

***CLUSTERING* PENYAKIT KEHAMILAN  
MENGUNAKAN METODE *K-MEANS* PADA  
PUSKESMAS BOLIYOHUTO**

Oleh  
**NAZLY IBRAHIM**  
**T3122223**

**SKRIPSI**

Untuk memenuhi salah satu syarat ujian  
guna memperoleh gelar Sarjana



**PROGRAM SARJANA  
TEKNIK INFORMATIKA  
UNIVERSITAS ICHSAN GORONTALO  
GORONTALO  
2024**

**CLUSTERING PENYAKIT KEHAMILAN  
MENGUNAKAN METODE K-MEANS PADA  
PUSKESMAS BOLIYOHUTO**

Oleh  
NAZLY IBRAHIM  
T3122223

**SKRIPSI**

Untuk memenuhi salah satu syarat ujian  
Guna memperoleh gelar Sarjana  
Program Studi Teknik Informatika,  
Ini telah disetujui oleh Tim Pembimbing

Gorontalo, Juni 2024

Pembimbing I

Pembimbing II



Rezqiwaty Ishak, M.Kom

NIDN : 0903087901



Yulivanti Lasena, M.Kom

NIDN : 0907078603

## PERSETUJUAN SKRIPSI

### **CLUSTERING PENYAKIT KEHAMILAN MENGUNAKAN METODE *K-MEANS* PADA PUSKESMAS BOLIOHUTO**

Oleh

NAZLY IBRAHIM

T3122223

Diperiksa oleh Panitia Ujian Sastra Satu (S1)

Universitas Ichsan Gorontalo


1. Ketua Penguji  
Haditsah Annur, M.Kom
2. Anggota  
Maryam Hasan, M.Kom
3. Anggota  
Syarifah Fitrah Ramadhani, S.Pd., M.Kom
4. Anggota  
Rezqiwati Ishak, M.Kom
5. Anggota  
Yulianty Lasena M.Kom

Mengetahui

Dekan Fakultas Ilmu Komputer

Ketua Program Studi

  
**Irvan Braham Saalihi, M.Kom**  
NIDN: 0928028101

  
**Sudirman S. Pana, M.Kom**  
NIDN: 094038205

## 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 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 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, .....2024  
Yang Membuat Pernyataan



Nazly Ibrahim

## **ABSTRACT**

### **NAZLY IBRAHIM. T3122223. CLUSTERING OF PREGNANCY DISEASES USING THE K-MEANS METHOD AT BOLIYOHUTO PUBLIC HEALTH CENTER**

*This study examines the application of the K-Means method for clustering pregnancy diseases to identify categories or classes of diseases that are susceptible to pregnant women based on age and type of disease. The problem faced is the difficulty in knowing the diseases that often occur in pregnant women, considering the large amount of data inputted manually using Microsoft Excel. The clustering of pregnancy diseases is carried out using the K-Means method to overcome it. The K-Means algorithm used in this study is a non-hierarchical algorithm of the data clustering method. This algorithm begins with cluster partitions at the beginning, followed by the partitions iteratively repaired until there is no significant change in the cluster. The data analyzed include the age of the pregnant woman and the type of disease suffered. Applying the K-Means method shows the results of three optimum clusters with a Davies-Bouldin Index (DBI) value of 1.65, indicating good clustering validity. The clustering analysis reveals that the first cluster is dominated by Hepatitis cases (0), showing that pregnant women in this cluster may be healthier or only suffer from Hepatitis. The second cluster shows a more even distribution of diseases with high frequencies of Hypertension (4), Constipation (5), and Flu (6), indicating the need for comprehensive medical intervention for these diseases. The third cluster has a wider distribution of diseases with two primary peaks, reflecting a more diverse variety of diseases, including chronic diseases or significant complications that require more intensive medical attention. Based on the results, the K-Means method has proven to be effective in grouping pregnancy diseases so that it can provide useful guidance for medical intervention and decision-making to improve the health of pregnant women.*

**Keywords** : clustering, pregnancy diseases, K-Means, Davies-Bouldin Index (DBI), pregnant women's health



## ABSTRAK

### **NAZLY IBRAHIM. T3122223. CLUSTERING PENYAKIT KEHAMILAN MENGGUNAKAN METODE K-MEANS PADA PUSKESMAS BOLIYOHUTO**

Penelitian ini mengkaji penerapan metode *K-Means* untuk *clustering* penyakit kehamilan dengan tujuan mempermudah identifikasi kategori atau kelas penyakit yang rentan dialami oleh ibu hamil berdasarkan usia dan jenis penyakit. Masalah yang dihadapi adalah kesulitan dalam mengetahui penyakit yang sering terjadi pada ibu hamil mengingat banyaknya data yang diinput secara manual menggunakan Microsoft Excel. Untuk mengatasi masalah ini, dilakukan *clustering* penyakit kehamilan menggunakan metode *K-Means*. Algoritma *K-Means* yang digunakan dalam penelitian ini merupakan algoritma non-hirarki dari metode data *clustering*. Algoritma ini dimulai dengan pembentukan partisi klaster di awal, kemudian secara iteratif partisi tersebut diperbaiki hingga tidak terjadi perubahan signifikan pada klaster. Data yang dianalisis meliputi usia ibu hamil dan jenis penyakit yang diderita. Hasil dari penerapan metode *K-Means* menunjukkan bahwa terdapat tiga *cluster* yang optimum dengan nilai Davies-Bouldin Index (DBI) sebesar 1,65, menandakan validitas *clustering* yang baik. Analisis *clustering* mengungkap bahwa *cluster* pertama didominasi oleh kasus Hepatitis (0), mengindikasikan bahwa ibu hamil dalam *cluster* ini mungkin lebih sehat atau hanya menderita Hepatitis. *Cluster* kedua menunjukkan distribusi penyakit yang lebih merata dengan frekuensi tinggi pada penyakit Hipertensi (4), Sembelit (5), dan Flu (6), yang menunjukkan perlunya intervensi medis yang komprehensif untuk penyakit-penyakit tersebut. *Cluster* ketiga memiliki distribusi penyakit yang lebih luas dengan dua puncak utama, mencerminkan variasi penyakit yang lebih beragam dan mencakup penyakit kronis atau komplikasi signifikan, yang memerlukan perhatian medis lebih intensif. Dengan hasil ini, metode *K-Means* terbukti efektif dalam mengelompokkan penyakit kehamilan, sehingga dapat memberikan panduan yang berguna untuk intervensi medis dan pengambilan keputusan dalam upaya meningkatkan kesehatan ibu hamil.

Kata kunci : *clustering*, penyakit kehamilan, *K-Means*, *Davies-Bouldin Index (DBI)*, kesehatan ibu hamil

## KATA PEGANTAR

Puji syukur penulis panjatkan atas segala rahmat dan karunia yang telah diberikan oleh Allah SWT, karena pada kesempatan ini penulis dapat menyelesaikan Skripsi ini dengan judul “***Clustering Penyakit Kehamilan Menggunakan Metode K-Means Pada Puskesmas Boliyohuto***” untuk memnuhi salah satu syarat Ujian Akhir agar bisa mendapatkan gelar Sarjana Komputer di Universitas Ichsan Gorontalo.

Terselesaikannya Skripsi ini tentu berkat dukungan dan doa dari banyak pihak yang membantu penulis. Oleh karena itu, pada kesempatan kali ini penulis ingin mengucapkan banyak terima kasih khususnya kepada :

1. Dr. Yuriko 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, M.Kom, selaku Dekan Fakultas Ilmu Komputer Universitas Ichsan Gorontalo;
4. Bapak Sudirman Melangi M.Kom selaku Pembantu Dekan I Bidang Akademik Fakultas Ilmu Komputer Universitas Ichsan Gorontalo;
5. Bapak Irma Surya Kumala Idris, M.Kom, selaku Pembantu Dekan II Bidang Administrasi Umum dan Keuangan Fakultas Ilmu Komputer Universitas Ichsan Gorontalo;
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 Utama yang telah meluangkan waktu ditengah kesibukan beliau, memberi saran dan pengarahan kepada penulis dalam proses penulisan Skripsi;

8. Ibu Yulianty Lasena, M.Kom, selaku Pembimbing Pendamping yang telah meluangkan waktu ditengah kesibukan beliau, dalam penelitian ini yang telah membimbing penulis, memberikan arahan dan masukan selama penyusunan Skripsi;
  9. Bapak dan Ibu Dosen Universitas Ichsan Gorontalo yang telah mendidik dan mengajarkan berbagai disiplin ilmu kepada penulis;
  10. Kedua orang tua penulis, Anis Ibrahim dan Isna Nur, terima kasih atas segala kasih sayang, doa, dukungan, motivasi, didikan dan semangat yang sangat berarti. Semoga penulis dapat menjadi anak yang selalu berbakti kepada orang tua;
  11. Rekan-rekan seperjuangan yang telah banyak memberikan bantuan dan dukungan moril yang sangat besar kepada penulis;
  12. Kepada semua pihak yang ikut membantu dalam menyelesaikan Skripsi ini yang tak sempat penulis sebutkan satu-persatu;
- Semoga Allah SWT, melimpahkan balasan atas jasa-jasa mereka kepada kami. 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, Juni 2024

Penulis

## DAFTAR ISI

<b>PERSETUJUAN SKRIPSI.....</b>	<b>ii</b>
<b>PENGESAHAN SKRIPSI.....</b>	<b>iii</b>
<b>PERNYATAAN SKRIPSI.....</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>ix</b>
<b>DAFTAR GAMBAR.....</b>	<b>xii</b>
<b>DAFTAR TABEL .....</b>	<b>xiii</b>
<b>DAFTAR <i>PSEUDOCODE</i>.....</b>	<b>xiv</b>
<b>BAB I PENDAHULUAN.....</b>	<b>1</b>
1.1 Latar Belakang .....	1
1.2 Identifikasi Masalah .....	3
1.3 Rumusan Masalah .....	3
1.4 Tujuan Penelitian .....	3
1.5 Manfaat Penelitian .....	4
<b>BAB II LANDASAN TEORI .....</b>	<b>5</b>
2.1 Tinjauan Studi .....	5
2.2 Tinjauan Pustaka .....	6
2.3 Perangkat Lunak Pendukung .....	16
2.4 Kerangka Berpikir .....	17
<b>BAB III METODE PENELITIAN .....</b>	<b>18</b>
3.1 Jenis, Metode, Subjek, Waktu dan Lokasi Penelitian .....	18
3.2 Pengumpulan Data .....	18

3.3 Pemodelan .....	19
<b>BAB IV HASIL PENELITIAN.....</b>	<b>20</b>
4.1 Pengumpulan Data .....	20
4.2 Hasil Pemodelan .....	20
4.3 Implementasi Hasil Pemodelan.....	25
<b>BAB V HASIL DAN PEMBAHASAN .....</b>	<b>33</b>
5.1 Pembahasan Kinerja Model .....	33
5.2 Pembahasan Hasil Pemodelan .....	35
<b>BAB VI.....</b>	<b>41</b>
6.1 Kesimpulan .....	41
6.2 Saran.....	41
<b>DAFTAR PUSTAKA .....</b>	<b>42</b>
<b>LAMPIRAN</b>	

## DAFTAR GAMBAR

Gambar 2.1 : Proses <i>Knowledge Discovery In Database</i> (KDD).....	7
Gambar 2.2 : Visualisasi Hasil Metode <i>Elbow</i> .....	13
Gambar 2.3 : Visualisasi Hasil <i>Cluster</i> .....	14
Gambar 2.4 : Informasi Hasil <i>Cluster</i> .....	14
Gambar 2.5 : Kerangka Berpikir .....	17
Gambar 5.1 : Visualisasi Teknik <i>Elbow</i> .....	33
Gambar 5.2 : Visualisasi Perhitungan DBI .....	34
Gambar 5.3 : Visualisasi Hasil <i>Cluster</i> Untuk $K=3$ .....	35
Gambar 5.8 : Hasil Proses Perhitungan Deskripsi Statistik .....	37
Gambar 5.9 : Distribusi Penyakit Setiap <i>Cluster</i> .....	38

