, M.Kom**  
NIDN.0918077302

  
Ketua Program Studi  
**Sudirman S. Panna**  
NIDN.0924038205

## LEMBAR 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 penelitian saya sendiri, tanpa bantuan pihak lain, kecuali dari arahan Tim Pembimbing.
3. Dalam karya tulis ( Skripsi ) saya ini tidak terdapat karya atau pendapat yang telah dipublikasi orang lain, kecuali serta tertulis dicantumkan sebagai acuan/ sitasi dalam naskah dan dicantumkan pula dalam daftar pustaka.
4. Pernyataan isi saya buat dengan sesungguhnya dan apabila dikemudian 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 normanorma yang berlaku di Universitas Ichsan Gorontalo.

Gorontalo, Juni 2022

Yang membuat pernyataan

  
  
Maimun S. Sompia

## ABSTRACT

***MAIMUN S. SOMPA. T3115070. THE CLUSTERING OF THE ECONOMIC LEVEL OF PROSPECTIVE RECIPIENTS OF THE SMART INDONESIA CARD (KIP) FOR HIGHER EDUCATION USING THE K-MEANS METHOD***

The Smart Indonesia Card (KIP) for Higher Education by the government is under the auspices of the Ministry of Education and Culture. The Smart Indonesia Card (KIP) for Higher Education aims to help provide tuition assistance, especially for poor students to continue their studies. It prevents children from dropping out of education. Universitas Ichsan Gorontalo is one of the private universities granted a quota of the Smart Indonesia Card (KIP) for Higher Education. The limited number of student admissions (quota) of the Smart Indonesia Card (KIP) for Higher Education requires special attention in determining the right students as recipients on target to get the number of quotas that are not commensurate with the number of applicants. In seeing that, clusters are carried out based on the economic level of students to get a group of students prioritized to get the Smart Indonesia Card (KIP) for Higher Education. The K-Means method gets clustering results using the Elbow technique, namely 5 clusters. The results of clustering for each cluster indicate that Cluster 1 is a group of students with medium economic level and taken the second priority for recipients of assistance. Cluster 2 is a group of students with low economic levels and becomes the first priority of recipients of assistance. Cluster 3 is a group of students with middle to high economic levels and becomes the third priority for recipients of assistance. Cluster 4 is a group of students with middle economic level and become the fourth priority for recipients of assistance. Cluster 5 is a group of students with middle to upper economic levels and is the fifth priority for recipients of assistance.

Keywords: The Smart Indonesia Card (KIP) for Higher Education, Clustering, Elbow, K-Means





## ABSTRAK

### **MAIMUN S. SOMPA. T3115070. CLUSTERING TINGKAT EKONOMI MAHASISWA CALON PENERIMA KARTU INDONESIA PINTAR (KIP) KULIAH DENGAN METODE *K-MEANS***

Kartu Indonesia Pintar (KIP) Kuliah diluncurkan oleh pemerintah dibawah naungan Kementerian Pendidikan dan Kebudayaan. KIP kuliah bertujuan untuk membantu menyediakan bantuan Kuliah terfkhusus untuk siswa miskin sehingga dapat melanjutkan kuliahnya dengan sasaran untuk mencegah anak putus sekolah. Universitas Ichsan Gorontalo merupakan salah satu Perguruan Tinggi swasta yang diberikan kuota KIP Kuliah, terbatasnya jumlah penerimaan (Kuota) mahasiswa KIP Kuliah yang akan diterima membutuhkan perhatian khusus dalam menetapkan mahasiswa yang tepat sasaran untuk mendapatkan jumlah kuota yang tidak sepadan dengan jumlah pemohon, maka dilakukan *cluster* terhadap tingkat ekonomi mahasiswa demi mendapatkan kelompok mahasiswa yang diprioritaskan untuk mendapatkan bantuan KIP Kuliah menggunakan metode K-Means mendapat hasil clustering menggunakan teknik Elbow sebanyak 5 *cluster*. Hasil clustering untuk masing-masing cluster yaitu cluster 1 merupakan kelompok mahasiswa yang tingkat ekonominya sedang dan menjadi prioroitas kedua penerima bantuan, cluster 2 merupakan kelompok mahasiswa yang tingkat ekonominya kecil dan menjadi prioroitas pertama penerima bantuan, cluster 3 merupakan kelompok mahasiswa yang tingkat ekonominya menengah ke bawah dan menjadi prioroitas ketiga penerima bantuan, cluster 4 merupakan kelompok mahasiswa yang tingkat ekonominya menengah dan menjadi prioroitas keempat penerima bantuan, dan cluster 5 merupakan kelompok mahasiswa yang tingkat ekonominya menengah ke atas dan menjadi prioroitas kelima penerima bantuan.

Kata kunci: KIP, Clustering, Elbow, K-Means



## KATA PENGANTAR

Assalamualaikum warahmatullahi wabarakatuh.

Alhamdulillah penulis dapat menyelesaikan Skripsi dengan judul **“Clustering Tingkat Ekonomi Mahasiswa Calon Penerima Kartu Indonesia Pintar (KIP) Kuliah Dengan Metode K-Means”**.

untuk memenuhi salah satu syarat penyusunan Skripsi 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 mengucapkan banyak terima kasih dan penghargaan yang 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. Pak Jorry Karim, 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 Universitas Ichsan Gorontalo;
5. Ibu Irma Surya Kumala Idris, M.Kom, selaku Wakil Dekan II Bidang Administrasi Umum dan Keuangan Fakultas Ilmu Komputer Universitas Ichsan
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. Ibu Yulianty Lasena, 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 disiplin ilmu kepadva penulis;

10. Kedua Orang Tua saya yang tercinta, atas segala kasih sayang, jerih payah dan doa restunya dalam membesarkan dan mendidik penulis;
11. Rekan-rekan seperjuangan yang telah banyak memberikan bantuan dan moril yang sangat besar kepada penulis;
12. Kepada semua pihak yang ikut membantu dalam penyelesaian Skripsi ini yang tak sempat penulis sebutkan satu-persatu.

Semoga Allah SWT, melimpahkan balasan atas jasa-jasa mereka kepada kami. Penulis menyadari sepenuhnya bahwa dalam penyusunan ini masih jauh dari kesempurnaan dan masih banyak 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. Aamiinn.

Gorontalo, Juni 2022

Maimun S. Sompa



## DAFTAR ISI

	<b>Halaman</b>
<b>LEMBAR PERSETUJUAN .....</b>	<b>ii</b>
<b>LEMBAR PENGESAHAN .....</b>	<b>iii</b>
<b>LEMBAR PERNYATAAN .....</b>	<b>iv</b>
<b>ABSTRACT .....</b>	<b>v</b>
<b>ABSTRAK .....</b>	<b>vi</b>
<b>KATA PENGANTAR.....</b>	<b>vii</b>
<b>DAFTAR ISI.....</b>	<b>viii</b>
<b>DAFTAR TABEL .....</b>	<b>ix</b>
<b>DAFTAR GAMBAR.....</b>	<b>x</b>
<b>BAB I PENDAHULUAN</b>	
1.1. Latar Belakang Masalah .....	1
1.2. Identifikasi Masalah .....	3
1.3. Rumusan Masalah .....	4
1.4. Tujuan Penelitian.....	4
1.5. Manfaat Penelitian.....	4
<b>BAB II LANDASAN TEORI</b>	
2.1. Tinjauan Studi .....	5
2.2. Tinjauan Pustaka .....	7
2.3. Perangkat Lunak Pendukung.....	18
2.4. Kerangka Pikir.....	19
<b>BAB III METODOLOGI</b>	
3.1. Jenis, Metode, Subjek, Waktu dan Lokasi Penelitian .....	20
3.2. Pengumpulan Data .....	20
3.3. Tahapan Clustering.....	21
<b>BAB IV HASIL PENELITIAN</b>	
4.1. Hasil Pengumpulan Data .....	23
4.2. Tahapan Clustering K-Means.....	24
4.3. Hasil Implentasi K-Means dengan Python.....	28

## **BAB V PEMBAHASAN PENELITIAN**

5.1. Kinerja Model .....	32
5.2. Pembahasan Model .....	33

## **BAB VI KESIMPULAN DAN SARAN**

6.1. Kesimpulan.....	41
6.2. Saran .....	41

## **DAFTAR PUSTAKA**

## ***LAMPIRAN***

## DAFTAR TABEL

Tabel	Halaman
Tabel 1.1. Data Kuota Penerima dan Pemohon .....	2
Tabel 2.1. Penelitian Terkait .....	5
Tabel 2.2. Sample Dataset Nilai Mata Kuliah Kalkulus .....	13
Tabel 2.3. Dataset Nilai Mata Kuliah Matematika Diskrit .....	14
Tabel 2.4. Sample Hasil Penggabungan Dataset.....	14
Tabel 2.5. Sample Hasil Konversi Nilai Huruf Ke Nilai Angka.....	15
Tabel 2.6. <i>Centroids</i> Dan Jumlah Data .....	17
Tabel 2.7. Sampel Hasil <i>Clustering</i> .....	17
Tabel 2.8. Perangkat Lunak Pendukung .....	18
Tabel 3.1 Variabel Data .....	21
Tabel 3.2. Konversi Nilai Atribut .....	22
Tabel 4.1. Hasil Pengumpulan Data.....	23
Tabel 4.2. Dataset Hasil Olahan.....	25
Tabel 4.3. Konversi Nilai Atribut Penghasilan Ortu.....	26
Tabel 4.4. Konversi Nilai Atribut Fasilitas Rumah.....	26
Tabel 4.5. Dataset Akhir .....	27
Tabel 5.1. Pusat Cluster dan Jumlah Data.....	37
Tabel 5.2. Hasil Pelabelan Cluster .....	39

## DAFTAR GAMBAR

<b>Gambar</b>	<b>Halaman</b>
Gambar 2.1. Proses Knowledge Discoveryin Database (KDD) .....	9
Gambar 2.2. Proses Algoritma <i>K-Means</i> .....	12
Gambar 2.3. Contoh Visualisasi Penggunaan Metode <i>Elbow</i> .....	13
Gambar 2.4. Visualisasi Hasil Metode <i>Elbow</i> .....	16
Gambar 2.5. Visualisasi Hasil <i>Cluster</i> .....	16
Gambar 2.6. Informasi Pusat <i>Cluster</i> .....	17
Gambar 2.7. Bagan Kerangka Pikir .....	19
Gambar 5.1. Visualisasi Teknik Elbow .....	33
Gambar 5.2. Visualisasi Hasil Clustering .....	35

# **BAB I**

## **PENDAHULUAN**

### ***1.1 Latar Belakang***

Perbaikan dalam suatu Daerah yang seiring dengan penataan Pendidikan yang berkelanjutan dapat dikategorikan sebagai tolak ukur kemajuan dari suatu daerah. Meratanya pendidikan tentunya terlihat dengan adanya support dari pemerintah dalam memantau, meninjau dan mengevaluasi jalannya proses dari pelaksanaan pendidikan sebagai upaya untuk meningkatkan mutu pendidikan. Pemerataan pendidikan ini sudahlah barang tentu merupakan cakupan dari program desentralisasi, dimana pemerintah daerah memiliki tanggungjawab untuk menetapkan dan menentukan kebijakan yang disesuaikan dengan kebutuhan daerah tersebut demi mencapai suatu tujuan.[1]

Kartu Indonesia Pintar (KIP) Kuliah yang diluncurkan oleh pemerintah dibawah naungan Kementerian Pendidikan dan Kebudayaan sebagaimana diatur dalam “Peraturan Menteri Pendidikan dan Kebudayaan RI Nomor 9 Tahun 2018 tentang Perubahan atas Peraturan Menteri Pendidikan dan Kebudayaan Nomor 19 Tahun 2016 tentang Petunjuk Teknis Program Indonesia Pintar”. KIP kuliah ini bertujuan untuk membantu menyediakan bantuan Kuliah terkhusus untuk siswa miskin sehingga dapat melanjutkan kuliahnya dengan sasaran untuk mencegah anak putus sekolah. Awal masuknya bantuan Pendidikan ini, dikenal dengan nama BIDIKMISI hingga tahun 2020 berganti nama menjadi Kartu Indonesia Pintar (KIP) Kuliah.[1] Bantuan KIP kuliah ini diberikan kepada baik pada Perguruan Tinggi Negeri maupun Perguruan Tinggi Swasta, untuk Perguruan Tinggi Swasta penbagian kuota KIP Kuliah ditentukan dari Lembaga Layanan Pendidikan Tinggi (LLDIKTI ) di setiap Wilayah, khususnya daerah Gorontalo berada di wilayah LLDIKTI XVI.