## DAFTAR TABEL

Tabel 1.1 : Data Pasien Ibu Hamil .....	2
Tabel 2.1 : Tinjauan Studi.....	5
Tabel 2.2 : Sampel Dataset Nilai Kalkulus .....	11
Tabel 2.3 : Dataset Nilai Matematika Diskrit .....	11
Tabel 2.4 : Sampel Hasil Penggabungan Dataset.....	12
Tabel 2.5 : Sampel Hasil Nilai Huruf Ke Angka .....	12
Tabel 2.6 : <i>Centroid</i> dan Jumlah Data .....	14
Tabel 2.7 : Sampel Hasil <i>Clustering</i> .....	15
Tabel 2.8 : Perangkat Lunak Pendukung .....	16
Tabel 3.1 : Atribut/Variabel Data .....	18
Tabel 4.1 : Dataset Awal Pasien Ibu Hamil .....	20
Tabel 4.2 : Transformasi Data Atribut Penyakit Kehamilan.....	21
Tabel 4.3 : Menghapus Satuan Data .....	22
Tabel 4.4 : Dataset Akhir Penyakit Kehamilan.....	22
Tabel 4.5 : Standarisasi Data Sebelum <i>Clustering</i> .....	21
Tabel 4.6 : Hasil Pemodelan Dengan <i>K-Means</i> .....	24
Tabel 4.7 : Hasil Clustering Dengan <i>K-Means</i> .....	24
Tabel 4.8 : Deskripsi <i>Pseudocode Import Library Python</i> .....	25
Tabel 5.1 : Hasil Perhitungan DBI Untuk Masing-masing <i>Cluster</i> .....	34
Tabel 5.2 : Deskripsi Statistic Hasil <i>Clustering</i> .....	37
Tabel 5.3 : Hasil Analisis Clustering K=3 .....	38
Tabel 5.4 : Daftar Penyakit Ibu Hamil .....	39
Tabel 5.5 : Analisa Distribusi Jenis Penyakit .....	39
Tabel 5.6 : Detail Hasil <i>Clustering</i> .....	40

## DAFTAR PSEUDOCODE

Pseudocode 4.1 : <i>Import Library Python</i> .....	25
Pseudocode 4.2 : <i>Meenghubungkan Google Colab dengan Google Drive</i> .....	26
Pseudocode 4.3 : <i>Load Dataset</i> .....	26
Pseudocode 4.4 : <i>Menghapus Satuan Pada Kolom Umur</i> .....	26
Pseudocode 4.5 : <i>Menghapus Satuan Pada Kolom Berat Badan</i> .....	27
Pseudocode 4.6 : <i>Menghapus Satuan Pada Kolom Tinggi Badan</i> .....	27
Pseudocode 4.7 : <i>Menghapus Satuan Pada Kolom Usia Kehamilan</i> .....	27
Pseudocode 4.8 : <i>Menghapus Satuan Kolom Pada Sistole</i> .....	28
Pseudocode 4.9 : <i>Menghapus Satuan Kolom Pada Diastole</i> .....	28
Pseudocode 4.10 : <i>Mengkonversi Kolom Penyakit Ke Angka Secara Otomatis</i> .....	28
Pseudocode 4.11 : <i>Lihat Dataset</i> .....	28
Pseudocode 4.12 : <i>Baca Dataset Pasien Ibu Hamil Reprocessing</i> .....	29
Pseudocode 4.13 : <i>Baca Dataset Standarisasi Sebelum Klastering</i> .....	29
Pseudocode 4.14 : <i>Data Frame Dari Data Yang Telah Distandarisasi</i> .....	29
Pseudocode 4.15 : <i>Pemodelan Melakukan K-Means</i> .....	30
Pseudocode 4.16 : <i>Melakukan PCA Untuk Visualisasi</i> .....	30
Pseudocode 4.17 : <i>Plot Hasil Clustering</i> .....	30
Pseudocode 4.18 : <i>Menentukan Jumlah Cluster Yang Optimum Dengan Teknik Elbow</i> .....	31
Pseudocode 4.19 : <i>Menentukan Jumlah Cluster Yang Optimum Dengan Teknik David-Bouldin-Index</i> .....	32
Pseudocode 4.20 : <i>Melakukan Clustering K-Means</i> .....	32

# **BAB I**

## **PENDAHULUAN**

### **1.1 Latar Belakang**

Kesehatan sangat penting selama kehamilan, baik sebelum maupun selama masa kehamilan. Risiko kehamilan bisa meningkat jika wanita tersebut dalam keadaan sakit. Jika seorang wanita hamil memiliki kondisi kesehatan yang kurang baik atau sedang menderita penyakit tertentu, maka penting baginya untuk mendapatkan bantuan medis guna merencanakan kebutuhan yang diperlukan serta menentukan tempat yang aman untuk persalinan[1].

Upaya pemerintah dalam memberikan pelayanan kesehatan kepada masyarakat sangat bergantung pada peran puskesmas. Sebagai unit pelayanan kesehatan yang berada di garis depan, puskesmas memiliki misi untuk menjadi pusat pengembangan pelayanan kesehatan. Tugas utamanya adalah melaksanakan pembinaan dan memberikan pelayanan kesehatan yang menyeluruh dan terpadu kepada masyarakat di wilayah tertentu[2].

Tujuan pembangunan kesehatan yang diselenggarakan oleh puskesmas adalah mendukung tercapainya tujuan pembangunan kesehatan nasional, yakni meningkatkan kesadaran, kemauan dan kemampuan hidup sehat bagi setiap orang yang bertempat tinggal diwilayah kerja puskesmas, agar terwujud derajat kesehatan yang setinggi-tingginya dalam rangka mewujudkan indonesia sehat[2].

Puskesmas Boliyohuto merupakan salah satu puskesmas yang berada di kabupaten Gorontalo yang setiap harinya melayani banyak pasien ibu hamil dengan berbagai jenis penyakit. Pada penelitian ini penulis ingin mengelompokkan penyakit kehamilan dengan menggunakan metode *k-means* agar bisa terkelompok dengan kategori masing-masing jenis penyakit kehamilan.

Metode *k-means* merupakan salah satu metode data *clustering* non hirarki yang berusaha mempartisi data yang ada kedalam bentuk satu atau lebih *cluster*/kelompok. Metode ini mempartisi data ke dalam *cluster*/kelompok sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam satu *cluster* yang sama dan data yang mempunyai karakteristik yang berbeda dikelompokkan

ke dalam kelompok yang lain[3]. Penggunaan metode k-means pada penelitian ini digunakan untuk mengetahui clusterisasi data penyakit kehamilan dengan mengelompokkan penyakit kedalam kelompok-kelompok berdasarkan Pengelompokkan penyakit pasien ibu hamil dilakukan berdasarkan usia, gejala, dan diagnosa penyakit.

Adapun data pasien ibu hamil sebagai berikut :

**Tabel 1.1** Data Pasien Ibu Hamil

No	Usia Ibu Hamil	Berat Badan	Tinggi Badan	Usia Kehamilan	Tekanan Darah		Penyakit
					Sistole	Diastole	
1	28 Tahun	70 Kg	166 Cm	37 Minggu	140 mm	90 Hg	Hepatitis
2	35 Tahun	77 Kg	169 Cm	38 Minggu	75 mm	50 Hg	Anemia
3	23 Tahun	67 Kg	153 Cm	36 Minggu	70 mm	50 Hg	Anemia
4	23 Tahun	60 Kg	165 Cm	9 Minggu	121 mm	80 Hg	Plasenta Previa
5	24 Tahun	75 Kg	160 Cm	36 Minggu	122 mm	81 Hg	Infeksi Saluran Kemih
6	20 Tahun	79 Kg	171 Cm	38 Minggu	150 mm	10 Hg	Hipertensi
7	23 Tahun	72 Kg	164 Cm	36 Minggu	120 mm	79 Hg	Sembelit
8	29 Tahun	66 Kg	152 Cm	11 Minggu	110 mm	77 Hg	Flu
9	22 Tahun	59 Kg	160 Cm	14 Minggu	120 mm	79 Hg	Keputihan
10	25 Tahun	71 Kg	167 Cm	19 Minggu	80 mm	60 Hg	Pendarahan
11	30 Tahun	78 Kg	166 Cm	36 Minggu	140 mm	90 Hg	Hepatitis
12	19 Tahun	53 Kg	147 Cm	34 Minggu	160 mm	10 Hg	Hipertensi
13	25 Tahun	60 Kg	154 Cm	36 Minggu	100 mm	80 Hg	Diabetes Gestasional
14	31 Tahun	70 Kg	160 Cm	34 Minggu	84 mm	69 Hg	Anemia
15	28 Tahun	78 Kg	150 Cm	26 Minggu	130 mm	80 Hg	Kehamilan Ektopik
16	25 Tahun	69 Kg	149 Cm	34 Minggu	120 mm	90 Hg	Diabetes Gestasional
17	25 Tahun	71 Kg	154 Cm	34 Minggu	110 mm	80 Hg	Diare
18	26 Tahun	63 Kg	148 Cm	14 Minggu	80 mm	50 Hg	Anemia
...	...	...	...	...	...	...	...
274	30 Tahun	68 Kg	152 Cm	20 Minggu	148 mm	97 Hg	Hipertensi

Sumber : Puskesmas Boliyohuto

Berdasarkan tabel diatas dapat dilihat berbagai macam usia ibu hamil yang terkena penyakit serta jenis penyakit yang diderita ibu hamil. Data pemasukan ibu hamil di input secara manual menggunakan *Microsoft Excel*. Masalah dari penelitian ini, dengan banyaknya data yang ada akan sulit dalam mengetahui penyakit yang rentan dialami oleh ibu hamil. Maka dengan adanya *clustering* penyakit kehamilan dengan menggunakan metode k-means bisa mempermudah dalam mengetahui kategori atau kelas-kelas dari masing-masing pasien ibu hamil.

Pada penelitian sebelumnya juga pernah dilakukan oleh Rezqiwati Ishak dan Amiruddin yang berjudul “*Clustering* Tingkat Pemahaman Dasar Mahasiswa Pada Pra Perkuliahan Probabilitas Statistika Dengan Metode *K-means*”, yang

hasilnya terdapat 3 *cluster* Penelitian ini bertujuan menyelesaikan pengelompokan tingkat pemahaman dasar mahasiswa berdasarkan mata kuliah prasyarat yaitu kalkulus dan Matematika Diskrit. Hasil penelitian diperoleh hasil sejumlah 3 *cluster*, dimana *cluster* 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 110 mahasiswa yang menjadi sebagai dataset, masing-masing ditemukan untuk *cluster* 1 sejumlah 30 mahasiswa, *cluster* 2 sejumlah 58 mahasiswa dan *cluster* 3 sejumlah 22 mahasiswa. Berdasarkan *cluster* tersebut maka mahasiswa yang masuk pada 3 *cluster* perlu dilakukan tambahan perkuliahan sebelum mereka mengikuti mata kuliah probabilitas statistik[4]. Dengan demikian penerapan metode k-means dapat digunakan dalam proses *clustering* penyakit kehamilan pada puskesmas baliyohuto.

Berdasarkan uraian diatas , maka peneliti tertarik untuk melakukan penelitian dengan judul : **“Clustering Penyakit Kehamilan Menggunakan Metode K-Means Pada Puskesmas Baliyohuto”**

### **1.1 Identifikasi Masalah**

Berdasarkan yang telah diuraikan diatas maka di identifikasikan permasalahan dalam penelitian ini :

Sulitnya melakukan *clustering* dalam penyakit yang rentan dialami oleh ibu hamil.

### **1.2 Rumusan Masalah**

1. Bagaimana hasil penerapan metode k-means untuk *Clustering* penyakit kehamilan pada puskesmas baliyohuto?
2. Bagaimana hasil analisa *Clustering* penyakit kehamilan pada puskesmas baliyohuto?