Universitas Ichsan Gorontalo adalah salah satu Perguruan Tinggi swasta di wilayah LLDIKTI XVI yang diberikan kuota KIP Kuliah. Universitas Ichsan Gorontalo yang berdiri sejak tahun 2001 dengan membawahi 6 Fakultas dan 11

program Studi pada tahun akademik ini 2021/2022 tercatat Mahasiswa aktif kurang lebih 3700 mahasiswa. Dalam dua tahun terakhir ini, Universitas Ichsan Gorontalo menerima Mahasiswa Baru dengan rerata penerimaan sebanyak 600 hingga 700-an Mahasiswa Baru yang tentunya dengan keanekaragaman tingkat perekonomian orang tua mereka. Sesuai dengan adanya bantuan Kartu Indonesia Pintar (KIP) Kuliah diberikan ke Mahasiswa Baru pada semester satu, maka mereka diberikan hak untuk mendapatkan bantuan tersebut dengan ketentuan dan beberapa syarat yang harus diperhatikan utamanya dari golongan ekonomi menengah kebawah atau masyarakat miskin. [2]

Terbatasnya jumlah penerimaan (Kuota) mahasiswa KIP Kuliah yang akan diterima tentunya membutuhkan perhatian khusus dalam menetapkan mahasiswa yang tepat sasaran untuk mendapatkan bantuan dikarenakan keberagaman status perekonomian orang tua mahasiswa yang hampir sama. Berikut perbandingan data kuota dan pemohon KIP Kuliah :

**Tabel 1.1.** Data Kuota Penerima dan Pemohon

<b>Tahun</b>	<b>Jumlah Pemohon</b>	<b>Jumlah Kuota</b>
2020	97	56
2021	134	87

Sumber: Baak Unisan Gorontalo, 2022

Dari tabel diatas terlihat bahwa jumlah kuota yang tidak sepadan dengan jumlah pemohon, sehingganya perlu dilakukan pengelompokan atau *cluster* terhadap tingkat ekonomi mahasiswa demi mendapatkan kelompok mahasiswa yang diprioritaskan untuk mendapatkan bantuan KIP Kuliah. Hal ini juga tentunya bertujuan agar mempermudah pihak Baak Unisan dalam pemilihan mahasiswa yang lebih berhak untuk mendapatkan bantuan dikarenakan telah adanya pengelompokan terkait tingkat ekonomi mahasiswa.

*Cluster* disini disarankan karena belum diketahui masing-masing tingkat perekonomian dari masing-masing mahasiswa sehingganya dari persyaratan yang dimasukkan, dengan sendirinya akan terbentuk *cluster* dari masing-masing tingkat ekonomi mahasiswa, dan dari *cluster* tersebut akan diketahui kelompok tingkat ekonomi mahasiswa yang mana menjadi prioritas untuk mendapatkan bantuan KIP

Kuliah. Tentunya hal ini berdasar pada persyaratan yang diajukan diantaranya (1). Pekerjaan Orang Tua, (2). Penghasilan Orang Tua, (3). Jumlah tanggungan, (4). Fasilitas Rumah (Kepemilikan rumah, Sumber listrik, Sumber air dan MCK).

*K-Means* merupakan sebuah tatanan algoritma dari teknik data mining yang mampu melakukan cluster atau pengelompokan terhadap data heterogen karena pada dasarnya algoritma pengelompokan hanya mampu mengenali nilai atribut homogen saja. Proses algoritma *K-Means* berbeda dengan dengan algoritma data mining lainnya seperti Algoritma Apriori yang mencari *frequent item sets* yang sering muncul dengan model data heterogen juga kemudian dipangkas dan dilakukan perhitungan sesuai jumlah K item setnya[3]

Dalam beberapa penelitian sebelumnya yang dilakukan oleh Muliono dengan judul “Data Mining *Clustering* menggunakan Algoritma *K-Means* Untuk Klasterisasi Tingkat Tridarma Pengajaran Dosen” berhasil mendapatkan klasterisasi dengan nilai yang mendekati karakteristik menjadi lebih efektif. Hal ini membuktikan jika metode *K-Means* dapat digunakan dalam proses pengclusteran mahasiswa calon penerima KIP Kuliah.[3]

Berdasarkan uraian latar belakang diatas maka peneliti tertarik untuk mengangkat judul “***Clustering Tingkat Ekonomi Mahasiswa Calon Penerima Bantuan Kartu Indonesia Pintar (KIP) Kuliah Dengan Metode K-Means***” Diharapkan penelitian ini dapat bermanfaat bagi pihak Baak Universitas Ichsan Gorontalo untuk mengetahui *cluster* tingkat ekonomi mahasiswa calon penerima KIP Kuliah.

## **1.2 Identifikasi Masalah**

Berdasarkan latar belakang diatas dapat diidentifikasi permasalahannya adalah sulitnya menentukan kelompok prioritas calon penerima bantuan KIP Kuliah karena pemohon memiliki tingkat perekonomian hampir sama antara pemohon yang satu dengan yang lainnya.



### **1.3 Rumusan Masalah**

Adapun rumusan masalah yang dapat kita lihat dari identifikasi diatas yaitu :

1. Bagaimana menentukan jumlah *cluster* yang optimal untuk menentukan jumlah kelompok tingkat ekonomi mahasiswa calon penerima KIP Kuliah sebagai kelompok prioritas penerima bantuan.
2. Bagaimana hasil analisis setiap *clustering* tingkat ekonomi mahasiswa sehingga bisa dijadikan keputusan penentuan kelompok prioritas penerima bantuan KIP Kuliah

### **1.4 Tujuan Penelitian**

Adapun tujuan penelitian ini adalah :

1. Mengetahui jumlah cluster yang optimal dalam penentuan tingkat ekonomi mahasiswa calon penerima KIP Kuliah sebagai kelompok prioritas penerima bantuan
2. Mengetahui hasil analisis setiap *clustering* tingkat ekonomi mahasiswa sehingga bisa dijadikan keputusan penentuan kelompok prioritas penerima bantuan KIP Kuliah.

### **1.5 Manfaat Penelitian**

1. Manfaat Teoritis

Memberikan masukan bagi perkembangan ilmu pengetahuan dan teknologi, khususnya pada bidang ilmu computer, yaitu berupa pemuktahiran Metode *K-Means* dalam pengolahan data.

2. Manfaat Praktis

Sumbangan pemikiran, karya, bahan pertimbangan agar dapat menghasilkan sistem yang berkualitas.

## BAB II

### LANDASAN TEORI

#### 2.1 Tinjauan Studi

Berikut adalah penelitian yang terkait dengan metode *K-Means*, yaitu:

**Tabel 2.1.** Penelitian Terkait

No	Peneliti	Judul	Tahun	Metode	Hasil
1	Rezqiwati dan Amiruddin [1]	Clustering Tingkat Pemahaman Dasar Mahasiswa Pada Pra-Perkuliahan Probabilitas Statistika Dengan Metode K-Means	2022	<i>K-Means</i>	Probabilitas Statistik merupakan mata kuliah logika perhitungan dimana sebagian Mahasiswa merasa kesulitan dalam penerimaan materi sehingga bagi Dosen pengampu mata kuliah. Tujuan penelitian ini adalah untuk melakukan pengelompokan tingkat pemahaman dasar mahasiswa berdasarkan nilai mata kuliah prasyarat Metode yang digunakan yaitu <i>K-Means</i> untuk Clustering. Pada penelitian tersebut

No	Peneliti	Judul	Tahun	Metode	Hasil
					didapatkan hasil sebanyak 3 cluster, cluster 1 merupakan kelompok Mahasiswa yang tingkat pemahamannya digolongkan kategori Baik, Cluster 2 merupakan kelompok kategori Sangat Baik dan Cluster 3 merupakan kategori Kurang. Dari data tersebut masing-masing diperoleh untuk cluster 1 sebanyak 30 Mahasiswa, cluster 2 sebanyak 58 Mahasiswa dan cluster 3 sebanyak 22 Mahasiswa.
2.	Muliono [2]	Data Mining Clustering Menggunakan Algoritma K-Means Untuk Klasterisasi Tingkat	2019	<i>K-Means</i>	Universitas medan area memiliki dosen dengan jumlah yang banyak dimana setiap dosen mengampuh matakuliah sesuai bidang keahliannya masing-masing. Hasil

No	Peneliti	Judul	Tahun	Metode	Hasil
		Tridarma Pengajaran Dosen			klasterisasi tersebut selanjutnya akan merujuk pada pemberian besaran tunjangan yang akan diberikan ke masing2 dosen yang membuat dan mengumpulkan dokumen-dokumen pengajaran tersebut. Penelitian ini diharapkan dapat membantu proses klasterisasi dengan nilai yang mendekati karakteristik menjadi lebih efektif.

No	Peneliti	Judul	Tahun	Metode	Hasil
3.	Mega Pradnya [3]	<i>Clustering</i> Menggunakan Metode <i>K-Means</i> Untuk Menentukan Status Gizi Balita	2015	<i>K-Means</i>	Kekurangan gizi atau yang biasa disebut malnutrisi merupakan salah satu masalah kesehatan yang cukup sering menimpa balita-balita di Indonesia. Pada Penelitian ini mencoba untuk melakukan pengelompokan 60 balita di desa Karang Songo kedalam 5 cluster status gizi, dengan hasil pengelompokan status gizi balita yaitu cluster 1 - gizi buruk; cluster 2 - gizi kurang; cluster 3 - gizi baik; cluster 4 - gizi lebih; cluster 5 – obesitas.

## 2.2 Tinjauan Pustaka

### 2.2.1 Kartu Indonesia Pintar (KIP) Kuliah

Kementerian Pendidikan dan Kebudayaan (Kemendikbud) mencanangkan program Merdeka Belajar Episode IX Kartu Indonesia Pintar (KIP) Kuliah Merdeka. Beasiswa yang diberikan melalui KIP Kuliah bertujuan untuk meningkatkan akses gerak masyarakat yang tidak mampu dengan kata lain status

ekonomi menengah kebawah pada pendidikan tinggi yang lebih merata dan berkualitas, sehingga visi misi dari Bapak Presiden Joko Widodo terkait SDM unggul Indonesia dapat segera terlaksana[4].

KIP Kuliah Merdeka sebagai wujud komitmen serta keputusan dari Kemendikbud dalam memberikan akses pemerataan pendidikan tinggi yang berkualitas dan berkelanjutan. Pendidikan tinggi berpotensi memberikan dampak positif tercepat dalam pembangunan SDM unggul sesuai visi Presiden Joko Widodo. Kemendikbud mengubah skema KIP Kuliah dengan memberikan bantuan biaya pendidikan (uang kuliah) dan biaya hidup yang jauh lebih tinggi. Perubahan ini berlaku untuk mahasiswa baru yang menerima KIP Kuliah pada tahun 2021. Anggaran yang dialokasikan untuk KIP Kuliah meningkat signifikan dari Rp1,3 triliun pada 2020, menjadi sebesar Rp 2,5 triliun. KIP Kuliah akan diberikan kepada 200 ribu mahasiswa baru pada perguruan tinggi negeri (PTN) dan perguruan tinggi swasta (PTS) di bawah naungan Kemendikbud. Adapun biaya pendidikan akan disesuaikan dengan prodi masing-masing[5].

Kebijakan Program Indonesia Pintar (PIP) melalui Kartu Indonesia Pintar (KIP) dicanangkan oleh pemerintah Indonesia dibawah naungan Kementerian Pendidikan dan Kebudayaan (Kemendikbud) melalui Tim Nasional Percepatan Penanggulangan Kemiskinan (TNP2K). Tujuan dari program tersebut adalah untuk membantu mahasiswa dengan status ekenomi menengah kebawah untuk memperoleh pendidikan yang layak, mencegah anak putus sekolah, serta untuk memenuhi kebutuhan sekolah mereka. Bantuan ini diharapkan untuk dimanfaatkan oleh Mahasiswa dalam memenuhi kebutuhan sekolah seperti biaya transportasi siswa pulang pergi ke sekolah, biaya perlengkapan sekolah, dan uang saku siswa.[6]

### **2.2.2 Data Mining**

Data mining merupakan rangkaian proses yang menggali nilai tambah dari suatu kumpulan data dalam bentuk pengetahuan. Menurut Han dan Kamber, “Data mining merupakan proses menggali pengetahuan dari sejumlah data besar [7].