### **1.3 Tujuan Penelitian**

1. Menerapkan metode *K-Means* untuk *Clustering* Penyakit Kehamilan Pada Puskesmas Baliyohuto

2. Menganalisa hasil *Clustering* Penyakit Kehamilan Pada Puskesmas Boliyohuto

#### **1.4 Manfaat Penelitian**

hasil penelitian ini diharapkan dapat berguna, adapun manfaatnya adalah sebagai berikut :

1. Secara teoritis, menambah dan mengembangkan ilmu pengetahuan tentang sistem yang dibuat dan serta sebagai modal dasar untuk karya yulis kedepannya.
2. Secara Praktis, diharapkan dapat memberikan banyak keuntungan yang meningkatkan efektivitas penelitian dan aplikasinya dalam menangani penyakit kehamilan

## BAB II

### LANDASAN TEORI

#### 2.1 Tinjauan Studi

Tinjauan studi terkait dengan penelitian ini dapat dilihat pada tabel 2.1 sebagai berikut :

**Tabel 2.1** Tinjauan Studi

Penelitian / Tahun	Judul	Metode	Hasil
Rezqiwati Ishak, Amiruddin / 2022[4]	<i>Clustering</i> Tingkat Pemahaman Dasar Mahasiswa Pada Pra Perkuliahan Probabilitas Statistika Dengan Metode <i>K-means</i> [4].	Metode K-Means	Penelitian ini bertujuan menyelesaikan pengelompokan tingkat pemahaman dasar mahasiswa berdasarkan mata kuliah prasyarat yaitu kalkulus dan Matematika Diskrit. Hasil penelitian diperoleh hasil sejumlah 3 <i>cluster</i> , dimana <i>cluster</i> 1 ialah kelompok mahasiswa yang kualitas pemahamannya dikategorikan baik, <i>cluster</i> 2 ialah kelompok yang kategorinya sangat baik dan <i>cluster</i> 3 ialah yang kategorinya kurang. Dari 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].
Muhamad Rivalda, Erti Maulina Hidayat, Muhamad Azhar Gunawan & Diaz Defriyanto / 2023[5]	Penerapan Metode <i>Clustering</i> Dalam Upaya Pencegahan Penyakit Demam Berdarah Menggunakan Algoritma <i>K-means</i> (Studi Kasus :Kota Tasikmalaya)[5]	Metode K-Means	Algoritma <i>K-means</i> dapat digunakan untuk penerapan upaya pencegahan penyakit DBD dengan mempertimbangkan tingkat kejadian DBD di suatu wilayah[5]

Faisal Dikarya & Sita Muharni / 2022[6]	Penerapan Algoritma <i>K-means Clustering</i> Untuk Pengelompokan Universitas Terbaik Didunia[6]	Metode K-Means	Pengujian dari penelitian ini dilakukan iterasi <i>clustering</i> dengan data 2000 universitas teratas didunia terjadi sebanyak 10 kali terdapat tiga <i>cluster</i> yaitu <i>cluster</i> rendah, <i>cluster</i> sedang, <i>cluster</i> tinggi[6].
Gustientiedina, M. Hasnil Adiya & Yenny Desnelita / 2019[7]	Penerapan Algoritma <i>K-means</i> untuk <i>Clustering</i> Data Obat-obatan pada RSUD Pekanbaru[7]	Metode K-Means	Dari hasil <i>clusterisasi</i> pada obat-obatan dapat ditarik kesimpulan bahwa kelompok obat yang termasuk pemakaian sedikit rata-rata permintaan obat setiap tahunnya kurang dari 18000 buah, dan obat yang termasuk pemakaian sedang rata-rata permintaan obat setiap tahunnya diantara 18000-70000, sedangkan obat yang termasuk kedalam kelompok obat yang pemakaian tinggi rata-rata permintaan obat setiap tahunnya diatas 70000 buah[7].

## 2.2 Tinjauan Pustaka

### 2.2.1 Penyakit Kehamilan

Kehamilan adalah dikandungnya janin hasil pembuahan sel telur oleh sel sperma kehamilan merupakan suatu proses yang alamiah dan fisiologi. Kehamilan didefinisikan sebagai *fertilisasi* atau penyatuan dari *spermatozoa* dan *ovum* dan dilanjutkan dengan nidasi atau implantasi. Bila dihitung dari saat fertilisasi hingga lahirnya bayi, kelahiran normal akan berlangsung dalam waktu 40 minggu atau 10 bulan lunar atau 9 bulan menurut kalender internasional.

Berbagai macam penyakit yang rentan terjadi pada masa sebelum dan selama kehamilan, baik itu penyakit yang bersifat sedang atau yang dapat menyebabkan kematian[8].

### 2.2.2 Data Mining

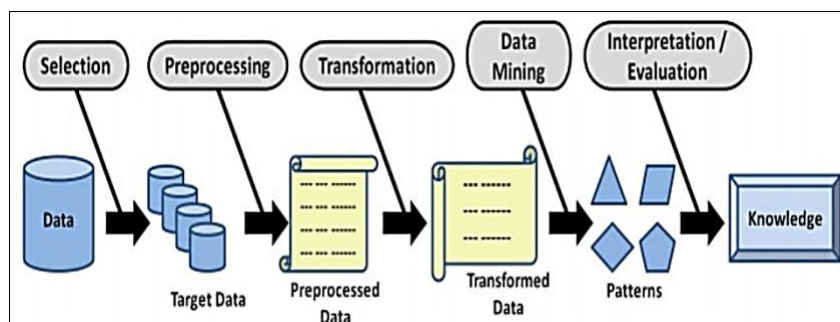
Data mining adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan machine learning untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar. Data mining merupakan serangkaian proses untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui secara manual[9].

Data mining dibagi menjadi dua kategori utama yaitu[9];

- Prediktif, tujuan dari tugas prediktif adalah untuk memprediksi nilai dari atribut tertentu berdasarkan pada nilai atribut-atribut lain.
- Deskriptif, tujuan dari tugas deskriptif adalah untuk menurunkan pola-pola (korelasi, trend, *cluster*, teritori, dan anomali) yang meringkas hubungan yang pokok dalam data.

### 2.2.3 Knowledge Discovery In Database (KDD)

Kdd adalah metode yang digunakan untuk dapat memperoleh pengetahuan yang berasal dari database yang ada. Hasil pengetahuan yang diperoleh dapat dimanfaatkan untuk basis pengetahuan (*knowledge base*) yang digunakan dalam keperluan mengambil keputusan secara lebih detail proses KDD[7], seperti pada gambar dibawah ini :



**Gambar 2.1** Proses *Knowledge Discovery In Database* (KDD)

1. *selection*

digunakan untuk menentukan variabel yang akan diambil agar tidak ada kesamaan dan terjadi perulangan yang tidak diperlukan dalam pengolahan data mining.

2. *Preprocessing*

Pada preprocessing terdapat dua tahap, yaitu sebagai berikut :

a. *Data Cleaning*

Menghilangkan data yang tidak diperlukan seperti menangani data-data yang tidak konsisten dan relevan

b. *Data Integration*

Dilakukan terhadap atribut yang mengidentifikasi entitas yang unik.

3. *Transformation*

Merubah data sesuai format ekstention yang sesuai dalam pengolahan data mining memerlukan format khusus sebelum dapat diproses pada data mining.

4. *Data Mining*

Proses utama pada metode yang diterapkan untuk mendapatkan pengetahuan baru dari data yang diproses. Pada penelitian ini diterapkan teknik *clustering* yaitu metode *k-means clustering*.

5. *Evaluation/Interpretation*

Mengidentifikasi pola-pola yang menarik kedalam *knowledge base* yang diidentifikasi . Pada tahap ini menghasilkan pola-pola khas maupun mode prediksi yang dievaluasi untuk menilai kajian yang ada sudah memenuhi target yang diinginkan.

#### **2.2.4 Clustering**

Metode *clustering* adalah mempartisi data kedalam kelompok sehingga data yang dimiliki karakteristik yang sama dikelompokkan kedalam satu *cluster* yang sama. Tujuan dari *clustering* ini adalah untuk meminimalisir fungsi tujuan yang diterapkan dalam proses *clustering*. Yang umumnya berusaha meminimalisir variasi dalam suatu *cluster* dan memaksimalkan variasi antar *cluster*[5].

### 2.2.5 Algoritma K-Means

Algoritma *k-means* merupakan algoritma non hirarki yang berasal dari metode data *clustering*. Algoritma *k-means* dimulai dengan pembentukan partisi kluster diawal kemudian secara iteratif partisi *cluster* ini diperbaiki hingga tidak terjadi perubahan yang signifikan pada partisi *cluster*[10].

Langkah-langkah melakukan clustering dengan metode *k-means* sebagai berikut[10] :

1. tentukan nilai k sebagai jumlah kluster yang ingin dibentuk.
2. Inisialisasi k pusat kluster ini bisa dilakukan dengan berbagai cara, namun yang paling sering dilakukan adalah dengan cara random yang diambil dari data yang ada.
3. Menghitung jarak setiap data input terhadap masing-masing centroid menggunakan rumus jarak Euclidean (Euclidean Distance) hingga ditemukan jarak yang paling dekat dari setiap data dengan centroid. Berikut adalah persamaan Euclidean Distance :

$$De = \sqrt{(x_i - s_i)^2 + (y_i - t_i)^2}$$

Dimana :

De adalah *Euclidean Distance*

I adalah banyaknya objek

(x,y) merupakan koordinat object dan

(s,t) merupakan koordinat centroid.

4. mengklasifikasikan setiap data berdasarkan kedekatan dengan centroid (jarak terkecil).
5. memperbaharui nilai centroid. Nilai *centroid* baru diperoleh dari rata-rata *cluster* yang bersangkutan dengan menggunakan rumus :

$$v_{ij} = \frac{1}{N_i} \sum_{k=0}^{N_i} X_{kj}$$

dimana :

vij adalah centroid/ rata-rata *cluster* ke-i untuk variable ke-j

ni adalah jumlah data yang menjadi anggota *cluster* ke-i

i,k adalah indeks dari *cluster*

j adalah indeks dari variable

x<sub>kj</sub> adalah nilai data ke-k yang ada didalam *cluster* tersebut untuk variable ke-j

6. melakukan perulangan dari langkah 2 hingga langkah 5 sampai anggota tiap *cluster* tidak ada yang berubah.

### 2.2.5 Metode *Elbow*

Metode elbow merupakan salah satu metode yang dapat digunakan untuk menentukan jumlah cluster terbaik, yaitu dengan cara melihat presentase setiap cluster yang akan membentuk siku pada suatu titik tertentu. Metode elbow biasa disajikan dalam bentuk grafik untuk mengetahui lebih jelas siku yang terbentuk. Tujuan dari metode elbow adalah untuk memilih nilai k yang kecil dan masih memiliki nilai withinss yang rendah. Nilai k pada kombinasi siku dengan k-means adalah grafik hubungan cluster dengan penurunan error. Jumlah cluster k yang dihasilkan dari pengujian dengan k-means dievaluasi dengan teknik SSE. SSE (Sum of Square Error) merupakan rumus yang digunakan untuk mengukur perbedaan antara data yang telah dilakukan sebelumnya[11].

### 2.2.6 Contoh Penerapan Algoritma *K-Means*

Berikut ini adalah contoh penerapan algoritma k-means pada penelitian sebelumnya oleh Rezkiwati Ishak dan Amiruddin yang berjudul “ *Clustering* Tingkat Pemahaman Dasar Mahasiswa Pada Pra-Perkuliahan Probabilitas Statistika Dengan Metode *K-Means*”[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	T3220004	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*

Pada 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 :

**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*

Pada tahapan ini melakukan konversi data nilai ialah nilai maata kuliah yang mana bentuk huruf dikonversi kebentuk angka yaitu nilai A=4, B=3, C=2, D=1, keadaan ini dilakukan agar bisa menggunakan pemodelan metode k-means. Tahapan ini lagsung dilakukan di dalam Python dengan hasil akhir seperti berikut:

**Tabel 2.5** Sampel Hasil Nilai Huruf ke Angka

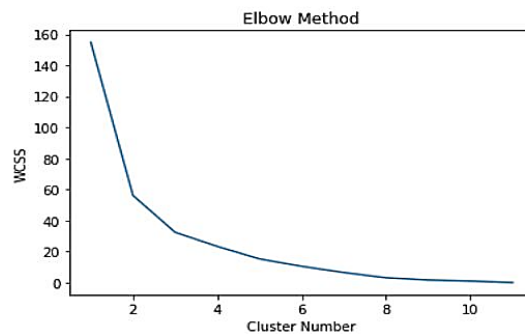
NO	NIM	KALKULUS	M_DISKRIT
1	T3119101	4	4
2	T3120002	3	2
3	T3120003	4	1
4	T3120004	4	4
5	T3120005	3	1
6	T3120006	4	4
7	T3120007	4	3
8	T3120008	4	3
9	T3120009	4	4
10	T3120010	3	4

4. Tahap *Data Mining*

Pada tahap ini 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 :



**Gambar 2.2** Visualisasi Hasil Metode Elbow

Pada gambar diatas 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 dimasukan kedalam fungsi k-means yang tersedia pada library scikit learn dengan tahapan sebagai berikut:

1. baca dataset yang sudah diproses:

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

2. panggil fungsi k-means:

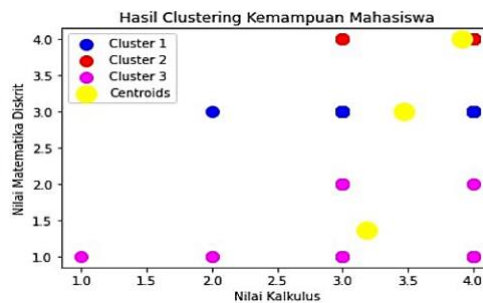
```
k-means = Kmeans(n_clusters = 3, init =k-means++)
```

3. terapkan hasil pemodelan terhadap dataset:

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

4. lakukan visualisasi hasil *clusterin* atau pemodelan

berikut visualisai hasil *clustering* berdasarkan pemodelan k-means yang sudah dilakukan:



**Gambar 2.3** Visualisasi Hasil *cluster*

Sesuai pada gambar diatas, dapat ditemukan 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 2.4** Informasi Hasil *Cluster*

Pada jumlah anggota masing-masing *cluster* dapat diketahui dengan menggunakan perintah “dfData[‘Kluster’].value\_counts()”. Berdasarkan kedua perintah diatas dapat dimodel hasilnya dalam bentuk tabel seperti pada tabel berikut ini:

**Tabel 2.6** Centroid dan Jumlah data

CLUSTER	CENTROID		JUMLAH MAHASISWA
	KALKULUS	MATEMATIKA DIKSTRIT	
1	3,46	3	30
2	3,91	4	58
3	3,18	1,36	22

Berdasarkan tabel diatas, 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 didownload dari hasil pemodelan di aplikasi Python bisa dilihat pada tabel dilihat ini:

**Tabel 2.7** Sampel Hasil *Clustering*

NIM	KAKULUS	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
T3120008	4	3	1
T3120009	4	3	1
T3120010	4	4	2
T3120011	3	4	1

Berdasarkan informasi tabel diatas, 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.

### 2.2.8 Python

*Python* merupakan salah satu bahasa pemrograman yang banyak digunakan oleh perusahaan besar maupun para developer untuk mengembangkan berbagai macam aplikasi berbasis desktop, web dan mobile. *Python* diciptakan oleh *Guido van Rossum* di Belanda pada tahun 1990 dan namanya diambil dari acara televisi kesukaan *Guido Monty Python's Flying Circus*. *Van Rossum* mengembangkan *Python* sebagai hobi, kemudian *Python* menjadi bahasa pemrograman yang dipakai secara luas dalam industri dan pendidikan karena sederhana, ringkas, sintak intuitif dan memiliki pustaka yang luas[12].

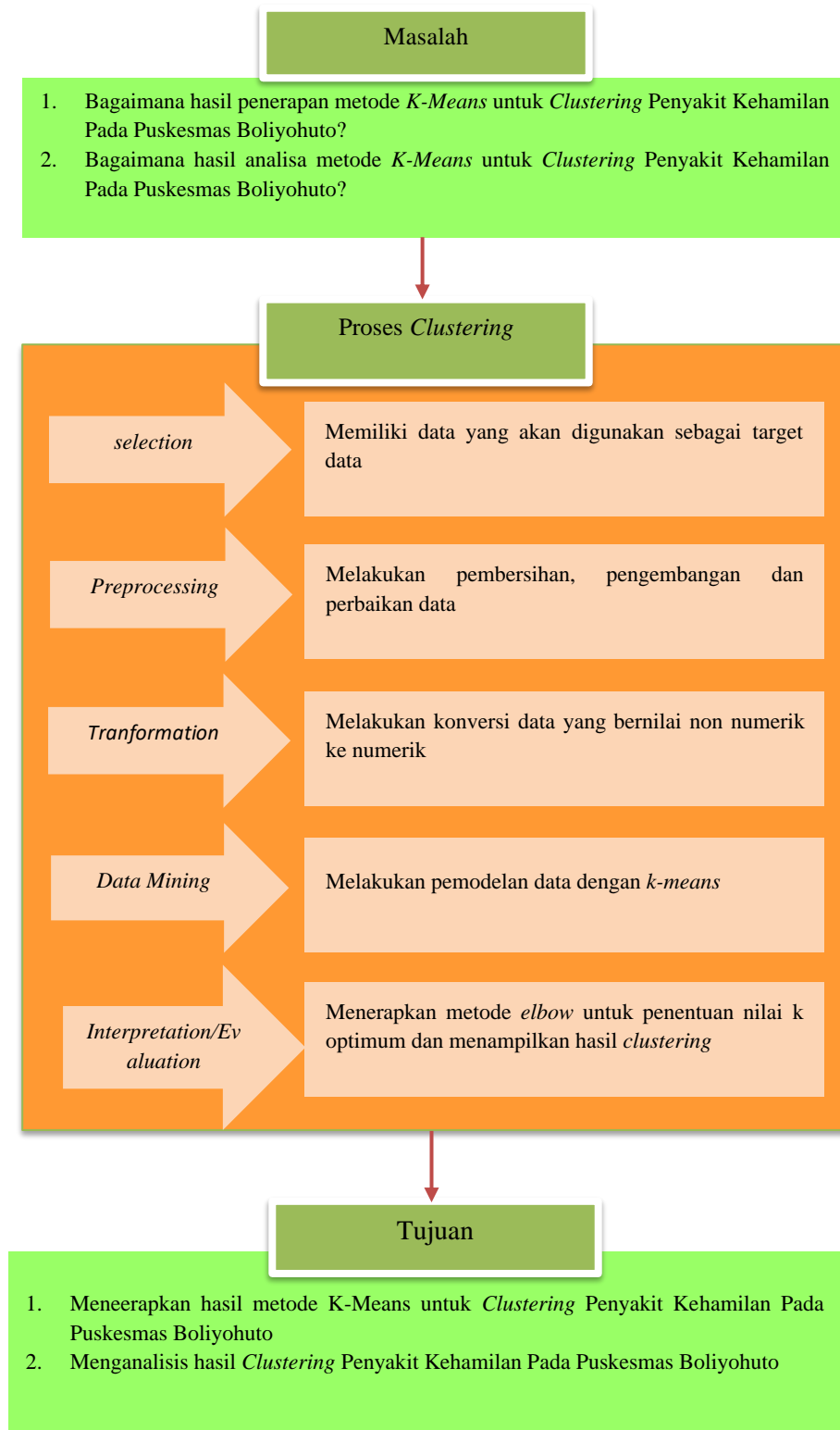
### 2.2.9 Perangkat Lunak Pendukung

Perangkat lunak pendukung yang digunakan dalam membuat *clustering*, seperti pada tabel dibawah ini:

**Tabel 2.8** Perangkat Lunak Pendukung

No	Perangkat Lunak	Keterangan
1	Python	Python merupakan bahasa pemograman yang digunakan dalam membut aplikasi dan bisa dihunakan untuk melakukan analisis data. Juga digunakan untuk membuat berbagai maam program yang berbeda
2	Google colab	Untuk digunakan menjalankan kode python tanpa perlu melakukan proses instalasi

## 2.4 Kerangka Berpikir



Gambar 2.5 Kerangka Berpikir

## BAB III

### METODE PENELITIAN

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

Penelitian ini menggunakan metode penelitian kuantitatif. Dengan demikian jenis penelitian ini adalah penelitian deskriptif yaitu suatu jenis penelitian yang menggambarkan suatu keadaan yang sementara berjalan pada saat penelitian dilakukan, dan melakukan perancangan sistem pakar berdasarkan data yang ada.

Berdasarkan latar belakang dan kerangka pikir seperti yang sudah diuraikan dalam Bab I, Bab II, maka yang menjadi subjek penelitian ini adalah “*Clustering Penyakit Kehamilan Menggunakan Metode K-Means Pada Puskesmas Boliyohuto*”. Penelitian ini dimulai dari 29 Januari 2024. Yang berlokasi di Puskesmas Boliyohuto, Kab. Gorontalo.

#### 3.2 Pengumpulan data

Adapun jenis pengumpulan 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 *clustering* menggunakan metode *k-means*.

Adapun variabel/atribut dengan tipe datanya masing-masing pada tabel berikut.

**Tabel 3.1** Atribut/Variabel Data

No.	Variable/Atribut	Type	Keterangan
1	Usia ibu hamil	Varchar	Variabel input
2	Berat badan	Varchar	Variabel input
3	Tinggi badan	Varchar	Variabel input
4	Usia kehamilan	Varchar	Variabel input
5	Tekanan darah	Varchar	Variabel input
6	Penyakit	Varchar	Variabel input

### 3.3 Pemodelan

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

1. *Selection* : tahap ini dilakukan seleksi terhadap dataset yang dikumpulkan dari data jumlah pasien ibu hamil di Puskesmas Boliyohuto yang akan dijadikan target data.
2. *Preprocessing* : tahap ini melakukan proses penghapusan data yang sama, pemeriksaan data yang tidak konsisten dan melengkapi data yang belum ada.
3. *Transformation* : pada tahap ini dilaksanakan perubahan atau konversi data yaitu bentuk huruf dikonversi ke bentuk angka agar bisa digunakan untuk pemodelan *k-means*.
4. Data mining : tahap selanjutnya dilaksanakan pengembangan model model *k-means* agar menghasilkan *cluster* yang akan dibentuk.
5. *Interpretation/Evaluation* : pada tahap ini dilaksanakan penentuan beberapa hasil *cluster* yang optimal saat menggunakan metode *elbow*. Hasil nilai *K* optimum dari metode *elbow* yang kemudian diterapkan kedalam metode *k-means* kembali. Hasil dari *clustering* tersebut divisualisasikan jumlah *cluster* beserta member *clusternya*, maka dari itu mudah dipahami arti dari masing-masing kelompok *cluster* tersebut yang kemudian dijadikan sebagai pengetahuan atau *clustering* penyakit kehamilan.

## BAB IV

### HASIL PENELITIAN

#### 4.1 Pengumpulan Data

Dataset Penyakit ibu hamil didapatkan langsung dari sumbernya yaitu Puskesmas Boliyohuto sebanyak 274 *record* dalam bentuk file *Excel*. Bentuk dataset ditunjukkan pada Tabel 4.1

**Tabel 4.1** Dataset Awal Pasien Ibu Hamil

No	Usia Ibu Hamil	Berat Badan	Tinggi Badan	Usia Kehamilan	Tekanan Darah		Penyakit
					Sistole	Diastole	
1	28 Tahun	70 Kg	166 Cm	37 Minggu	140 mm	90 Hg	Hepatitis
2	35 Tahun	77 Kg	169 Cm	38 Minggu	75 mm	50 Hg	Anemia
3	23 Tahun	67 Kg	153 Cm	36 Minggu	70 mm	50 Hg	Anemia
4	23 Tahun	60 Kg	165 Cm	9 Minggu	121 mm	80 Hg	Plasenta Previa
5	24 Tahun	75 Kg	160 Cm	36 Minggu	122 mm	81 Hg	Infeksi Saluran Kemih
6	20 Tahun	79 Kg	171 Cm	38 Minggu	150 mm	10 Hg	Hipertensi
7	23 Tahun	72 Kg	164 Cm	36 Minggu	120 mm	79 Hg	Sembelit
8	29 Tahun	66 Kg	152 Cm	11 Minggu	110 mm	77 Hg	Flu
9	22 Tahun	59 Kg	160 Cm	14 Minggu	120 mm	79 Hg	Keputihan
10	25 Tahun	71 Kg	167 Cm	19 Minggu	80 mm	60 Hg	Pendarahan
11	30 Tahun	78 Kg	166 Cm	36 Minggu	140 mm	90 Hg	Hepatitis
12	19 Tahun	53 Kg	147 Cm	34 Minggu	160 mm	10 Hg	Hipertensi
13	25 Tahun	60 Kg	154 Cm	36 Minggu	100 mm	80 Hg	Diabetes Gestasional
14	31 Tahun	70 Kg	160 Cm	34 Minggu	84 mm	69 Hg	Anemia
15	28 Tahun	78 Kg	150 Cm	26 Minggu	130 mm	80 Hg	Kehamilan Ektopik
16	25 Tahun	69 Kg	149 Cm	34 Minggu	120 mm	90 Hg	Diabetes Gestasional
17	25 Tahun	71 Kg	154 Cm	34 Minggu	110 mm	80 Hg	Diare
18	26 Tahun	63 Kg	148 Cm	14 Minggu	80 mm	50 Hg	Anemia
...	...	...	...	...	...	...	...
274	30 Tahun	68 Kg	152 Cm	20 Minggu	148 mm	97 Hg	Hipertensi

#### 4.2 Hasil Pemodelan

Sebelum dilakukan pemodelan dengan metode *k-means* terlebih dahulu dilakukan beberapa *preprocessing* data terhadap dataset pada Tabel 4.1 di atas khususnya untuk atribut penyakit kehamilan dilakukan transformasi data dari jenis kategorikal menjadi numerik supaya datanya menjadi satu jenis *type* data yaitu data numerik.