Tujuan dari data mining adalah untuk mengelompokkan pola yang harus

ditemukan dalam data mining. Secara umum data mining dapat dibedakan menjadi dua kategori yaitu deskriptif dan prediktif. Adapun operasi – operasi dan teknik-teknik terkait:

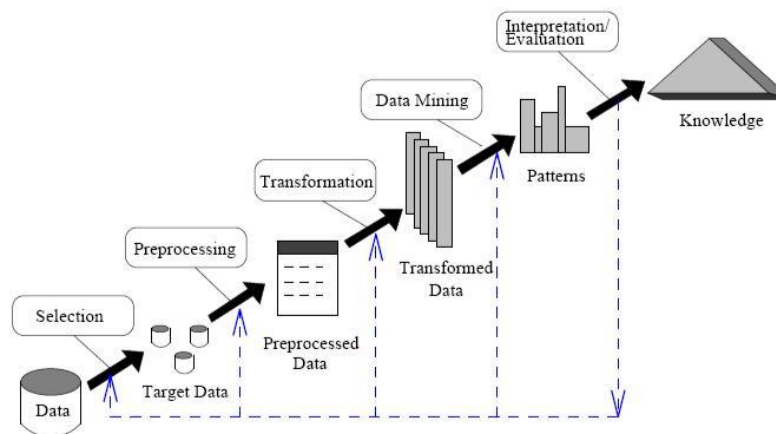
1. *Operasi Predictive modeling : (classification, value prediction)*
2. *Database segmentation : (demographic clustering, neural clustering)*
3. *Link Analysis : (association discovery, sequential pattern discovery, similar times equence discovery)*
4. *Deviation detection: (statistics, visualization)*

Hasil data mining biasanya diintegrasikan dengan sistem pendukung keputusan (DSS). Misalnya, dalam aplikasi bisnis, informasi yang dihasilkan oleh data mining dapat diintegrasikan dengan alat manajemen aktivitas produk untuk mengimplementasikan dan menguji promosi pemasaran yang efektif. Integrasi ini membutuhkan langkah pasca-pemrosesan untuk memastikan hanya hasil yang efektif dan bermanfaat yang dapat digabungkan dengan DSS. Salah satu pekerjaan dan *post-processing* adalah visualisasi, yang memungkinkan *analist* untuk mengeksplorasi data dan hasil data mining dari berbagai sudut[5].

#### **2.2.4 Tahapan Data Mining**

Tahapan yang dilakukan dalam proses data mining dimulai dengan pemilihan data, mulai dari data sumber hingga data target, tahap pemrosesan untuk meningkatkan kualitas data, pengalihan data, data mining serta tahap interpretasi dan pengetesan atau evaluasi. Diharapkan dengan dihasilkannya output berupa pengetahuan baru dapat memberikan kontribusi yang lebih baik. Dengan rincian sebagai berikut [8].





**Gambar 2. 1** Proses Knowledge Discoveryin Database (KDD)

1. **Data selection**  
 sebelum tahap penggalian informasi dalam *Knowledge Discovery in Database* dimulai, diperlukan pemilihan (seleksi) data dari sekumpulan data operasional. Data yang dipilih digunakan untuk proses data mining disimpan dalam satu. File, terpisah dari database operasional.
2. ***Pre-processing / cleaning***  
 Sebelum proses data mining, perlu dilakukan pembersihan data yang menjadi fokus KDD. Proses pembersihan termasuk menghapus data duplikat, memeriksa data yang tidak konsisten, dan memperbaiki kesalahan pada data..
3. **Transformation**  
*coding* adalah proses mengubah data yang dipilih, jadi data ini cocok untuk proses data mining. Proses *coding* di KDD 11 adalah proses kreatif dan sebenarnya bergantung pada jenis atau mode informasi yang akan dicari di database.
4. **Data mining**  
 Data mining adalah proses menggunakan teknik atau metode tertentu untuk menemukan pola atau informasi yang menarik dalam data yang dipilih. Teknik, metode, atau algoritma dalam data mining sangat bervariasi. Pilihan metode atau algoritma yang benar sangat bergantung pada tujuan dan proses keseluruhan KDD.

## 5. *Interpretation / evaluation*

Pola informasi yang dihasilkan dari proses data mining perlu ditampilkan dalam bentuk yang mudah dipahami oleh peserta yang tertarik. Tahap ini adalah bagian dari proses KDD (disebut interpretasi). Tahapan ini meliputi pengecekan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau asumsi sebelumnya.

### 2.2.5 Clustering

*Clustering* adalah merupakan suatu tahapan untuk mengelompokkan himpunan data yang atribut kelasnya belum diketahui, secara konsep *clustering* adalah untuk mengoptimalkan dan meminimalkan kesamaan antar kelas. sebagai contoh, ada suatu himpunan obyek, proses pertama dapat diklaster menjadi beberapa himpunan kelas selanjutnya menjadi sebuah himpunan beraturan sehingga dapat diturunkan berdasarkan kelompok cluster tertentu. *Cluster* juga dapat diartikan sebagai kelompok. Sehingga analisa clustering pada dasarnya akan menghasilkan sejumlah *cluster* (kelompok). Sebelum melakukan analisa perlu diterapkan pemahaman bahwa suatu himpunan dari data tertentu sebenarnya sudah memiliki kesamaan di antara anggota lainnya. Olehnya itu, setiap anggota yang memiliki kesamaan karakteristik dikelompokkan ke dalam satu atau lebih dari suatu kelompok. Tujuan dari pada *clustering* data adalah untuk meminimalisasikan fungsi tujuan yang ditetapkan didalam proses *clustering*, umumnya selalu meminimalisasi variasi suatu *cluster* dan memaksimalisasikan variasi antar *cluster* [3].

### 2.2.6 Algoritma K-Means

Algoritma K-Means atau *Clustering K-Means* adalah “merupakan salah satu teknik *clustering* pada Data Mining proses pemodelan tanpa supervisi dan metode pengelompokkan data secara berkelompok”. Data yang dikelompokkan metode *K-Means* menjadi beberapa kelompok dan setiap kelompok memiliki karakteristik yang memiliki kesamaan dengan yang lainnya tetapi dengan kelompok lainnya memiliki karakteristik yang berbeda. Dengan tujuan menimalisasi perbedaan setiap data didalam satu cluster serta memaksimalkan perbedaan dengan cluster yang lain.

Istilah-istilah didalam algoritma *k-means clustering*[9] :

1. Cluster adalah kelompok atau grup.
2. Cendroid adalah titik pusat untuk menentukan *auclidian distance*.
3. Iterasi adalah pengulangan proses, berhenti ketika hasil iterasi telah konvergen.

Secara umum tahapan algoritma *K-Means* adalah sebagai berikut:

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 *Euclidean distance*:

$$dik = \sqrt{\sum (C_{ij} - C_{kj})^2} \quad n \ i=1 \quad (2.1)$$

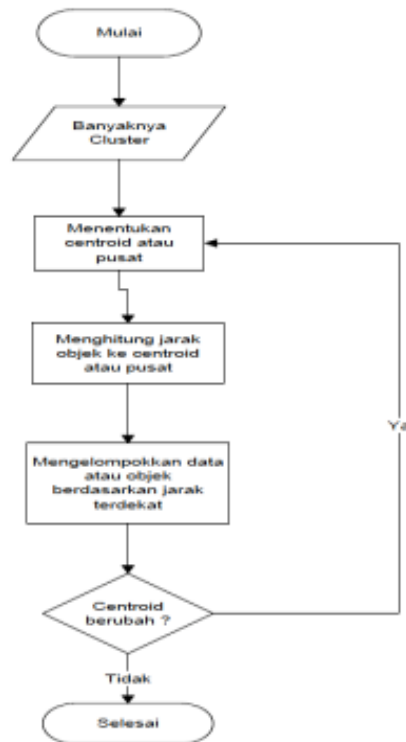
Dimana:

$C_{ij}$  = pusat cluster

$C_{kj}$  = data

5. Hitung Kembali semua titik centroid

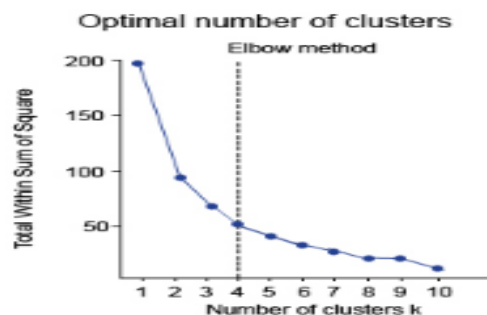
6. Ulang kembali Langkah 3-5 hingga nilai titik centroid tidak lagi berubah.



**Gambar 2.2.** Proses Algoritma *K-Means*

### 2.2.7 Metode Elbow

Metode *Elbow* merupakan salah satu metode atau teknik yang bisa diterapkan pada algoritma *K-Means* untuk menghitung dan menentukan jumlah  $K$  yang optimal. Penentuan jumlah  $K$  yang optimal pada metode Elbow hanya dengan melihat titik pada graph dimana penurunan inersia tidak lagi signifikan. Hal ini bisa dilihat pada gambar di bawah ini dimana jumlah  $K$  yang optimal adalah  $K = 4$ .



**Gambar 2.3.** Contoh Visualisasi Penggunaan Metode *Elbow*

### 2.2.8 Penerapan Algoritma *K-Means*

Berikut contoh penerapan algoritma *K-Means* dengan judul “Clustering Tingkat Pemahaman Dasar Mahasiswa Pada Pra-Perkuliahan Probabilitas Statistika dengan Metode *K-Means*”. Tools yang digunakan dalam proses *clustering* ini menggunakan bahasa pemrograman *Python* dan beberapa *library* seperti *Numpy*, *Pandas*, *Matplotlib* dan *Scikit Learn*[10].

#### 1. Tahap *Selection*

Pada tahapan ini melakukan penyeleksian dataset yang akan digunakan yaitu nilai mahasiswa pada mata kuliah prasyarat yaitu Kalkulus pada tabel 2.2 dan mata kuliah Matematika Diskrit pada tabel 2.3.

**Tabel 2.2.** Sample Dataset Nilai Mata Kuliah Kalkulus

No	NIM	Nilai Huruf
1	T31 19 101	D
2	T31 20 002	A
3	T31 20 003	B
4	T31 20 004	A
5	T31 20 005	A
6	T31 20 006	B
7	T31 20 007	A
8	T31 20 008	A
9	T31 20 009	A
10	T31 20 010	A

**Tabel 2.3.** Dataset Nilai Mata Kuliah Matematika Diskrit

No	NIM	Nilai Huruf
1	T31 14 097	B
2	T31 16 126	D
3	T31 16 342	D
4	T31 17 006	B

No	NIM	Nilai Huruf
5	T31 17 057	B
6	T31 17 069	D
7	T31 17 149	B
8	T31 17 168	C
9	T31 18 003	D
10	T31 20 010	A

## 2. Tahap *Preprocessing*

Pada tahapan ini melakukan proses penghapusan data yang *double*, melengkapi data yang masih kosong atau null dan melakukan penggabungan data nilai antara nilai mata kuliah Kalkulus dan nilai mata kuliah Matematika Diskrit, sehingga hasil penggabungan dataset tersebut seperti pada tabel 2.4 berikut ini :

**Tabel 2.4.** Sample Hasil Penggabungan Dataset

No	NIM	Nilai Kalkulus	Nilai M_Diskrit
1	T31 20 002	A	A
2	T31 20 003	B	C
3	T31 20 004	A	D
4	T31 20 005	A	A
5	T31 20 006	B	D
6	T31 20 007	A	A
7	T31 20 008	A	B
8	T31 20 009	A	B
9	T31 20 010	A	A
10	T31 20 011	B	A

## 3. Tahap *Transformation*

Pada tahapan ini melakukan konversi data nilai yaitu nilai mata kuliah bentuk huruf dikonversikan ke bentuk nilai angka yaitu nilai A = 4, B = 3, C = 2, D = 1,

hal ini perlu dilakukan supaya dapat digunakan dalam pemodelan metode *K-Means*.

**Tabel 2.5.** Sample Hasil Konversi Nilai Huruf Ke Nilai Angka

No	NIM	Nilai Kalkulus	Nilai M_Diskrit
1	T31 20 002	4	4
2	T31 20 003	3	2
3	T31 20 004	4	1
4	T31 20 005	4	4
5	T31 20 006	3	1
6	T31 20 007	4	4
7	T31 20 008	4	3
8	T31 20 009	4	3
9	T31 20 010	4	4
10	T31 20 131	3	4

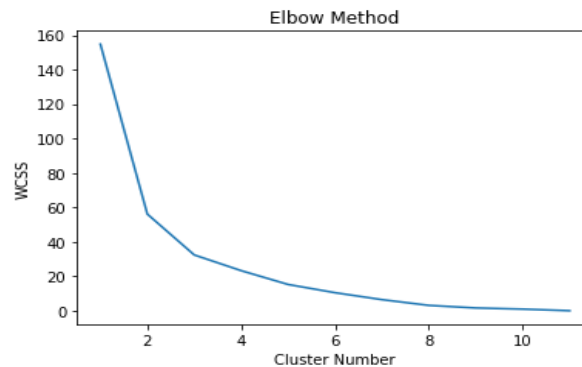
4. Tahap *Data Mining* :

Pada tahapan ini melakukan pengembangan model data mining yaitu menggunakan model *K-Means* untuk melakukan *clustering*. Perintah yang digunakan untuk proses pemodelan metode *K-Means* pada penelitian ini menggunakan *library* dari *Scikit Learn*.

5. Tahap *Interpretation/Evaluation* :

Pada tahapan ini melakukan evaluasi terhadap jumlah *cluster* yang optimum dengan menggunakan metode atau teknik *elbow*, hasil dari metode tersebut dibuatkan dalam bentuk grafik atau visualisasi dengan menggunakan *library Matplotlib*, sehingga hasil visualisasi dapat dilihat pada gambar 4 berikut ini :

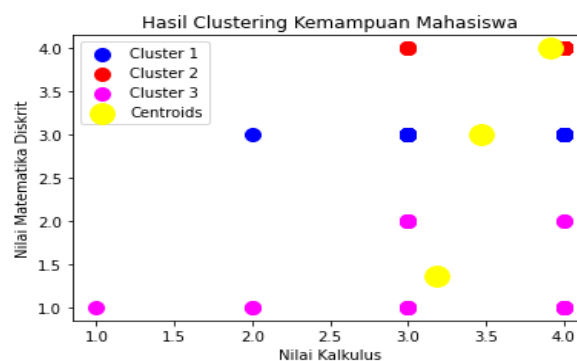




**Gambar 2.4.** Visualisasi Hasil Metode *Elbow*

Pada gambar 2.2 di atas dapat dilihat bahwa jumlah *cluster* yang optimum pada penelitian ini adalah sebanyak 3 *cluster*. Melalui metode *Elbow* ini, jumlah *cluster* yang optimum bisa langsung diketahui hasilnya tanpa harus melakukan pengujian jumlah *cluster* secara berulang-ulang. Jumlah *cluster* optimum yang sudah didapatkan, selanjutnya jumlah *cluster* tersebut dimasukkan ke dalam fungsi *K-Means* yang pada *library* yang digunakan yaitu *Scikit Learn*.

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



**Gambar 2.5.** Visualisasi Hasil *Cluster*

Berdasarkan pada gambar 2.3 di atas, dapat ditentukan titik *Centroids* untuk masing-masing *cluster* dengan melihat posisi koordinat pusat *cluster* yang berwarna kuning atau dapat juga diketahui titik *Centroids* secara langsung dengan menggunakan perintah “(*kmeans.cluster\_centers\_*)” sehingga didapatkan hasil titik *Centroids* sebagai berikut :

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

**Gambar 2.6.** Informasi Pusat *Cluster*

Banyaknya anggota untuk masing-masing *cluster* yang sudah dihasilkan dapat diketahui dengan menggunakan perintah “`dfData['Kluster'].value_counts()`”, sehingga hasilnya bisa dilihat seperti pada tabel 2.6 berikut ini :

**Tabel 2.6.** *Centroids* Dan Jumlah Data

<i>Cluster</i>	<i>Centroids</i>		Jumlah Mahasiswa
	Kalkulus	Matematika Diskrit	
1	3,46	3	30
2	3,91	4	58
3	3,18	1,36	22

Pada tabel 2.6 di atas, berdasarkan titik *Centroids* pada setiap *cluster*, maka dapat dianalisis bahwa untuk *cluster* 1 merupakan kategori tingkat pemahaman Mahasiswa kategori baik, untuk *cluster* 2 merupakan kategori sangat baik dan untuk *cluster* 3 merupakan kategori kurang, sehingga bisa disimpulkan bahwa *cluster* 3 merupakan kelompok Mahasiswa yang perlu diberikan materi tambahan sebelum mereka mengikuti perkuliahan Mata Kuliah Probabilitas Statistik yaitu sebanyak 22 Mahasiswa. Hasil pelabelan *clustering* untuk masing-masing Mahasiswa dapat dilihat pada tabel 2.7 di bawah ini :

**Tabel 2.7.** Sampel Hasil *Clustering*

NIM	Kalkulus	M_Diskrit	Kluster
T31 20 002	4	4	2
T31 20 003	3	2	3
T31 20 004	4	1	3
T31 20 005	4	4	2
T31 20 006	3	1	3

NIM	Kalkulus	M_Diskrit	Kluster
T31 20 007	4	4	2
T31 20 008	4	3	1
T31 20 009	4	3	1
T31 20 010	4	4	2
T31 20 011	3	4	1

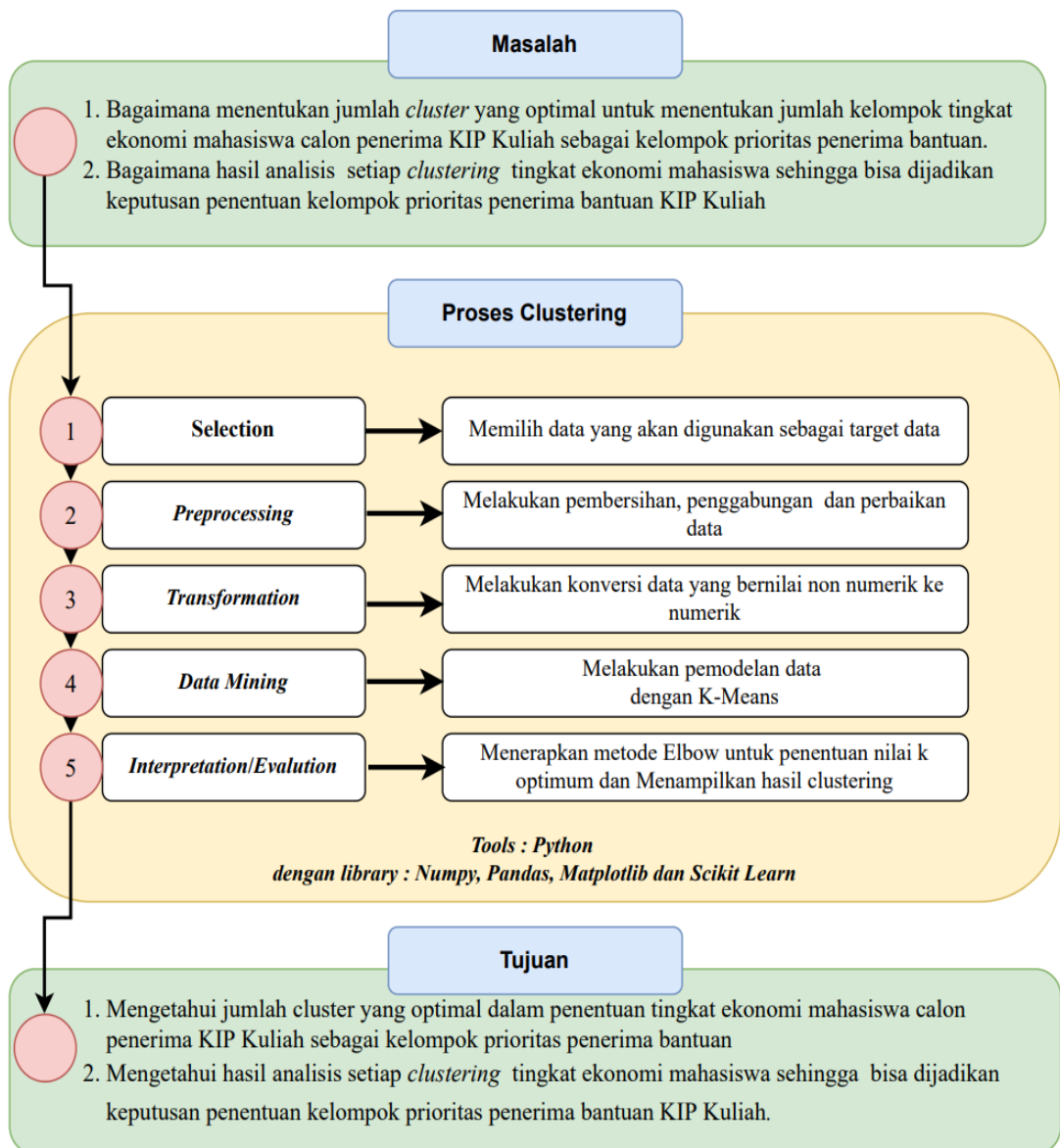
### 2.3 Perangkat Lunak Pendukung

Adapun perangkat lunak pendukung yang digunakan oleh penulis dalam melakukan clustering pada skripsi ini diantaranya adalah:

**Tabel 2.8.** Perangkat Lunak Pendukung

No.	Perangkat Lunak Pendukung	Keterangan
1.	<i>Python</i>	Bahasa pemrograman tingkat tinggi, Penulisan kode/sintaks lebih sederhana, Bersifat <i>open-source</i> , <i>cross-platform</i> dan cocok digunakan untuk Data Scientist serta didukung banyak library seperti : <i>Numpy</i> , <i>SciPy</i> , <i>Pandas</i> , <i>Scikit-Learn</i> , dan <i>Matplotlib</i>
2.	<i>Google Colaboratory</i> (Google Colab)	<i>Integrated Development Environments</i> (IDE) berbasis Web yang akan digunakan untuk menuliskan coding Bahasa <i>Python</i>

## 2.4 Kerangka Pikir



**Gambar 2.7.** Bagan Kerangka Pikir

## **BAB III**

### **METODE PENELITIAN**

#### **3.1 *Jenis, Metode, Subjek, Waktu dan Lokasi Penelitian***

Berdasarkan dari tingkat penerapan maka, Penelitian ini merupakan penelitian terapan. Di lihat dari jenis informasi yang diolah, penelitian ini adalah penelitian kuantitatif. Pada saat yang sama, dari perilaku ke data, studi ini adalah studi konfirmatori.

Penelitian ini menggunakan metode penelitian studi kasus pada Universitas Ichsan Gorontalo. Dengan demikian jenis penelitian ini adalah deskriptif. Subjek penelitian ini adalah Clustering Tingkat Ekonomi Mahasiswa Calon Penerima KIP Kuliah Menggunakan Metode *K-Means*. Penelitian ini dimulai dari Januari hingga April 2022 berlokasi di Universitas Ichsan Gorontalo.

#### **3.2 *Pengumpulan Data***

Untuk pengumpulan data, digunakan dua jenis data yaitu data primer dan sekunder. Data primer merupakan data yang diperoleh atau didapatkan dengan terjun langsung ke lapangan. Sedangkan data sekunder merupakan data yang diperoleh atau didapatkan dari penelitian yang sudah ada atau kepustakaan.

##### **1. Penelitian Data Primer (Lapangan)**

Untuk mendapatkan data primer yang merupakan data lapangan atau langsung keobjek penelitian yaitu bertempat di Unisan Gorontalo. Teknik yang digunakan:

- a. Observasi yaitu memungkinkan untuk analis, meninjau atau mengamati secara langsung. Adapun pada penelitian ini dilakukan dengan cara mengumpulkan data Mahasiswa calon penerima KIP.
- b. Wawancara, dengan cara mengajukan beberapa pertanyaan kepada bagian yang terkait pada BAAK Unisan Gorontalo dalam proses prediksi Calon Penerima KIP.