##### 4.2.1 Data *Preprocessing*

###### 1. Transformasi Data untuk Atribut Penyakit Kehamilan

Transformasi nilai atribut untuk atribut penyakit kehamilan ditunjukkan pada Tabel 4.2

**Tabel 4.2** Transformasi Data Atribut Penyakit Kehamilan

No	Nilai Atribut	Nilai Konversi
1	Hepatitis	0
2	Anemia	1
3	Plasenta Previa	2
4	Infeksi Saluran Kemih	3
5	Hipertensi	4
6	Sembelit	5
7	Flu	6
8	Keputihan	7
9	Pendarahan	8
10	Hepatitis	9
11	Diabetes Gestasional	10
12	Kehamilan Ektopik	11
13	Diare	12
14	Keguguran	13
15	Rubella	14
16	Campak	15
17	Asma	16
18	Demam Berdarah (DBD)	17
19	Maag	18
20	Tipes	19
...	...	...
36	Rematik	35

Pada Tabel 4.2 diatas menunjukkan transformasi data atribut untuk atribut penyakit kehamilan. Tabel ini mencantumkan daftar penyakit kehamilan dan nilai konversi yang sesuai. Setiap penyakit kehamilan memiliki nilai atribut tertentu yang sesuai dengan nilai konversi numerik yang ditentukan. Misalnya, Hepatitis memiliki nilai atribut 0, Anemia memiliki nilai atribut 1, dan seterusnya.

## 2. Menghapus satuan data

Beberapa atribut dataset Pasien Ibu Hamil disertai dengan satuan seperti atribut Usia Ibu Hamil terdapat satuan Tahun begitu juga dengan atribut lainnya seperti yang ditunjukkan pada Tabel 4.3 berikut.

**Tabel 4.3** Menghapus Satuan Data

No	Atribut	Nilai Atribut (Satuan)	Nilai Atribut2
1	Usia Ibu Hamil	28 Tahun	28
2	Berat Badan	70 Kg	70
3	Tinggi Badan	166 Cm	166
4	Usia Kehamilan	37 Minggu	37
5	Tekanan Darah	140/90 mmHg	140/90
6	Penyakit	Hepatitis	0

Pada Tabel 4.3 diatas adalah Proses menghapus satuan dari data. Setiap baris dalam tabel mencakup atribut tertentu beserta nilai atributnya dalam satuan awal dan nilai atribut yang sudah diubah tanpa satuan.

Setelah dilakukan transformasi untuk semua atribut yang bernilai kategorikal dan penghapusan satuan atribut, maka hasil akhir dataset yang akan digunakan ditunjukkan pada Tabel 4.4 berikut.

**Tabel 4.4** Dataset Akhir Penyakit Kehamilan

NO	Usia Ibu Hamil	Berat Badan	Tinggi Badan	Usia Kehamilan	Sistole	Diastole	Penyakit
1	28	70	166	37	140	90	0
2	35	77	169	38	75	50	1
3	23	67	153	36	70	50	1
4	23	60	165	9	121	80	2
5	24	75	160	36	122	81	3
6	20	79	171	38	150	10	4
7	23	72	164	36	120	79	5
8	29	66	152	11	110	77	6
9	22	59	160	14	120	79	7
10	25	71	167	19	80	60	8
...	...	...	...	...	...	...	...
274	30	68	152	20	148	97	4

Setelah didapatkan dataset akhir seperti pada tabel 4.4 diatas dilakukan standarisasi agar ranges nilai setiap atribut memiliki range yang sama.

**Tabel 4.5** Standarisasi Data Sebelum *Clustering*

No	Usia Ibu Hamil	Berat Badan	Tinggi Badan	Usia Kehamilan	Sistole	Diastole	Penyakit	Cluster
1	0,189229	0,009032	0,705086	1,323757	1,245796	0,994956	-1,05547	-1,17524
2	1,168002	0,833699	1,082811	1,441355	-1,68708	-1,42341	-0,94978	-0,04881

No	Usia Ibu Hamil	Berat Badan	Tinggi Badan	Usia Kehamilan	Sistole	Diastole	Penyakit	Cluster
3	-0,50989	-0,3444	-0,93172	1,206159	-1,91269	-1,42341	-0,94978	-0,04881
4	-0,50989	-1,16906	0,579178	-1,96898	0,388493	0,390365	-0,84409	1,077616
5	-0,37007	0,59808	-0,05036	1,206159	0,433615	0,450824	-0,7384	-1,17524
6	-0,92937	1,069318	1,334627	1,441355	1,697008	-3,84178	-0,63272	-0,04881
7	-0,50989	0,244651	0,45327	1,206159	0,343372	0,329905	-0,52703	-1,17524
8	0,329054	-0,46221	-1,05763	-1,73378	-0,10784	0,208987	-0,42134	1,077616
9	-0,64972	-1,28687	-0,05036	-1,38099	0,343372	0,329905	-0,31565	1,077616
10	-0,23024	0,126842	0,830994	-0,793	-1,46148	-0,81882	-0,20997	-0,04881
	...	...	...	...	...	...	...	...
274	0,468879	-0,22659	-1,05763	-0,6754	1,606766	1,41817	-0,63272	1,077616

Tabel di atas menggambarkan data ibu hamil dengan beberapa variabel, yang kemudian dikelompokkan ke dalam beberapa *cluster*. data ini menunjukkan karakteristik ibu hamil pada berbagai variabel, yang telah dinormalisasi. setiap baris mewakili satu individu ibu hamil dengan nilai masing-masing variabel. **Cluster**: Menunjukkan kelompok di mana setiap individu diklasifikasikan berdasarkan kemiripan atribut mereka. Misalnya, *Cluster* -1.17524 dan *Cluster* 1.077616.

#### 4.2.2 Pemodelan dengan *k-means*

*Clustering* Penyakit Kehamilan dengan menggunakan metode *K-Means*, pemodelannya menggunakan Bahasa Pemrograman *Python* dengan *tools Google Colab*, hasil dari pemodelan didapatkan informasi yang ditunjukkan pada Tabel 4.6 berikut :

**Tabel 4.6** Hasil Pemodelan Dengan *K-Means*

Cluster	Pusat Cluster							Jumlah data
	Usia Ibu Hamil	Berat Badan	Tinggi Badan	Usia Kehamilan	Sistole	Diastole	Penyakit	
0	0.2215 1757	0.208107 89	0.225835 52	0.158418 95	- 1.184208 54	- 1.641524 35	- 0.873 7766 5	55
1	0.2225 2087	0.154891 56	0.175472 12	0.153380 87	- 1.115546 34	- 1.655649 55	- 0.844 092	116
2	- 0.4729 8003	- 0.793958 07	- 0.677889 97	- 0.144801 84	0.095386 21	0.364246 16	0.218 7029 7	103

Tabel diatas menggambarkan data hasil pengelompokan (*clustering*) dari data ibu hamil berdasarkan beberapa variabel. *Cluster* 0 terdiri dari 55 ibu hamil dengan nilai rata-rata pada masing-masing variabel yang cenderung rendah pada tekanan darah sistolik dan diastolik, serta penyakit. *Cluster* 1 terdiri dari 116 ibu hamil dengan karakteristik yang mirip dengan *Cluster* 0 namun dengan sedikit variasi pada nilai rata-rata tiap variabel. *Cluster* 2 terdiri dari 103 ibu hamil dengan nilai rata-rata pada beberapa variabel yang berbeda signifikan dibandingkan *Cluster* 0 dan 1, misalnya nilai rata-rata tekanan darah yang lebih tinggi.

Hasil *clustering* dengan *K-Means* hasilnya ditunjukkan pada Tabel 4.5 berikut :

**Tabel 4.7** Hasil *Clustering* dengan *K-Means*

No	Usia Ibu Hamil	Berat Badan	Tinggi Badan	Usia Kehamilan	Sistole	Diastole	Penyakit	Cluster
1	28	70	166	37	140	90	0	0
2	35	77	169	38	75	50	1	1
3	23	67	153	36	70	50	1	1
4	23	60	165	9	121	80	2	2
5	24	75	160	36	122	81	3	0
6	20	79	171	38	150	10	4	1
7	23	72	164	36	120	79	5	0
8	29	66	152	11	110	77	6	2
9	22	59	160	14	120	79	7	2
10	25	71	167	19	80	60	8	1
...	...	...	...	...	...	...	...	...
274	30	68	152	20	148	97	4	2

Tabel diatas menunjukkan data ibu hamil dengan beberapa variabel klinis dan kategori klaster yang dihasilkan dari analisis *clustering*. data ini menunjukkan karakteristik fisik dan klinis ibu hamil, serta klasifikasi mereka ke dalam beberapa *cluster* berdasarkan kesamaan karakteristik. *Cluster* : Menunjukkan kelompok di mana setiap individu diklasifikasikan berdasarkan analisis *clustering*. Misalnya, *Cluster* 0, 1, dan 2.

### 4.3. Implementasi Hasil Pemodelan

Pada sub-bab ini akan menjelaskan tentang tahapan implementasi dari uji coba pemodelan K-Means untuk *Clustering*

#### 4.3.1 Implementasi Data Preprocessing

Sebelum dilakukan pemodelan ke dalam metode K-Means terlebih dahulu dilakukan beberapa pengolahan data sebagai berikut :

##### 1. *Pseudocode Import Library Python*

```
1 import numpy as np
2 import pandas as pd
3 import seaborn as sns
4 import matplotlib.pyplot as plt
5 from sklearn.preprocessing import StandardScaler
6 from sklearn.decomposition import PCA
7 from sklearn.cluster import KMeans
8 from sklearn.metrics import silhouette_score
```

##### *Pseudocode 4.1* Import Lybrary Python

Penjelasan mengenai lybrary di atas ditunjukkan pada Tabel 4.6 dibawah ini.

**Tabel 4.8** Deskripsi *Pseudocode Import Library Python*

Nama Lybray	Deskripsi
numpy	Library untuk komputasi numerik, terutama digunakan untuk array dan operasi matematika.
pandas	Library untuk manipulasi dan analisis data, terutama bekerja dengan data dalam bentuk tabel (DataFrame)
seaborn	Library untuk visualisasi data yang dibangun di atas matplotlib, menyediakan antarmuka tingkat tinggi untuk menarik grafik statistik yang menarik
matplotlib.pyplot	Sub-library dari matplotlib untuk membuat plot dan visualisasi data.
sklearn.preprocessing.StandardScaler	Digunakan untuk standarisasi fitur dengan menghapus rata-rata dan menskalakan ke varians unit.
sklearn.decomposition.PCA	Digunakan untuk melakukan analisis komponen utama, teknik pengurangan dimensi.
sklearn.cluster.KMeans	Algoritma <i>clustering</i> yang membagi data ke dalam K kluster berdasarkan jarak terdekat.
sklearn.metrics.silhouette_score	igunakan untuk mengukur seberapa mirip sebuah objek dengan kluster-nya sendiri dibandingkan dengan kluster lainnya.

##### 2. *Pseudocode* menghubungkan *Google Colab* dengan *Google Drive*

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

##### *Pseudocode 4.2* Menghubungkan *Google Colab* dengan *Google Drive*

Perintah di atas digunakan untuk menghubungkan *Google Colab* dengan *Google Drive* pengguna. Dengan cara ini, pengguna dapat dengan mudah menyimpan dan mengakses data dan model langsung dari *Google Drive* mereka, memanfaatkan kapasitas penyimpanan yang besar dan kemampuan berbagi file yang disediakan oleh *Google Drive*.

### 3. *Pseudocode Load Dataset* penyakit ibu hamil

1	dfPasien = pd.read_excel('/content/drive/MyDrive/SKRIPSI/Data
2	Pasien Ibu Hamil.xlsx')
3	dfPasien.head()

#### ***Pseudocode 4.3 Load Dataset***

Pada *Pseudocode 4.3* melakukan pembacaan dataset penyakit ibu hamil yang berupa file excel yang simpan ke dalam variable data yang selanjutnya akan digunakan dalam proses selanjutnya

### 4. *Pseudocode Menghapus Satuan Pada Kolom Umur*

1	dfPasien['Usia Ibu Hamil'] = dfPasien['Usia Ibu
2	Hamil'].str.replace(' Tahun', '').astype(int)
3	dfPasien

#### ***Pseudocode 4.4 Menghapus Satuan Pada Kolom Umur***

Perintah untuk ini membersihkan kolom 'Usia Ibu Hamil' dari string ' Tahun' dan mengkonversi nilai-nilai dalam kolom tersebut menjadi tipe data integer. Dengan konversi ini, nilai usia sekarang dapat digunakan untuk perhitungan numerik seperti analisis statistik atau grafik

### 5. *Pseudocode Menghapus Satuan Kolom Pada Berat Badan*

1	dfPasien['Berat Badan'] = dfPasien['Berat Badan'].str.replace('
2	Kg', '').astype(float)
3	dfPasien

#### ***Pseudocode 4.5 Menghapus Satuan Kolom Pada Berat Badan***

Perintah untuk ini membersihkan kolom 'Berat Badan' dari string ' Kg' dan mengkonversi nilai-nilai dalam kolom tersebut menjadi tipe data integer. Dengan

konversi ini, nilai Berat Badan sekarang dapat digunakan untuk perhitungan numerik seperti analisis statistik atau grafik.

#### 6. *Pseudocode* Menghapus Satuan Kolom Pada Tinggi Badan

```
1 dfPasien['Tinggi Badan'] = dfPasien['Tinggi Badan'].str.replace('
2 Cm', '').astype(float)
3 dfPasien
```

#### ***Pseudocode 4.6*** Menghapus Satuan Kolom Pada Tinggi Badan

Perintah untuk ini membersihkan kolom 'Tinggi Badan' dari string ' Cm' dan mengkonversi nilai-nilai dalam kolom tersebut menjadi tipe data integer. Dengan konversi ini, nilai Tinggi Badan sekarang dapat digunakan untuk perhitungan numerik seperti analisis statistik atau grafik.

#### 7. *Pseudocode* Menghapus Satuan Kolom Pada Usia Kehamilan

```
1 dfPasien['Usia Kehamilan'] = dfPasien['Usia
2 Kehamilan'].str.replace(' Minggu', '').astype(float)
3 dfPasien
```

#### ***Pseudocode 4.7*** Menghapus Satuan Kolom Pada Usia Kehamilan

Perintah untuk ini membersihkan kolom 'Usia Kehamilan' dari string ' Minggu' dan mengkonversi nilai-nilai dalam kolom tersebut menjadi tipe data integer. Dengan konversi ini, nilai Usia Kehamilan sekarang dapat digunakan untuk perhitungan numerik seperti analisis statistik atau grafik.

#### 8. *Pseudocode* Menghapus Satuan Kolom Pada Sistol

```
1 dfPasien['Sistol'] = dfPasien['Sistol'].str.replace(' mm',
2 '').astype(float)
3 dfPasien
```

#### ***Pseudocode 4.8*** Menghapus Satuan Kolom Pada Sistol

Perintah untuk ini membersihkan kolom 'Sistol' dari string ' Mm' dan mengkonversi nilai-nilai dalam kolom tersebut menjadi tipe data integer. Dengan

konversi ini, nilai Sistolik sekarang dapat digunakan untuk perhitungan numerik seperti analisis statistik atau grafik.

#### 9. *Pseudocode* Menghapus Satuan Pada Kolom Diastole

```
1 dfPasien['Diastole'] = dfPasien['Diastole'].str.replace(' Hg',  
2 '').astype(int)  
3 dfPasien
```

#### ***Pseudocode 4.9*** Menghapus Satuan Kolom Pada Diastole

Perintah untuk ini membersihkan kolom 'Diastole' dari string ' Hg' dan mengkonversi nilai-nilai dalam kolom tersebut menjadi tipe data integer. Dengan konversi ini, nilai Diastole sekarang dapat digunakan untuk perhitungan numerik seperti analisis statistik atau grafik.

#### 10. *Pseudocode* Mengkonversi Kolom Penyakit Ke Angka secara Otomatis

```
1 dfPasien['Penyakit'], unique = pd.factorize(dfPasien['Penyakit'])  
2 dfPasien
```

#### ***Pseudocode 4.10*** Mengkonversi Kolom Penyakit Ke Angka secara Otomatis

Perintah ini untuk mengubah kolom 'Penyakit' dari representasi string ke representasi numerik, di mana setiap kategori penyakit diberi kode numerik yang unik. Array `unique` berisi daftar dari kategori penyakit yang unik.

#### 11. *Pseudocode* Lihat Dataset

```
1 dfPasien.info()
```

#### ***Pseudocode 4.11*** Lihat Dataset

Perintah diatas adalah bahasa pemrograman Python yang digunakan untuk mendapatkan informasi ringkas mengenai DataFrame `dfPasien` dalam pustaka `pandas`.

#### 12. *Pseudocode* Baca Dataset Pasien Ibu Hamil Hasil Preprocessing

```
1 dfPasien =  
2 pd.read_csv('/content/drive/MyDrive/SKRIPSI/data_ibuhamil_fix.csv')  
3 dfPasien.head()
```

#### ***Pseudocode 4.12*** Baca Dataset Pasien Ibu Hamil Hasil *Preprocessing*

Perintah tersebut digunakan untuk membaca dataset dari file CSV yang disimpan di Google Drive ke dalam sebuah DataFrame pandas, lalu menampilkan beberapa baris pertama dari DataFrame tersebut.

### 13. *Pseudocode* Standarisasi Sebelum Klustering

```
1 scaler = StandardScaler()
2 data_scaled =
3 scaler.fit_transform(dfPasien.drop(columns=['No'])) # Drop kolom
4 non-numerik
5 data_scaled
```

#### ***Pseudocode 4.13*** Baca Dataset Standarisasi Sebelum Klustering

Perintah diatas digunakan untuk menormalkan data numerik dalam sebuah DataFrame pandas menggunakan metode *Standard Scaler* dari pustaka *scikit-learn*.

### 14. *Pseudocode* Buat DataFrame dari Data yang Telah Distandarisasi

```
1 # Buat DataFrame dari data yang telah distandarisasi
2 df_scaled = pd.DataFrame(data_scaled,
3 columns=dfPasien.drop(columns=['No']).columns)
4
5 # Menambahkan kembali kolom 'No' ke DataFrame yang telah distandarisasi
6 df_scaled['No'] = dfPasien['No'].values
7
8 # Menyimpan DataFrame yang telah distandarisasi ke file Excel
9 df_scaled.to_excel('/content/drive/MyDrive/SKRIPSI/data_scaled.xlsx',
10 index=False)
11
12 print("Data yang telah distandarisasi berhasil disimpan ke
13 data_scaled.xlsx")
```

#### ***Pseudocode 4.14*** DataFrame dari Data yang Telah Distandarisasi

Perintah di atas digunakan untuk membuat Data Frame dari data yang telah distandarisasi, menambahkan kembali kolom yang sebelumnya dihapus, dan menyimpan Data Frame tersebut ke dalam file Excel.

## 4.3.2 Implementasi Pemodelan

### 1. *Pseudocode* Pemodelan Melakukan klustering *K-Means*

```
1 kmeans = KMeans(n_clusters=3, random_state=42)
2 dfPasien['Cluster'] = kmeans.fit_predict(data_scaled)
3 labels = kmeans.fit_predict(data_scaled)
```

#### ***Pseudocode 4.15*** Pemodelan Melakukan klustering *K-Means*

Perintah digunakan untuk melakukan klastering menggunakan algoritma *K-Means*, mengurangi dimensi data untuk visualisasi menggunakan PCA (*Principal Component Analysis*), dan memplot hasil klastering.

## 2. *Pseudocode* Melakukan PCA untuk visualisasi

1	pca = PCA(n_components=2)
2	data_pca = pca.fit_transform(data_scaled)

### ***Pseudocode 4.16*** Melakukan PCA untuk visualisasi

Tujuan dari koding berikut adalah untuk melakukan reduksi dimensi pada dataset yang telah diskalakan menggunakan *Principal Component Analysis* (PCA).

## 3. *Pseudocode* Plot Hasil Klastering

1	# Plot hasil klastering
2	plt.figure(figsize=(10, 8))
3	plt.scatter(data_pca[:, 0], data_pca[:, 1],
4	c=dfPasien['Cluster'], cmap='viridis')
5	plt.xlabel('PCA Component 1')
6	plt.ylabel('PCA Component 2')
7	plt.title('PCA of Clustered Data')
8	plt.colorbar(label='Cluster')
9	plt.show()

### ***Pseudocode 4.17*** Plot Hasil Klastering

Perintah diatas membuat visualisasi scatter plot untuk data yang telah diklasterkan dan direduksi dimensinya menggunakan PCA (*Principal Component Analysis*). Visualisasi bertujuan untuk menunjukkan bagaimana data dikelompokkan ke dalam beberapa klaster setelah reduksi dimensi dengan PCA. memvisualisasikan distribusi data dalam dua dimensi utama yang menangkap sebagian besar variabilitas dalam data. mempermudah identifikasi pola dan struktur dalam data berdasarkan klaster-klaster yang telah terbentuk.

#### 4.4 Implementasi Evaluasi Kinerja Model

1. *Pseudocode* Menentukan Jumlah *Cluster* Yang Optimum Dengan Teknik Elbow

```
1 # Menggunakan metode Elbow untuk menemukan jumlah kluster yang
2 optimal
3 wcss = []
4 for i in range(1, 11):
5     kmeans = KMeans(n_clusters=i, random_state=42)
6     kmeans.fit(data_scaled)
7     wcss.append(kmeans.inertia_)
8
9 # Plot hasil metode Elbow
10 plt.figure(figsize=(10, 8))
11 plt.plot(range(1, 11), wcss, marker='o')
12 plt.title('Metode Elbow')
13 plt.xlabel('Jumlah Kluster')
14 plt.ylabel('WCSS')
15 plt.show()
```

#### ***Pseudocode 4.18*** Menentukan Jumlah *Cluster* Yang Optimum Dengan Teknik Elbow

Kode tersebut menggunakan metode Elbow untuk menemukan jumlah kluster yang optimal dalam data yang telah diskalakan (*data\_scaled*).

2. *Pseudocode* Menentukan Jumlah *Cluster* Yang Optimum Dengan Teknik *Davies-Bouldin Index*

```
1 from sklearn.metrics import davies_bouldin_score
2
3 # Menggunakan Davies-Bouldin Index untuk menemukan jumlah
4 kluster yang optimal
5 davies_bouldin_scores = []
6 for i in range(2, 11):
7     kmeans = KMeans(n_clusters=i, random_state=42)
8     cluster_labels = kmeans.fit_predict(data_scaled)
9     db_score = davies_bouldin_score(data_scaled, cluster_labels)
10    davies_bouldin_scores.append(db_score)
11
12 # Plot hasil Davies-Bouldin Index
13 plt.figure(figsize=(10, 8))
14 plt.plot(range(2, 11), davies_bouldin_scores, marker='o')
15 plt.title('Davies-Bouldin Index')
16 plt.xlabel('Jumlah Kluster')
17 plt.ylabel('Davies-Bouldin Score')
18 plt.show()
```

#### ***Pseudocode 4.19*** Menentukan Jumlah *Cluster* Yang Optimum Dengan Teknik *Davies-Bouldin Index*

Kode ini melakukan evaluasi kluster menggunakan *Davies-Bouldin Index* untuk berbagai jumlah kluster (2 hingga 10) dan memvisualisasikan hasilnya. Titik dengan nilai *Davies-Bouldin Index* terendah dalam plot menunjukkan jumlah kluster yang optimal untuk data tersebut.

#### 4.5 Implementasi Interpretasi Model

##### 1. Pseudocode Melakukan klastering *K-Means*

```
1 # Melakukan klastering K-Means
2 kmeans = KMeans(n_clusters=3, random_state=42)
3 dfPasien['Cluster'] = kmeans.fit_predict(data_scaled)
4 labels = kmeans.fit_predict(data_scaled)
5
6 # Melakukan PCA untuk visualisasi
7 pca = PCA(n_components=2)
8 data_pca = pca.fit_transform(data_scaled)
9
10 # Plot hasil klastering
11 plt.figure(figsize=(10, 8))
12 plt.scatter(data_pca[:, 0], data_pca[:, 1],
13 c=dfPasien['Cluster'], cmap='viridis')
14 plt.xlabel('PCA Component 1')
15 plt.ylabel('PCA Component 2')
16 plt.title('PCA of Clustered Data')
17 plt.colorbar(label='Cluster')
18 plt.show()
```

##### ***Pseudocode 4.20*** Melakukan klastering *K-Means*

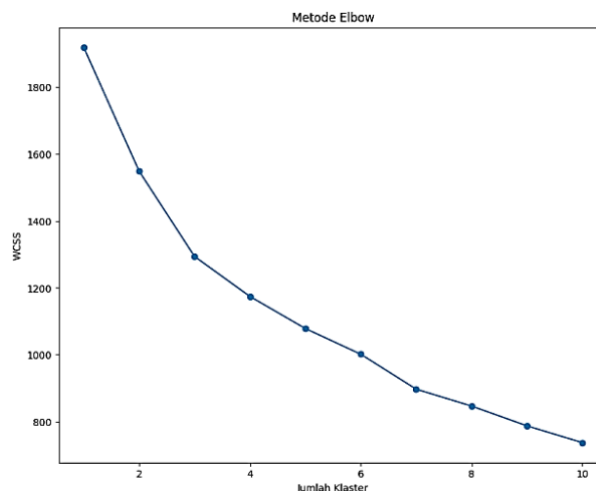
Tujuan dari kode diatas adalah **mengelompokkan Data**: Menggunakan *K-Means* untuk mengelompokkan data ke dalam kluster-kluster yang memiliki karakteristik serupa. Mereduksi dimensi data menggunakan PCA agar dapat divisualisasikan dalam dua dimensi sehingga pola-pola kluster dapat terlihat dengan jelas.