Adapun variabel dengan tipe datanya masing-masing ditunjukkan pada Tabel 3.1 berikut ini:

**Tabel 3. 1** Variabel Data

No	Variabel/Atribut	Type	Nilai	Ket
1	Pekerjaan Orang Tua	Nominal	- Tidak Bekerja - Nelayan - Petani - TNI/Polri - Wirausaha - Pegawai Swasta	Variabel Input
2	Penghasilan Orang Tua	Rasio	0 s/d 4.000.000	Variabel Input
3	Jumlah Tanggungan	Rasio	1 s/d 50	Variabel Input
4	Fasilitas Rumah (Kepemilikan Rumah, Sumber Listrik, Sumber Air, MCK, Luas Bangunan)	Nominal	- Memadai - Cukup Memadai - Kurang Memadai - Tidak Memadai	Variabel Input

## 2. Penelitian Data Sekunder (Kepustakaan)

Data sekunder di dapatkan dari pengkajian kepustakaan yang didasarkan pada dasar-dasar teori untuk digunakan sebagai untuk melengkapi data primer. Metode ini analisis sistem digunakan untuk memperoleh contoh-contoh dokumen atau catatan yang berkaitan dengan materi penelitian. Analisis sistem mencari data atau hal-hal yang berhubungan dengan penelitian baik dari sumber majalah, buku dan lain sebagainya.

### 3.3 Tahapan Clustering

Pada tahap ini dilakukan serangkaian proses sesuai dengan kerangka pikir sebagai berikut :

1. Tahap *Selection* : pada tahapan ini akan dilakukan seleksi data terhadap dataset yang dikumpulkan dari data pemohon KIP Kuliah pada Universitas Ichsan Gorontalo dua tahun terakhir untuk dijadikan sebagai target data (dataset)

2. Tahap *Preprocessing* : pada tahapan ini dilakukan beberapa proses yaitu penghapusan data yang duplikat, pengecekan data yang tidak konsisten, melengkapi data yang kosong dan menghapus atribut yang tidak diperlukan dalam pemodelan.
3. Tahap *Transformation* : pada tahapan ini dilakukan perubahan atau konversi data yaitu atribut yang bernilai data nominal akan dikonversi menjadi nilai rasio agar dapat digunakan dalam pemodelan *K-Means*. Contoh untuk atribut pekerjaan ayah.

**Tabel 3.2** Konversi Nilai Atribut

Nilai Atribut	Nilai Konversi (Inialisasi)
Tidak Bekerja	1
Nelayan	2
Petani	3
TNI/Polri	4
Wirausaha	5
Pegawai Swasta	6

Semua atribut yang bertipe nominal akan dilakukan konversi nilai seperti pada table di atas.

4. Tahap *Data Mining* : pada tahapan ini dilakukan pengembangan model yaitu penggunaan metode *K-Means* untuk mendapatkan hasil *cluster*.
5. Tahap *Interpretation/Evaluation* : pada tahapan ini menentukan berapa jumlah *cluster* yang optimal dengan metode *Elbow*. Hasil yang didapat dari nilai *K* optimum pada metode *Elbow* tersebut diterapkan ke dalam metode *K-Means* kembali. Hasilnya divisualisasikan jumlah kluster beserta anggota klusternya, sehingga berdasarkan visualisasi tersebut dengan mudah dapat dipahami arti dari setiap kelompok *cluster* tersebut yang selanjutnya dijadikan sebagai pengetahuan atau keputusan untuk *clustering* tingkat ekonomi mahasiswa calon penerima KIP Kuliah.



## BAB IV

### HASIL PENELITIAN

#### 4.1 Hasil Pengumpulan Data

**Tabel 4.1. Hasil Pengumpulan Data**

No	Nama Siswa	Pekerjaan Ortu	Penghasilan Ortu	Jumlah Tanggungan
1	Fajri Lasena	Petani	< Rp. 250.000	2 Orang
2	Cristian Dandi Bryvan Rondonuwu	Lainnya	Rp. 500.000 - Rp. 750.000	3 Orang
3	Ricardo Suli	TIDAK BEKERJA	Tidak Berpenghasilan	1 Orang
4	Adam Halid	Petani	< Rp. 250.000	3 Orang
5	FATRA LAIYA	Petani	< Rp. 250.000	2 Orang
6	Sartina Abd R Ongku	Petani	< Rp. 250.000	3 Orang
7	ROYAN SAPII	Petani	< Rp. 250.000	2 Orang
8	Sri Nelviawati Noiyo	Petani	< Rp. 250.000	3 Orang
9	WAHYUNINGSIH YUNUS	Petani	< Rp. 250.000	3 Orang
10	RAFLIANSYAH LINGUDE	Petani	< Rp. 250.000	3 Orang
11	RIRIN SAPUTRI RAJAWALI	Petani	< Rp. 250.000	5 Orang
12	Refawanti Ibrahim	Lainnya	Rp. 500.000 - Rp. 750.000	5 Orang
13	ALFIRAH AHMAD	Petani	< Rp. 250.000	2 Orang
14	CHIAN ARDIKA PUTRA YANTU	Wirausaha	Rp. 2.000.000 - Rp. 3.000.000	2 Orang
15	ANSJAR FAJRIE SENEWE	Lainnya	Rp. 500.000 - Rp. 750.000	4 Orang
16	RENALDI ADAM	Petani	< Rp. 250.000	2 Orang
17	FEBRIYANTI MANSUR	Nelayan	Rp. 250.000 - Rp. 500.000	3 Orang
18	Anggi Mustapa	Petani	< Rp. 250.000	1 Orang
19	Regita Cahyani Ismail	Wirausaha	Rp. 2.000.000 - Rp. 3.000.000	3 Orang
20	Nuryati Abdullah	PNS	Rp. 2.000.000 - Rp. 3.000.000	2 Orang
21	...	...	...	...
192	SRISKA ASKIA AHIRI	Nelayan	Rp. 500.000 - Rp. 750.000	4 Orang

No	Kepemilikan Rumah	Sumber Listrik	Sumber Air	MCK	Luas Bangunan
1	Sendiri	PLN	Sungai/Mata Air	Berbagi Pakai	< 25 M2
2	Sendiri	PLN	Sungai/Mata Air	Berbagi Pakai	25-50 M2
3	Sendiri	PLN	Sungai/Mata Air	Berbagi Pakai	< 25 M2
4	Sendiri	PLN	Sumur	Kepemilikan Sendiri Didalam	< 25 M2
5	Sendiri	PLN	Sumur	Kepemilikan Sendiri Didalam	< 25 M2
6	Sendiri	PLN	Sumur	Kepemilikan Sendiri Didalam	< 25 M2
7	Sendiri	PLN	Sumur	Kepemilikan Sendiri Didalam	< 25 M2
8	Sendiri	PLN	Sumur	Kepemilikan Sendiri Didalam	< 25 M2
9	Sendiri	PLN	Sumur	Kepemilikan Sendiri Didalam	< 25 M2
10	Sendiri	PLN	Sumur	Kepemilikan Sendiri Didalam	< 25 M2
11	Menumpang	PLN	Sumur	Kepemilikan Sendiri Didalam	< 25 M2
12	Menumpang	PLN	PDAM	Kepemilikan Sendiri Didalam	25-50 M2
13	Menumpang	PLN	PDAM	Kepemilikan Sendiri Didalam	< 25 M2
14	Sendiri	PLN	Sumur	Kepemilikan Sendiri Didalam	50-99 M2
15	Sendiri	PLN	Sumur	Kepemilikan Sendiri Didalam	25-50 M2
16	Sendiri	PLN	Sumur	Kepemilikan Sendiri Didalam	< 25 M2
17	Sendiri	PLN	Sumur	Kepemilikan Sendiri Didalam	25-50 M2
18	Sendiri	PLN	Sumur	Kepemilikan Sendiri Didalam	< 25 M2
19	Sendiri	PLN	PDAM	Kepemilikan Sendiri Didalam	50-99 M2
20	Menumpang	PLN	PDAM	Kepemilikan Sendiri Didalam	50-99 M2
21	...	...	...	...	...
192	Sendiri	PLN	PDAM	Berbagi Pakai	100 - 200 M2

## 4.2 Tahapan Clustering K-Means

### 4.2.1 Tahapan Selection

Pada tahapan ini dilakukan seleksi data dari dataset yang sudah dikumpulkan untuk tahun 2020 – 2021 dengan jumlah data set sebanyak 192. Dari atribut yang ada hanya digunakan atribut penghasilan ortu dan jumlah tanggungan, untuk atribut kepemilikan rumah, sumber listrik, sumber air, mck dan luas bangunan diwakili

oleh atribut fasilitas rumah, sedangkan atribut pekerjaan orangtua tidak digunakan karena sudah terwakili oleh penghasilan orangtua (pekerjaan kategori dibawah penghasilannya juga dibawah).

#### 4.2.2 Tahap Preprocessing

Pada tahap ini dilakukan penggabungan atribut kepemilikan rumah, sumber listrik, sumber air, mck dan luas bangunan menjadi fasilitas rumah dengan mengisi nilai atribut yaitu : Tidak memadai, kurang memadai, cukup memadai dan memadai, penentuan nilai atribut tersebut berdasarkan nilai atribut gabungan diatas sehingga didapatkan dataset sebagai berikut :

**Tabel 4.2** Dataset Hasil Olahan

No	Nama Siswa	Pekerjaan Ortu	Penghasilan Ortu	Jumlah Tanggungan	Fasilitas Rumah
1	Fajri Lasena	Petani	< Rp. 250.000	2	Tidak Memadai
2	Cristian Dandi Bryvan Rondonuwu	Lainnya	Rp. 500.000 - Rp. 750.000	3	Tidak Memadai
3	Ricardo Suli	TIDAK BEKERJA	Tidak Berpenghasilan	2	Tidak Memadai
4	Adam Halid	Petani	< Rp. 250.000	3	Tidak Memadai
5	FATRA LAIYA	Petani	< Rp. 250.000	2	Tidak Memadai
6	Sartina Abd R Ongku	Petani	< Rp. 250.000	3	Tidak Memadai
7	ROYAN SAPII	Petani	< Rp. 250.000	2	Tidak Memadai
8	Sri Nelviawati Noiyo	Petani	< Rp. 250.000	3	Tidak Memadai
9	WAHYUNINGSIH YUNUS	Petani	< Rp. 250.000	3	Tidak Memadai
10	RAFLIANSYAH LINGUDE	Petani	< Rp. 250.000	3	Tidak Memadai
11	RIRIN SAPUTRI RAJAWALI	Petani	< Rp. 250.000	3	Tidak Memadai
12	Refawanti Ibrahim	Lainnya	Rp. 500.000 - Rp. 750.000	5	Tidak Memadai
13	ALFIRAH AHMAD	Petani	< Rp. 250.000	2	Tidak Memadai
14	CHIAN ARDIKA PUTRA YANTU	Wirausaha	Rp. 2.000.000 - Rp. 3.000.000	2	Kurang Memadai

15	ANSHAR FAJRIE SENEWE	Lainnya	Rp. 500.000 - Rp. 750.000	4	Tidak Memadai
16	RENALDI ADAM	Petani	< Rp. 250.000	2	Tidak Memadai
17	FEBRIYANTI MANSUR	Nelayan	Rp. 250.000 - Rp. 500.000	3	Tidak Memadai
18	Anggi Mustapa	Petani	< Rp. 250.000	2	Tidak Memadai
19	Regita Cahyani Ismail	Wirausaha	Rp. 2.000.000 - Rp. 3.000.000	3	Kurang Memadai
20	Nuryati Abdullah	PNS	Rp. 2.000.000 - Rp. 3.000.000	2	Kurang Memadai
...	...	...	...	...	...
192	SRISKA ASKIA AHIRI	Nelayan	Rp. 500.000 - Rp. 750.000	4	Cukup Memadai

#### 4.2.3 Tahap Transformasi

Pada tahap ini dilakukan tahap konversi untuk atribut nilai non numerik ke numerik yaitu atribut penghasilan orangtua dan fasilitas rumah dengan hasil sebagai berikut :

**Tabel 4.3** Konversi Nilai Atribut Penghasilan Ortu

No	Penghasilan Ortu	Nilai
1	Tidak berpenghasilan	1
2	< Rp. 250.000	2
3	Rp. 250.000 – Rp. 500.000	3
4	Rp. 500.000 – Rp. 750.000	4
5	Rp. 2.000.000 – Rp. 3.000.000	5
6	> Rp. 3.000.000	6

**Tabel 4.4** Konversi Nilai Atribut Fasilitas Rumah

No	Fasilitas Rumah	Nilai
1	Tidak Memadai	1
2	Kurang Memadai	2
3	Cukup Memadai	3
4	Memadai	4

Setelah dilakukan perbaikan data set maka didapatkan dataset akhir sebagai berikut :

**Tabel 4.5** Dataset Akhir

No	Nama Siswa	Pekerjaan Ortu	Penghasilan Ortu	Jumlah Tanggungan	Fasilitas Rumah
1	Fajri Lasena	Petani	2	2	1
2	Cristian Dandi Bryvan Rondonuwu	Lainnya	4	3	1
3	Ricardo Suli	TIDAK BEKERJA	1	2	1
4	Adam Halid	Petani	2	3	1
5	FATRA LAIYA	Petani	2	2	1
6	Sartina Abd R Ongku	Petani	2	3	1
7	ROYAN SAPII	Petani	2	2	1
8	Sri Nelviawati Noiyo	Petani	2	3	1
9	WAHYUNINGSIH YUNUS	Petani	2	3	1
10	RAFLIANSYAH LINGUDE	Petani	2	3	1
11	RIRIN SAPUTRI RAJAWALI	Petani	2	3	1
12	Refawanti Ibrahim	Lainnya	4	5	1
13	ALFIRAH AHMAD	Petani	2	2	1
14	CHIAN ARDIKA PUTRA YANTU	Wirausaha	5	2	2
15	ANSJAR FAJRIE SENEWE	Lainnya	4	4	1
16	RENALDI ADAM	Petani	2	2	1
17	FEBRIYANTI MANSUR	Nelayan	3	3	1
18	Anggi Mustapa	Petani	2	2	1
19	Regita Cahyani Ismail	Wirausaha	5	3	2
20	Nuryati Abdullah	PNS	5	2	2
...	...	...	...	...	...
192	SRISKA ASKIA AHIRI	Nelayan	4	4	3

Setelah dilakukan koversi nilai seperti dataset diatas maka tahap selanjutnya akan dilakukan pemodelan clustering dengan menggunakan metode k-means.

### 4.3 Hasil Implentasi K-Means Dengan Python

Implementasi pemodelan k-means menggunakan Bahasa pemograman python dengan memakai Tools Notebook Colaboratory dengan tahapan sebagai berikut :

#### 1. Import Library Python

Berikut Library Python yang perlu di import agar proses pemodelan k-means dapat dijalankan :

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn.cluster import KMeans
import seaborn as sns
from mpl_toolkits.mplot3d import Axes3D
import plotly.graph_objs as go
from plotly import tools
from plotly.subplots import make_subplots
import plotly.offline as py
```

#### 2. Pengaktifan Google Drive

Data set yang akan digunakan disimpan di google drive untuk itu diaktifkan dengan perintah sebagai berikut :

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

#### 3. Membaca Dataset

Dataset yang sudah dibersihkan yang akan digunakan pada proses klustering yang berbentuk file excel harus di upload terlebih dahulu kedalam google drive, selanjutnya dataset dibaca dengan menggunakan perintah sebagai berikut :

```
#Baca Dataset
df = pd.read_excel('/content/drive/MyDrive/SKRIPSI/Dataset KIP.xlsx')
df.head()
```

Setelah dijalankan perintah diatas maka didapatkan hasil sebagai berikut :

No	Nama Mahasiswa	Pekerjaan Ortu	Penghasilan Ortu	Jumlah Tanggungan	Fasilitas Rumah	
0	1	Fajri Lasena	Petani	< Rp. 250.000	2	Tidak Memadai
1	2	Cristian Dandi Bryvan Rondonuwu	Lainnya	Rp. 500.000 - Rp. 750.000	3	Tidak Memadai
2	3	Ricardo Suli	TIDAK BEKERJA	Tidak Berpenghasilan	2	Tidak Memadai
3	4	Adam Halid	Petani	< Rp. 250.000	3	Tidak Memadai
4	5	FATRA LAIYA	Petani	< Rp. 250.000	2	Tidak Memadai

#### 4. Konversi Dataset Excel Ke CSV

Dataset dalam format excel perlu dikonversi kebentuk CSV (COMMA Separated Values) agar bisa digunakan dalam pemodelan data dengan menggunakan perintah sebagai berikut :

```
#Konversi dataset Excel ke CSV
df.to_csv("/content/drive/MyDrive/SKRIPSI/DatasetKIP.
csv", index=False)
df.info()
```

Setelah dijalankan perintah diatas maka didapatkan hasil sebagai berikut :

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 192 entries, 0 to 191
Data columns (total 6 columns):
#   Column                Non-Null Count  Dtype
---  -
0   No                     192 non-null   int64
1   Nama Mahasiswa        192 non-null   object
2   Pekerjaan Ortu        192 non-null   object
3   Penghasilan Ortu      192 non-null   object
4   Jumlah Tanggungan     192 non-null   int64
5   Fasilitas Rumah       192 non-null   object
dtypes: int64(2), object(4)
memory usage: 9.1+ KB
```

#### 5. Menghitung Banyak Data

Sebelum melakukan konversi nilai atribut yang bernilai non numerik perlu di hitung banyaknya data untuk setiap nilai atribut. Berikut perintah untuk menghitung banyaknya data pada atribut penghasilan ortu :

```
df['Pekerjaan Ortu'].value_counts()
```

Setelah dijalankan perintah diatas maka didapatkan hasil sebagai berikut :

```
< Rp. 250.000          66
Rp. 500.000 - Rp. 750.000  59
Rp. 2.000.000 - Rp. 3.000.000  24
Tidak Berpenghasilan      16
> Rp. 3.000.000          16
Rp. 250.000 - Rp. 500.000    11
Name: Penghasilan Ortu, dtype: int64
```

Selanjutnya dihitung banyaknya data untuk atribut fasilitas rumah :

```
df['Fasilitas Rumah'].value_counts()
```

Setelah dijalankan perintah diatas maka didapatkan hasil sebagai berikut :

```
Tidak Memadai      129
Cukup Memadai      34
Memadai             16
Kurang Memadai      13
Name: Fasilitas Rumah, dtype: int64
```

### Konversi Nilai Atribut

Semua atribut yang bertipe string atau non numerik harus dilakukan konversi nilai agar bisa digunakan dalam pemodelan k-means.

Berikut perintah untuk konversi nilai atribut penghasilan ortu dan fasilitas rumah :

```
# Konversi Data String ke Angka
df['Penghasilan Ortu'].replace(['Tidak Berpenghasilan', '< Rp. 250.000',
'Rp. 250.000 - Rp. 500.000', 'Rp. 500.000 - Rp. 750.000',
'Rp. 2.000.000 - Rp. 3.000.000', '> Rp. 3.000.000'],
[1,2,3,4,5,6], inplace=True)
df['Fasilitas Rumah'].replace(['Tidak Memadai', 'Kurang Memadai', 'Cukup Memadai', 'Memadai'],
[1,2,3,4], inplace=True)
```



Berikut hasil konversi nilai atribut :

	No	Nama Mahasiswa	Pekerjaan Ortu	Penghasilan Ortu	Jumlah Tanggungan	Fasilitas Rumah
0	1	Fajri Lasena	Petani	2	2	1
1	2	Cristian Dandi Bryvan Rondonuwu	Lainnya	4	3	1
2	3	Ricardo Suli	TIDAK BEKERJA	1	2	1
3	4	Adam Halid	Petani	2	3	1
4	5	FATRA LAIYA	Petani	2	2	1
...	...	...	...	...	...	...
187	188	REFLAN PAKAYA	Wirausaha	5	2	3
188	189	VELIA WUISANG	Lainnya	4	3	3
189	190	MUHAMADUN BASHAR LABA	PNS	6	2	4
190	191	SILVIA SEPTIMIYANI KOBİ	PNS	6	2	4
191	192	SRISKAASKIAAHIRI	Nelayan	4	4	3

192 rows x 6 columns

### Penentuan Kolom Dataset

Sebelum dilakukan pemodelan k-means atribut yang akan digunakan perlu di tentukan terlebih dahulu yaitu atribut : Penghasilan ortu, jumlah tanggungan dan fasilitas rumah dengan perintah sebagai berikut :

```
# Memilih/menentukan kolom sebagai dataset
x = df[['Penghasilan Ortu', 'Jumlah Tanggungan', 'Fasilitas Rumah']].values
x
```

Setelah dijalankan perintah diatas mendapatkan hasil sebagai berikut :

```
array([[2, 2, 1],
       [4, 3, 1],
       [1, 2, 1],
       [2, 3, 1],
       [2, 2, 1],
       [2, 3, 1],
       [2, 2, 1],
       [2, 3, 1],
       [2, 3, 1],
       [2, 3, 1],
       [2, 3, 1],
       [2, 3, 1],
       [4, 5, 1],
       [2, 2, 1],
```

Selanjutnya dilakukan pemodelan metode K-Means dan akan dibahas pada bab selanjutnya.

## BAB V

### PEMBAHASAN PENELITIAN

#### 5.1 Kinerja Model

Kinerja metode K-Means dalam pengelompokkan data sangat ditentukan dari jumlah cluster yang ditentukan, agar tidak membutuhkan waktu yang lama dalam menguji coba berapa jumlah cluster yang tepat, maka diperlukan suatu teknik untuk menentukan jumlah cluster yang optimum yaitu dengan menggunakan teknik Elbow. Berikut potongan coding penggunaan teknik Elbow :

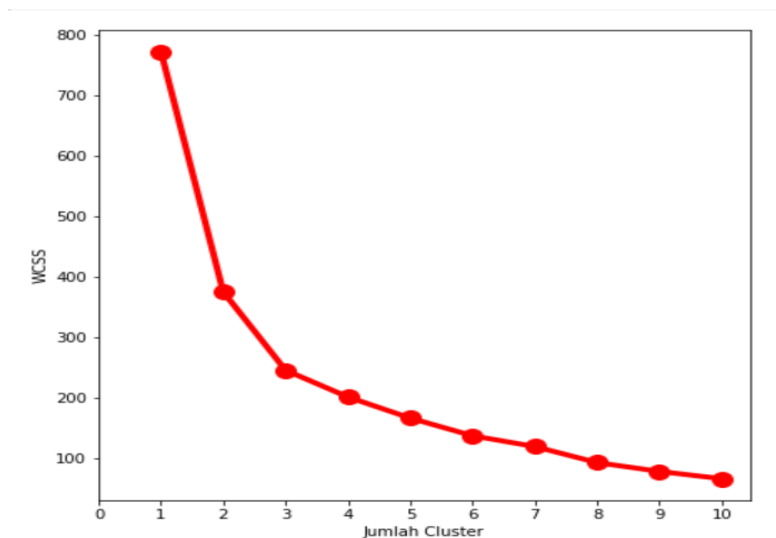
```
# Menentukan jumlah cluster yang optimum dengan teknik
Elbow

WCSS = []

for i in range(1,11):
    model = KMeans(n_clusters = i,init = 'k-means++')
    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()
```

Hasil dari potongan coding di atas adalah sebagai berikut :



Gambar 5.1 Visualisasi Teknik Elbow

Berdasarkan gambar 5.1 di atas dapat dilihat bahwa jumlah cluster yang optimum adalah 5 cluster. Sehingga untuk penerapan metode K-Menas dapat langsung digunakan jumlah cluster = 5.