## BAB V

### HASIL DAN PEMBAHASAN

#### 5.1 Pembahasan Kinerja Model

Hasil *clustering* penyakit kehamilan dengan metode *K-Means* dilakukan evaluasi kinerja model dengan menggunakan dua teknik yaitu teknik Elbow dan teknik DBI (*Davies-Bouldin Index*). Teknik Elbow digunakan untuk menentukan jumlah *cluster* yang optimum dalam bentuk grafik Elbow dan DBI digunakan untuk mengevaluasi atau mengukur kualitas hasil *clustering* dengan mengukur seberapa mirip *cluster* satu sama lain, dan seberapa tersebar data dalam masing-masing *cluster*. Hasil penentuan jumlah kluster yang optimum dengan teknik Elbow ditunjukkan pada Gambar 5.1 berikut :



**Gambar 5.1** Visualisasi Teknik Elbow

Dari gambar tersebut, terlihat bahwa *Within-Cluster Sum of Squares* menurun dengan tajam dari 1 hingga sekitar 3-4 kluster dan kemudian penurunannya mulai melambat. Titik "elbow" atau siku kemungkinan berada di sekitar jumlah kluster 3 atau 4, yang berarti ini mungkin jumlah kluster yang optimal untuk data yang dianalisis.

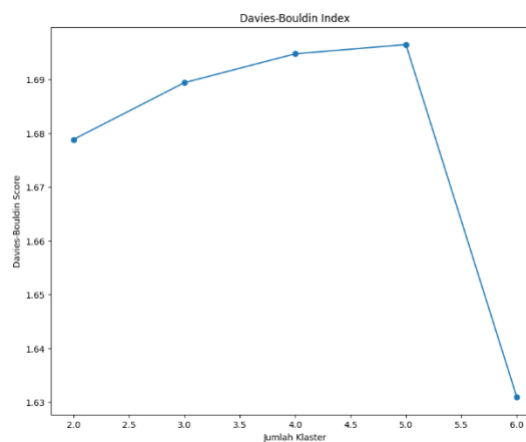
Berdasarkan Gambar 5.1 di atas, ditunjukkan bahwa jumlah *cluster* yang optimum yaitu sebanyak 3 *cluster* dengan cara melihat posisi titik garis Elbow yang penurunannya tidak signifikan lagi.

Berdasarkan jumlah *cluster* yang optimum selanjutnya dievaluasi apakah jumlah *cluster* tersebut merupakan jumlah *cluster* yang hasilnya lebih baik dengan menggunakan teknik DBI. Hasil perhitungan DBI untuk masing-masing jumlah *cluster* ditunjukkan pada Tabel 5.1 berikut :

**Tabel 5.1** Hasil perhitungan DBI untuk masing-masing jumlah *cluster*

Jumlah <i>Cluster</i>	Nilai DBI
2	1.68
3	1.69
4	1.69
5	1.70
6	1.63

Berdasarkan Tabel 5.1 di atas, ditunjukkan bahwa jumlah *cluster* 6 yang memiliki nilai DBI terkecil, namun jumlah *cluster* tersebut masih akan dievaluasi lebih lanjut dengan *visualisasi* hasil *clustering*. Hasil perhitungan DBI pada Tabel 5.1 di atas dapat juga dibuatkan dalam bentuk grafik seperti yang ditunjukkan pada Gambar 5.2 berikut :



**Gambar 5.2** Visualisasi Perhitungan DBI

Pada gambar di atas , dari 2 hingga 5 klaster, DBI meningkat atau tetap stabil, menunjukkan bahwa penambahan klaster tidak secara signifikan meningkatkan kualitas klaster dalam hal kompaksi dan separasi. Pada 6 klaster, terdapat penurunan yang signifikan dalam DBI, menunjukkan bahwa kualitas klaster meningkat secara drastis. Hal ini bisa mengindikasikan bahwa 6 klaster

adalah konfigurasi optimal, di mana kluster lebih kompak dan lebih terpisah satu sama lain dibandingkan dengan jumlah kluster lainnya.

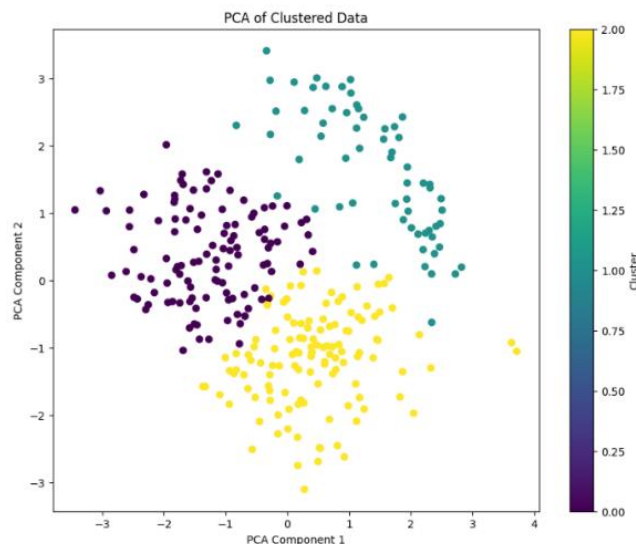
Berdasarkan analisis *Davies-Bouldin Index*, jumlah kluster optimal untuk dataset ini kemungkinan adalah 6, karena pada titik ini nilai DBI paling rendah, mengindikasikan kualitas kluster terbaik.

## 5.2 Pembahasan Hasil Pemodelan

Berdasarkan hasil evaluasi kinerja model pada *clustering* Penyakit Kehamilan dengan menggunakan metode K-Means dilakukan ujicoba dengan menggunakan jumlah *cluster* mulai dari 2 *cluster* sampai dengan enam *cluster* yang hasilnya dibuat dalam bentuk visualisasi dengan menggunakan grafik PCA, kemudian akan dievaluasi pengelompokan setiap *cluster* pada visualisasi tersebut.

### 5.2.1 Pemodelan K-Means dengan jumlah 3 *cluster* (k=3)

Hasil *clustering* dengan pemodelan menggunakan 3 *cluster* ditunjukkan pada Gambar 5.3 berikut :



**Gambar 5.3** Visualisasi Hasil *Cluster* untuk k=3

Berdasarkan Gambar 5.3 di atas, menunjukkan hasil pengelompokan (*clustering*) dari data yang divisualisasikan melalui PCA. Terdapat tiga kelompok

utama (klaster) yang dapat dikenali berdasarkan perbedaan warna, yang kemungkinan mewakili pola atau kelompok berbeda dalam data asli.

Warna pada titik-titik menggambarkan klaster yang berbeda. Skala warna di sebelah kanan (warna dari ungu ke kuning) menunjukkan pembagian klaster dalam data. Berdasarkan skala ini, ada tiga klaster berbeda yang diidentifikasi.

1. Klaster 0 ditampilkan dalam warna ungu tua. cenderung didominasi oleh kasus Hepatitis (0), menunjukkan ibu hamil dalam *cluster* ini mungkin lebih sehat atau hanya menderita Hepatitis.
2. Klaster 1 ditampilkan dalam warna kuning. menunjukkan distribusi penyakit yang lebih merata dengan frekuensi tinggi pada penyakit Hipertensi (4), Sembelit (5), dan Flu (6).
3. Klaster 2 ditampilkan dalam warna biru atau hijau. memiliki distribusi penyakit yang lebih merata dengan dua puncak utama, menunjukkan variasi penyakit yang lebih luas dan mencakup penyakit kronis atau komplikasi yang signifikan.

#### 5.2.6 Hasil Analisis Pemodelan

Setelah ditentukan jumlah *cluster*  $k=3$  yang akan digunakan dalam pemodelan *clustering* penyakit kehamilan selanjutnya untuk memudahkan dalam melakukan analisis terhadap hasil *clustering*, maka ditampilkan informasi terkait deskripsi statistik untuk masing-masing *cluster* yang ditunjukkan pada Tabel 5.2 Berikut :

**Tabel 5.2** Deskripsi Statistik Hasil *Clustering*

<i>cluster</i>	No count	Mean	std	min	25%	50%	75%	max
0	55.0	145.890909	90.931328	2.0	59.50	169.0	234.5	272.0
1	116.0	119.629310	73.852559	4.0	55.75	114.0	176.0	274.0

2	103.0	153.145631	75.039128	1.0	97.50	159.0	211.0	270.0
Cluster	Usia Ibu Hamil		diastole		penyakit			
	count	mean	75%	max	count		mean	
0	55.0	26.527273	52.5	69.0	55.0		1.781818	
1	116.0	23.077586	82.0	97.0	116.0		12.025862	
2	103.0	27.106796	82.0	95.0	103.0		11.786408	
Cluster	std		min		25%	50%	75%	max
0	1.959695		1.0		1.00	1.0	1.0	8.0
1	8.789778		0.0		5.75	9.0	20.0	32.0
2	9.864425		0.0		4.00	8.0	20.0	35.0

Tabel 5.1 di atas merupakan hasil perhitungan deksripsi statistik hasil *cluster*ng dengan k=3 menggunakan tools Python yang dikonversi ke Tabel. Adapun hasil proses perhitungannya dari Python ditunjukkan pada Gambar 5.. berikut :

	Usia Ibu Hamil		...	Diastole		Penyakit	
	count	mean	...	75%	max	count	mean
Cluster			...				
0	55.0	26.527273	...	52.5	69.0	55.0	1.781818
1	116.0	23.077586	...	82.0	97.0	116.0	12.025862
2	103.0	27.106796	...	82.0	95.0	103.0	11.786408
	std	min	25%	50%	75%	max	
Cluster							
0	1.959695	1.0	1.00	1.0	1.0	8.0	
1	8.789778	0.0	5.75	9.0	20.0	32.0	
2	9.864425	0.0	4.00	8.0	20.0	35.0	
[3 rows x 64 columns]							

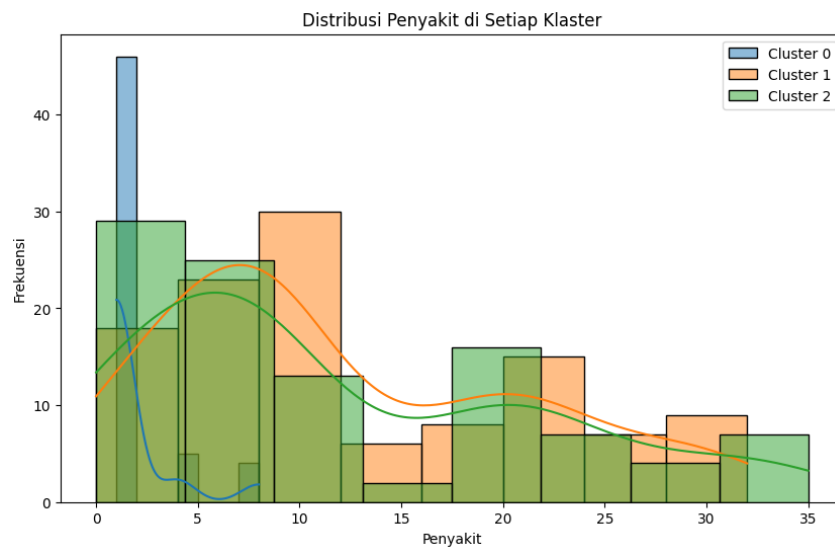
**Gambar 5.8** Hasil Proses Perhitungan deksripsi statistik

Berdasarkan hasil deskripsi statistik pada Tabel 5.2 dan Gambar 5.8 di atas dapat analisis untuk masing-masing *cluster* yang ditunjukkan pada Tabel 5.3 berikut :

**Tabel 5.3** Hasil Analisis *Clustering* k=3

Cluster	Hasil Analisis
0	Usia ibu hamil: Rata-rata 26.53 tahun dengan deviasi standar 1.96 tahun. Tekanan darah sistolik: Rata-rata 145.89 mmHg dengan deviasi standar 90.93 mmHg. Tekanan darah diastolik: Rata-rata 1.78 mmHg dengan deviasi standar 8.79 mmHg. Kondisi penyakit: Rata-rata 55% dengan deviasi standar 32%.
1	Usia ibu hamil: Rata-rata 23.08 tahun dengan deviasi standar 8.79 tahun. Tekanan darah sistolik: Rata-rata 119.63 mmHg dengan deviasi standar 73.85 mmHg. Tekanan darah diastolik: Rata-rata 12.03 mmHg dengan deviasi standar 8.79 mmHg. Kondisi penyakit: Rata-rata 82% dengan deviasi standar 20%.
2	Usia ibu hamil: Rata-rata 27.11 tahun dengan deviasi standar 9.86 tahun. Tekanan darah sistolik: Rata-rata 153.15 mmHg dengan deviasi standar 75.04 mmHg. Tekanan darah diastolik: Rata-rata 11.79 mmHg dengan deviasi standar 9.86 mmHg. Kondisi penyakit: Rata-rata 82% dengan deviasi standar 35%.