## 5.2 Pembahasan Model

### 5.2.1 Pemodelan Metode K-Means

Setelah didapatkan jumlah cluster yang optimum, maka tahap selanjutnya adalah menerapkan dalam pemodelan K-Means dengan menggunakan potongan coding sebagai berikut :

```
# Pemodelan K-Means dengan jumlah cluster yang optimum
model = KMeans(n_clusters = 5, init = "k-means++",
               max_iter = 300, n_init = 10, random_state = 0)
y_clusters = model.fit_predict(x)
```

### 5.2.2 Visualisasi Hasil Pemodelan K-Means

Setelah dilakukan pemodelan K-Means, maka selanjutnya dapat dibuatkan visualisasi setiap cluster yang terbentuk dengan menggunakan potongan coding sebagai berikut :

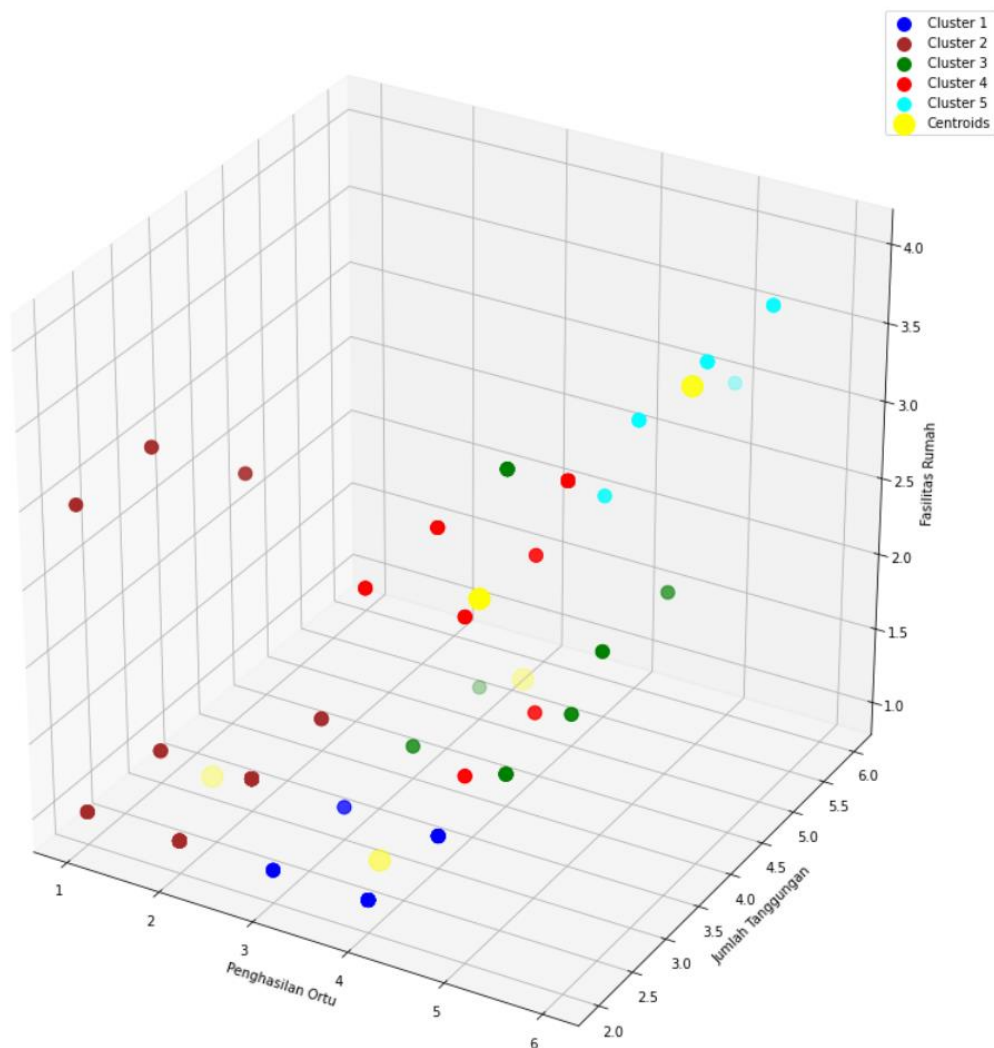
```

# Visualisasi hasil cluster dengan 3d scatterplot men
gunakan matplotlib

fig = plt.figure(figsize = (15,15))
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 = 100 , color = 'blue', label
= "Cluster 1")
ax.scatter(x[y_clusters == 1,0],x[y_clusters == 1,1],
x[y_clusters == 1,2], s = 100 , color = 'brown', labe
l = "Cluster 2")
ax.scatter(x[y_clusters == 2,0],x[y_clusters == 2,1],
x[y_clusters == 2,2], s = 100 , color = 'green', labe
l = "Cluster 3")
ax.scatter(x[y_clusters == 3,0],x[y_clusters == 3,1],
x[y_clusters == 3,2], s = 100 , color = 'red', label
= "Cluster 4")
ax.scatter(x[y_clusters == 4,0],x[y_clusters == 4,1],
x[y_clusters == 4,2], s = 100 , color = 'cyan', label
= "Cluster 5")
ax.scatter(model.cluster_centers_[ :,0],model.cluster_
centers_[ :,1],
            model.cluster_centers_[ :,2], s = 250, c =
"yellow", label = "Centroids")
ax.set_xlabel('Penghasilan Ortu')
ax.set_ylabel('Jumlah Tanggungan')
ax.set_zlabel('Fasilitas Rumah')
ax.legend()
plt.show()

```

Hasil potongan coding di atas dapat dilihat pada gambar berikut :



**Gambar 5.2 Visualisasi Hasil Clustering**

Berdasarkan gambar di atas, dapat dilihat bahwa setiap cluster diwakili titik berwarna yaitu cluster 1 warna biru, cluster 2 warna coklat, cluster 3 warna hijau, cluster 4 warna merah, cluster 5 warna cyan dan pusat cluster warna kuning. Pada gambar di atas juga dengan bantuan visualisasi 3 dimensi secara sejelas dapat dilihat keterkaitan ketiga atribut yang digunakan dalam clustering ....

### 5.2.3 Pusat Cluster dan Jumlah Data

Proses analisa untuk menarik kesimpulan terhadap setiap cluster, selain langsung melihat pada gambar di atas perlu diketahui nilai pusat cluster untuk masing-cluster beserta banyak data dalam setiap cluster. Berikut potongan coding untuk mengetahui pusat masing cluster.

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

Hasil output potongan coding di atas sebagai berikut :

```
[[3.75      2.5      1.       ]  
 [1.80487805 2.70731707 1.09756098]  
 [4.       4.25     1.53571429]  
 [4.84210526 2.42105263 2.92105263]  
 [5.75     4.16666667 3.75     ]]
```

Banyaknya data dalam setiap cluster dapat ditampilkan dengan potongan coding sebagai berikut :

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

Hasil output potongan coding di atas sebagai berikut :

```
1    82  
3    38  
0    32  
2    28  
4    12
```

Agar mudah dipahami data di atas, dibuatkan dalam bentuk tabel sebagai berikut :

**Tabel 5.1** Pusat Cluster dan Jumlah Data

Cluster	Pusat Cluster			Jumlah Data
	Penghasilan Ortu	Jumlah Tanggungan	Fasilitas Rumah	
1	3.75	2.5	1	32
2	1.80	2.70	1.09	82
3	4	4.25	1.53	28
4	4.84	2.42	2.92	38
5	5.75	4.16	3.75	12

Dari tabel di atas, dapat dianalisis atau disimpulkan bahwa :

**Cluster 1** adalah kelompok mahasiswa yang tingkat ekonominya sedang dan menjadi prioroitas kedua penerima bantuan (penghasilan orang tua Rp. 250.000 – Rp. 500.000, jumlah tanggungan 2 s/d 3 orang, fasilitas rumah tidak memadai)

**Cluster 2** adalah kelompok mahasiswa yang tingkat ekonominya kecil dan menjadi prioroitas pertama penerima bantuan (penghasilan orang tua kurang dari Rp. 250.000, jumlah tanggungan 2 s/d 3 orang, fasilitas rumah kurang memadai)

**Cluster 3** adalah kelompok mahasiswa yang tingkat ekonominya menengah ke bawah dan menjadi prioroitas ketiga penerima bantuan (penghasilan orang tua Rp. 500.000-Rp. 750.000, jumlah tanggungan 4 s/d 5 orang, fasilitas rumah kurang memadai)

**Cluster 4** adalah merupakan kelompok mahasiswa yang tingkat ekonominya menengah dan menjadi prioroitas keempat penerima bantuan (penghasilan orang tua Rp. 2.000.000-Rp.3.000.000, jumlah tanggungan 4 s/d 5 orang, fasilitas rumah cukup memadai)

**Cluster 5** adalah merupakan kelompok mahasiswa yang tingkat ekonominya menengah ke atas dan menjadi prioroitas kelima penerima bantuan (penghasilan orang tua di atas Rp.3.000.000, jumlah tanggungan 4 s/d 5 orang, fasilitas rumah memadai)

### 5.2.4 Pelabelan Data Hasil Cluster

Setiap data dapat diketahui masuk dalam cluster berapa dengan menggunakan potongan coding sebagai berikut :

```
# Menampilkan hasil kluster
print(model.labels_)

# Menambahkan kolom "kluster" dalam data frame Data N
ilai

df["Kluster"] = model.labels_
df
```

Hasil output potongan coding di atas adalah sebagai berikut :

```
[1 0 1 1 1 1 1 1 1 1 1 2 1 3 2 1 0 1 3 3 0 1 1 0 0 2 0 2 1 0 1 1 0 1 2 1 1
0 1 1 1 1 0 1 1 1 1 1 0 1 2 0 1 1 1 1 2 1 0 1 2 3 0 0 1 2 1 2 0 0 3 1 3
0 2 1 1 1 1 1 1 1 1 0 2 1 1 0 1 0 1 1 1 1 1 1 0 1 0 1 1 2 0 0 1 1 1 3 2 1
0 3 1 2 1 2 1 1 0 1 0 2 2 1 0 1 1 1 1 1 0 1 2 3 2 0 2 0 2 3 1 4 4 3 3 3 3
3 3 3 3 4 4 3 3 2 1 4 2 3 4 3 3 1 3 2 3 3 3 3 4 4 2 3 1 2 4 3 4 4 3 3 3 4
3 1 3 3 3 3 2]
```

	No	Nama Siswa	Pekerjaan Ortu	Penghasilan Ortu	Jumlah Tanggungan	Fasilitas Rumah	Kluster
0	1	Fajri Lasena	Petani	2	2	1	1
1	2	Cristian Dandi Bryvan Rondonuwu	Lainnya	4	3	1	0
2	3	Ricardo Suli	TIDAK BEKERJA	1	2	1	1
3	4	Adam Halid	Petani	2	3	1	1
4	5	FATRA LAIYA	Petani	2	2	1	1
...	...	...	...	...	...	...	...
187	46	NaN	Wirausaha	5	2	3	3
188	47	NaN	Lainnya	4	3	3	3
189	48	NaN	PNS	6	2	4	3
190	49	NaN	PNS	6	2	4	3
191	50	NaN	Nelayan	4	4	3	2

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

```
#Menambahkan kolom keterangan cluster

conditions = [
    (df['Kluster']==0),
    (df['Kluster']==1),
    (df['Kluster']==2),
    (df['Kluster']==3),
```



```
(df['Kluster']==4)]
choices = ['C1','C2','C3','C4','C5']
df['Ket'] = np.select(conditions, choices)
df
```

```
#Simpan Hasil Cluster ke Excel
df.to_excel("/content/drive/MyDrive/SKRIPSI/HasilKIP.
xlsx", index=False)
```

Hasil akhir setelah dilakukan kedua perintah di atas, maka hasilnya dapat dilihat pada tabel berikut :

**Tabel 5.2 Hasil Pelabelan Cluster**

No	Nama Mahasiswa	Pekerjaan Ortu	Penghasilan Ortu	Jumlah Tanggungan	Fasilitas Rumah	Kluster	Ket
1	Fajri Lasena	Petani	2	2	1	1	C2
2	Cristian Dandi Bryvan Rondonuwu	Lainnya	4	3	1	0	C1
3	Ricardo Suli	TIDAK BEKERJA	1	2	1	1	C2
4	Adam Halid	Petani	2	3	1	1	C2
5	FATRA LAIYA	Petani	2	2	1	1	C2
6	Sartina Abd R Ongku	Petani	2	3	1	1	C2
7	ROYAN SAPII	Petani	2	2	1	1	C2
8	Sri Nelviawati Noiyo	Petani	2	3	1	1	C2
9	WAHYUNINGSIH YUNUS	Petani	2	3	1	1	C2
10	RAFLIANSYAH LINGUDE	Petani	2	3	1	1	C2
11	RIRIN SAPUTRI RAJAWALI	Petani	2	3	1	1	C2
12	Refawanti Ibrahim	Lainnya	4	5	1	2	C3
13	ALFIRAH AHMAD	Petani	2	2	1	1	C2

No	Nama Mahasiswa	Pekerjaan Ortu	Penghasilan Ortu	Jumlah Tanggungan	Fasilitas Rumah	Kluster	Ket
14	CHIAN ARDIKA PUTRA YANTU	Wirausaha	5	2	2	3	C4
15	ANSHAR FAJRIE SENEWE	Lainnya	4	4	1	2	C3
16	RENALDI ADAM	Petani	2	2	1	1	C2
17	FEBRIYANTI MANSUR	Nelayan	3	3	1	0	C1
18	Anggi Mustapa	Petani	2	2	1	1	C2
19	Regita Cahyani Ismail	Wirausaha	5	3	2	3	C4
20	Nuryati Abdullah	PNS	5	2	2	3	C4
...	...	...	...	...	...	...	...
192	SRISKA ASKIA AHIRI	Nelayan	4	4	3	2	C3

## **BAB VI**

### **KESIMPULAN DAN SARAN**

#### **6.1. Kesimpulan**

Berdasarkan hasil penelitian yang sudah diuraikan di atas tentang clustering tingkat ekonomi mahasiswa dengan metode K-Means, maka dapat ditarik kesimpulan sebagai berikut :

1. Jumlah cluster yang optimum untuk melakukan clustering tingkat ekonomi mahasiswa calon penerima kartu indonesia pintar (KIP) yaitu sebanyak 5 cluster dengan menggunakan teknik Elbow.
2. Hasil clustering untuk masing-masing cluster yaitu cluster 1 merupakan kelompok mahasiswa yang tingkat ekonominya sedang dan menjadi prioritas kedua penerima bantuan, cluster 2 merupakan kelompok mahasiswa yang tingkat ekonominya kecil dan menjadi prioritas pertama penerima bantuan, cluster 3 merupakan kelompok mahasiswa yang tingkat ekonominya menengah ke bawah dan menjadi prioritas ketiga penerima bantuan, cluster 4 merupakan kelompok mahasiswa yang tingkat ekonominya menengah dan menjadi prioritas keempat penerima bantuan, dan cluster 5 merupakan kelompok mahasiswa yang tingkat ekonominya menengah ke atas dan menjadi prioritas kelima penerima bantuan.

#### **6.2. Saran**

Berdasarkan kesimpulan tersebut diatas, peneliti dapat memberikan saran untuk penelitian selanjutnya, yaitu :

1. Penelitian selanjutnya dapat menggunakan atribut lain atau menambah dataset lebih banyak lagi agar hasil clusteringnya bisa dibandingkan.
2. Dapat dikembangkan dengan menggunakan metode clustering yang lain agar hasil bisa dibandingkan mana yang lebih baik.

## DAFTAR PUSTAKA

- [1] D. R. Intansari and A. O. Penelitian, “Analisis Data Mining Memprediksi Jumlah Pernikahan di bawah Umur Menggunakan Metode Least Square ( Studi Kasus : KUA Tegalrejo ),” pp. 16–18, 2020.
- [2] H. Muhadzdzab, M. Asfi, and T. E. Putri, “Sistem Prediksi untuk Menentukan Jumlah Pendaftaran Mahasiswa Baru pada Universitas Catur Insan Cendekia Menggunakan Metode Least Square,” *J. Inform. Univ. Pamulang*, vol. 5, no. 3, p. 350, 2020, doi: 10.32493/informatika.v5i3.6598.
- [3] M. I. CHOLID, “Sistem Prediksi Industri Rumahan Menggunakan Metode Least Square,” 2018.
- [4] H. Priyatman, F. Sajid, and D. Haldivany, “Klasterisasi Menggunakan Algoritma K-Means Clustering untuk Memprediksi Waktu Kelulusan Mahasiswa,” *JEPIN (Jurnal Edukasi dan Penelit. Inform.)*, vol. 5, no. 1, pp. 62–66, Apr. 2019, doi: 10.26418/JP.V5I1.29611.
- [5] “PENERAPAN ALGORITMA K-MEANS PADA SISWA BARU SEKOLAH MENENGAH KEJURUAN UNTUK CLUSTERING JURUSAN | Nur | InfoTekJar : Jurnal Nasional Informatika dan Teknologi Jaringan.” <https://jurnal.uisu.ac.id/index.php/infotekjar/article/view/70/0> (accessed Jun. 17, 2022).
- [6] B. Mulyadi, Jaroji, and A. T, “Aplikasi Sistem Pemesanan Jasa Laundry (E-Laundry) Berbasis Android,” *Zo. J. Sist. Inf.*, vol. 1, no. 1, pp. 48–57, 2019.
- [7] Y. Mardi, “Data Mining : Klasifikasi Menggunakan Algoritma C4.5,” *J. Edik Inform.*, vol. 2, no. 2, pp. 213–219, 2017.
- [8] I. Adiwijaya, “Text Mining dan Knowledge Discovery,” *Kolokium bersama komunitas datamining Indonesia & soft-computing Indonesia*, 2006. .

- [9] G. Gustientiedina, M. H. Adiya, and Y. Desnelita, "Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan," *J. Nas. Teknol. dan Sist. Inf.*, vol. 5, no. 1, pp. 17–24, Apr. 2019, doi: 10.25077/TEKNOSI.V5I1.2019.17-24.
- [10] R. K. Dinata, S. Safwandi, N. Hasdyna, and N. Azizah, "Analisis K-Means Clustering pada Data Sepeda Motor," *INFORMAL Informatics J.*, vol. 5, no. 1, p. 10, 2020, doi: 10.19184/isj.v5i1.17071.

L A M P I R A N

## Kode Program

### 1. Konversi Data

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 192 entries, 0 to 191
Data columns (total 6 columns):
#   Column                Non-Null Count  Dtype
---  -
0   No                     192 non-null   int64
1   Nama Mahasiswa        192 non-null   object
2   Pekerjaan Ortu        192 non-null   object
3   Penghasilan Ortu      192 non-null   object
4   Jumlah Tanggungan    192 non-null   int64
5   Fasilitas Rumah       192 non-null   object
dtypes: int64(2), object(4)
memory usage: 9.1+ KB
```

### 2. Menentukan Jumlah Cluster Dengan Teknik Elbow

```
WCSS = []
for i in range(1,11):
    model = KMeans(n_clusters = i,init = 'k-means++')
    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()
```

### 3. Pemodelan K-Means Dengan Jumlah Cluster Yang Optimum

```
model = KMeans(n_clusters = 5, init = "k-means++",
               max_iter = 300, n_init = 10, random_state = 0)
y_clusters = model.fit_predict(x)
```

### 4. Visualisasi Hasil Cluster

```
fig = plt.figure(figsize = (15,15))
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 = 100 , color = 'blue', label = "
Cluster 1")
ax.scatter(x[y_clusters == 1,0],x[y_clusters == 1,1],x[
y_clusters == 1,2], s = 100 , color = 'brown', label =
"Cluster 2")
ax.scatter(x[y_clusters == 2,0],x[y_clusters == 2,1],x[
y_clusters == 2,2], s = 100 , color = 'green', label =
"Cluster 3")
ax.scatter(x[y_clusters == 3,0],x[y_clusters == 3,1],x[
y_clusters == 3,2], s = 100 , color = 'red', label = "C
luster 4")
ax.scatter(x[y_clusters == 4,0],x[y_clusters == 4,1],x[
y_clusters == 4,2], s = 100 , color = 'cyan', label = "
Cluster 5")
ax.scatter(model.cluster_centers_[:,0],model.cluster_ce
nters_[:,1],
            model.cluster_centers_[:,2], s = 250, c = "y
ellow", label = "Centroids")
ax.set_xlabel('Penghasilan Ortu')
ax.set_ylabel('Jumlah Tanggungan')
ax.set_zlabel('Fasilitas Rumah')
ax.legend()
plt.show()

```



## Surat Keterangan Telah Melakukan Penelitian



**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 : 088/UNISAN-G/VI/2022**

Yang bertanda tangan dibawah ini :

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

Dengan ini menerangkan bahwa :

Nama : Maimun S. Sompia  
Nim : T3115070  
Prodi : Teknik Informatika  
Fakultas : Ilmu Komputer  
Universitas : Ichsan Gorontalo

Adalah benar telah melakukan penelitian dalam rangka penulisan Skripsinya yang berjudul **“Clustering Tingkat Ekonomi Mahasiswa Calon Penerima KIP Kuliah Dengan Metode K-Means”** sejak tanggal 02 Maret – 20 April 2022, dan telah pula membahas materi hasil penelitiannya dengan kami.

Demikian surat keterangan ini dibuat untuk dapat dipergunakan sebagaimana mestinya,

Gorontalo, 13 Juni 2022

An. Rektor

Wakil Rektor Bidang Akademik, Tata Kelola & Sistem Informasi



Amiruddin, M.Kom  
NIDN. 0910097601

Tembusan:

1. Rektor Universitas Ichsan Gorontalo
2. Ketua Lemlit Universitas Ichsan Gorontalo
3. Mahasiswa yang bersangkutan
4. Arsip

## Hasil Turnitin



Similarity Report ID: oid:25211:18603439

PAPER NAME

SKRIPSI\_T3115070\_MAIMUN S. SOMPA.  
docx

AUTHOR

T3115070 - MAIMUN S. SOMPA mhayy2  
9@gmail.com

WORD COUNT

**7512 Words**

CHARACTER COUNT

**44443 Characters**

PAGE COUNT

**49 Pages**

FILE SIZE

**980.3KB**

SUBMISSION DATE

**Jun 11, 2022 11:05 AM GMT+8**

REPORT DATE

**Jun 11, 2022 11:06 AM GMT+8**

### ● 20% Overall Similarity

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

- 20% Internet database
- 2% Publications database
- Crossref database
- Crossref Posted Content database
- 2% Submitted Works database

### ● Excluded from Similarity Report

- Bibliographic material
- Small Matches (Less than 25 words)

### ● 20% Overall Similarity

Top sources found in the following databases:

- 20% Internet database
- 2% Publications database
- Crossref database
- Crossref Posted Content database
- 2% Submitted Works database

#### TOP SOURCES

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

1	<b>ejurnal.ung.ac.id</b> Internet	5%
2	<b>jurnal.unimed.ac.id</b> Internet	3%
3	<b>andi.ddns.net</b> Internet	2%
4	<b>es.scribd.com</b> Internet	2%
5	<b>putuekaputra2021.blogspot.com</b> Internet	1%
6	<b>scribd.com</b> Internet	1%
7	<b>kaggle.com</b> Internet	1%
8	<b>pakdosen.pengajar.co.id</b> Internet	<1%

9	jurnal.wicida.ac.id	<1%
	Internet	
10	jurnalsaintek.uinsby.ac.id	<1%
	Internet	
11	repository.iainbengkulu.ac.id	<1%
	Internet	
12	jurnal.unigal.ac.id	<1%
	Internet	
13	eprints.umpo.ac.id	<1%
	Internet	
14	humannesia.com	<1%
	Internet	
15	digilib.unhas.ac.id	<1%
	Internet	

## Surat Bebas Pustaka



**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 : 035/Perpustakaan-Fikom/VI/2022

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

Nama Anggota : Maimun S. Somp  
No. Induk : T3115070  
No. Anggota : M202276

Terhitung mulai hari, tanggal : Sabtu, 11 Juni 2022, dinyatakan telah bebas pinjam buku dan koleksi perpustakaan lainnya.

Demikian keterangan ini di buat untuk di gunakan sebagaimana mestinya.

**Gorontalo, 11 Juni 2022**

**Mengetahui,  
Kepala Perpustakaan**



**Apriyanto Alhamad, M.Kom**  
**NIDN : 0924048601**

### Daftar Riwayat Hidup

Nama : Maimun S. Somp  
NIM : T3115070  
Jenis Kelamin : Perempuan  
Tempat Tanggal Lahir: Gorontalo, 29 April 1997  
Status : Menikah  
Alamat : Jl.Beringin, Duingi Kota Gorontalo  
Kewarganegaraan : Indonesia  
Agama : Islam  
Email : [mhayy29@gmail.com](mailto:mhayy29@gmail.com)



#### Riwayat Pendidikan :

1. Tamatan SD Negeri 18 Kota Gorontalo pada tahun 2009
2. Tamatan SMP Negeri 4 Kota Gorontalo pada tahun 2012
3. Tamatan SMA Negeri 2 Kota Gorontalo pada Tahun 2015
4. Masuk dan diterima di Universitas Ichsan Gorontalo pada tahun 2015





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

**FAKULTAS ILMU KOMPUTER**

SK MENDIKNAS NOMOR 84/D/O/2001

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

**Berita Acara Perbaikan/Revisi Ujian SKRIPSI**

Pada hari ini, Selasa 14-Juni-2022, Pukul 10.00-12.00 Wita. Telah dilaksanakan Ujian SKRIPSI mahasiswa Program Studi Teknik Informatika Fakultas Ilmu Komputer Universitas Ichsan Gorontalo.

Nama : Maimun S. Somp  
Nim : T3115070  
Pembimbing I : Rezqiwati Ishak, M.Kom  
Pembimbing II : Yulianty Lasena, M.Kom  
Judul SKRIPSI : Clustering Tingkat Ekonomi Mahasiswa Calon Penerima Kartu Indonesia Pintar (KIP) Kuliah Dengan Metode K-Means

Oleh Komite Seminar sebagai berikut :

No	Komite Seminar	Status	Tanda Tangan
1	Amiruddin, M.Kom	Ketua	
2	Sudirman S. Panna, M.Kom	Anggota	
3	Roys Pakaya, M.Kom	Anggota	
4	Rezqiwati Ishak, M.Kom	Anggota	
5	Yulianty Lasena, M.Kom	Anggota	