Berdasarkan hasil analisa *clustering* pada Tabel 5.3, untuk memudahkan melakukan analisis terkait distribusi jenis penyakit yang diderita oleh ibu hamil dibuatkan dalam bentuk grafik distribusi yang ditunjukkan pada gambar 5.9



**Gambar 5.9** Distribusi Penyakit Setiap Kluster

Nama penyakit ibu hamil yang diberikan nomor index 0 sampai dengan 35 pada Gambar 5.9 diatas, di uraikan pada Tabel 5.4 berikut.

**Tabel 5.4** Daftar Penyakit Ibu Hamil

No Index 1	Daftar Penyakit 1	No Index 2	Daftar Penyakit 2
0	Hepatitis	19	Tipes
1	Anemia	20	Hyperemesis Gravidarum (Mual Muntah)
2	Plasenta Previa	21	Asam Urat
3	Infeksi Saluran Kemih	22	Kolesterol
4	Hipertensi	23	Cacar Air
5	Sembelit	24	Eklampsia
6	Flu	25	Stroke
7	Keputihan	26	Asam Lambung
8	Pendarahan	27	Konstipasi
9	Hepatitis	28	Tuberkulosis (TBC)
10	Diabetes Gestasional	29	Eklampsia
11	Kehamilan Ektopik	30	Preterm Labor (Prematur yang tidak wajar)
12	Diare	31	Gemelli (Kehamilan Ganda)
13	Keguguran	32	Keputihan
14	Rubella	33	Hidroamnios
15	Campak	34	Rematik
16	Asma	35	Ginjal Kronis
17	Demam Berdarah (DBD)	36	Malaria
18	Maag	37	Sembelit

Analisa distribusi jenis penyakit ibu hamil yang ditunjukkan pada gambar 5.9 diuraikan pada tabel 5.5 berikut.

**Tabel 5.5** Analisa Distribusi Jenis Penyakit

Cluster	Analisa
0 (Warna Biru)	Hepatitis (0) : <i>Cluster</i> 0 memiliki frekuensi yang sangat tinggi untuk Hepatitis, lebih dari 40 kasus. Hepatitis (0) : <i>Cluster</i> 0 memiliki frekuensi yang sangat tinggi untuk Hepatitis, lebih dari 40 kasus.
1 (Warna Orange)	Distribusi lebih merata : <i>Cluster</i> ini menunjukkan distribusi yang lebih merata dengan puncak kasus sekitar penyakit 5 hingga 10. Hipertensi : Terdapat puncak signifikan yang mungkin menunjukkan banyak kasus Hipertensi. Sembelit (5) dan flu (6) Mungkin termasuk dalam penyakit dengan frekuensi tinggi dalam <i>cluster</i> ini. Rentang penyakit lebih luas (5-10) : <i>Cluster</i> ini mungkin memiliki banyak kasus Anemia, Plasenta Previa, Infeksi Saluran Kemih, dan penyakit lainnya dalam rentang ini.
2 (Warna Hijau)	Distribusi ganda : Terdapat dua puncak utama pada penyakit sekitar 0-5 dan 20-25. Anemia (1) , Plasenta Previa (2), infeksi saluran kemih (3) : Penyakit-penyakit ini mungkin menjadi puncak awal. Hyperemesis gravidarum (mual muntah (20), asam urat (21), kolesterol (22) : Penyakit-penyakit ini mungkin menjadi puncak kedua.

Berdasarkan hasil analisa jenis penyakit pada Tabel 5.5 dapat disimpulkan sebagai berikut.

1. Cluster 0, cenderung didominasi oleh kasus Hepatitis (0), menunjukkan ibu hamil dalam *cluster* ini mungkin lebih sehat atau hanya menderita Hepatitis.

2. Cluster 1, menunjukkan distribusi penyakit yang lebih merata dengan frekuensi tinggi pada penyakit Hipertensi (4), Sembelit (5), dan Flu (6).
3. Cluster 2, Penyakit yang mungkin mendominasi puncak awal adalah Anemia, Plasenta Previa, dan Infeksi Saluran Kemih. Penyakit seperti Hyperemesis gravidarum, asam urat, dan kolesterol cenderung muncul pada puncak kedua.

Hasil *clustering* penyakit kehamilan menggunakan metode K-Means dengan jumlah *cluster*  $k=3$ , secara detail ditunjukkan pada Tabel 5.6 berikut :

**Tabel 5.6** Detail Hasil *Clustering*

No	Usia Ibu Hamil	Berat Badan	Tinggi Badan	Usia Kehamilan	Sistole	Diastole	Penyakit	Cluster
1	28	70	166	37	140	90	0	2
2	35	77	169	38	75	50	1	0
3	23	67	153	36	70	50	1	0
4	23	60	165	9	121	80	2	1
5	24	75	160	36	122	81	3	2
6	20	79	171	38	150	10	4	0
7	23	72	164	36	120	79	5	2
8	29	66	152	11	110	77	6	1
9	22	59	160	14	120	79	7	1
10	25	71	167	19	80	60	8	0
...	...	...	...	...	...	...	...	...
274	30	68	152	20	148	97	4	1

## **BAB VI**

### **KESIMPULAN**

#### **6.1 Kesimpulan**

Berdasarkan hasil dari penelitian yang sudah di uraikan di atas tentang *clustering* penyakit kehamilan dengan metode *K-Means*, maka dapat disimpulkan sebagai berikut:

1. Hasil penerapan metode K-Means dalam *Clustering* Penyakit Kehamilan didapatkan 3 *cluster* yang optimum dengan nilai DBI sebesar 1,65.
2. Hasil analisa *clustering* untuk masing-masing *cluster* yaitu, *cluster* 1 cenderung didominasi oleh kasus Hepatitis, menunjukkan ibu hamil dalam *cluster* ini mungkin lebih sehat atau hanya menderita Hepatitis. *Cluster* 2 menunjukkan distribusi penyakit yang lebih merata dengan frekuensi tinggi pada penyakit Hipertensi, Sembelit, dan Flu. dan *Cluster* 3 Penyakit yang mungkin mendominasi puncak awal adalah Anemia, Plasenta Previa, dan Infeksi Saluran Kemih. Penyakit seperti Hyperemesis gravidarum, asam urat, dan kolesterol cenderung muncul pada puncak kedua.

#### **6.2 Saran**

Berdasarkan kesimpulan di atas, saran kepada peneliti selanjutnya, yaitu:

1. Penelitian berikutnya dapat menggunakan metode seleksi atribut yang penting atau menggunakan teknik optimisasi mencari data outlier agar hasil *clustering* bisa dibandingkan hasilnya
2. Dapat menggunakan metode *clustering* yang lain agar bisa dibandingkan hasilnya mana yang lebih baik.

## DAFTAR PUSTAKA

- [1] M. K. . Aryunita, SST,.M.Kes, Nurkholidah. SST, *Asuhan Kehamilan*. PT Inovasi Pratama Internasional, 2021.
- [2] D. Ayi Nurhidayah, M.Keb., Lilis Zuniawati Setianingsih, S.S.T, Bd, M.H., Kusuma Estu Werdani, S.KM., M.Kes., *Manajemen Puskesmas*. Kaizen Media Publishing, 2023.
- [3] P. Yudi Agusta, “K-Means – Penerapan, Permasalahan dan Metode Terkait,” *J. Sist. dan Inform.*, vol. 3, no. Pebruari, pp. 47–60, 2007.
- [4] A. Rezqiwati Ishak, “Clustering Tingkat Pemahaman Dasar Mahasiswa Pada Pra-Perkuliahan Probabilitas Statistika Dengan,” *Jambura J. Electr. Electron. Eng.*, vol. 4, pp. 65–69, 2022.
- [5] D. D. Muhamad Rivalda, Erti Maulina Hidayat, Muhamad Azhar Gunawan, “Jurnal larik,” *J. Larik, Ldng. Artik. Ilmu Komput.*, vol. 3, no. 1, 2023.
- [6] S. M. Faisal Dikarya, “PENERAPAN ALGORITMA K-MEANS CLUSTERING UNTUK,” *J. Inform.*, vol. 22, no. 02, 2022.
- [7] M. H. Adiya and Y. Desnelita, “Jurnal Nasional Teknologi dan Sistem Informasi Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan Pada RSUD Pekanbaru,” vol. 01, pp. 17–24, 2019.
- [8] R. Agusli, M. Iqbal, and F. Saputra, “Sistem Pakar Diagnosa Penyakit Pada Ibu Hamil Dengan Metode Certainty Faktor Berbasis Web,” *Acad. J. Comput. Sci. Res.*, vol. 2, no. 1, 2020, doi: 10.38101/ajcsr.v2i1.264.
- [9] Yuli Mardi, “Data Mining : Klasifikasi Menggunakan Algoritma C4 . 5 Data mining merupakan bagian dari tahapan proses Knowledge Discovery in Database ( KDD ) . Jurnal Edik Informatika,” *J. Edik Inform.*, vol. 2, no. 2, pp. 213–219, 2019.
- [10] A. Sulistiyawati and E. Supriyanto, “Implementasi Algoritma K-means Clustering dalam Penentuan Siswa Kelas Unggulan,” vol. 15, no. 2, pp. 25–36, 2020.
- [11] N. A. Maori and E. Evanita, “Metode Elbow dalam Optimasi Jumlah Cluster pada K-Means Clustering,” *Simetris J. Tek. Mesin, Elektro dan Ilmu Komput.*, vol. 14, no. 2, pp. 277–288, 2023, doi: 10.24176/simet.v14i2.9630.
- [12] Muhammad Romzi and B. Kurniawan, “Pembelajaran Pemrograman Python Dengan Pendekatan Logika Algoritma,” *JTIM J. Tek. Inform. Mahakarya*, vol. 03, no. 2, pp. 37–44, 2020.

# **LAMPIRAN**

## DATA YANG DIKUMPULKAN

No	Usia Ibu Hamil	Berat Badan	Tinggi Badan	Usia Kehamilan	Tekanan Darah		Penyakit
					Sistole	Diastole	
1	28 Tahun	70 Kg	166 Cm	37 Minggu	140 mm	90 Hg	Hepatitis
2	35 Tahun	77 Kg	169 Cm	38 Minggu	75 mm	50 Hg	Anemia
3	23 Tahun	67 Kg	153 Cm	36 Minggu	70 mm	50 Hg	Anemia
4	23 Tahun	60 Kg	165 Cm	9 Minggu	121 mm	80 Hg	Plasenta Previa
5	24 Tahun	75 Kg	160 Cm	36 Minggu	122 mm	81 Hg	Infeksi Saluran Kemih
6	20 Tahun	79 Kg	171 Cm	38 Minggu	150 mm	10 Hg	Hipertensi
7	23 Tahun	72 Kg	164 Cm	36 Minggu	120 mm	79 Hg	Sembelit
8	29 Tahun	66 Kg	152 Cm	11 Minggu	110 mm	77 Hg	Flu
9	22 Tahun	59 Kg	160 Cm	14 Minggu	120 mm	79 Hg	Keputihan
10	25 Tahun	71 Kg	167 Cm	19 Minggu	80 mm	60 Hg	Pendarahan
11	30 Tahun	78 Kg	166 Cm	36 Minggu	140 mm	90 Hg	Hepatitis
12	19 Tahun	53 Kg	147 Cm	34 Minggu	160 mm	10 Hg	Hipertensi
13	25 Tahun	60 Kg	154 Cm	36 Minggu	100 mm	80 Hg	Diabetes Gestasional
14	31 Tahun	70 Kg	160 Cm	34 Minggu	84 mm	69 Hg	Anemia
15	28 Tahun	78 Kg	150 Cm	26 Minggu	130 mm	80 Hg	Kehamilan Ektopik
16	25 Tahun	69 Kg	149 Cm	34 Minggu	120 mm	90 Hg	Diabetes Gestasional
17	25 Tahun	71 Kg	154 Cm	34 Minggu	110 mm	80 Hg	Diare
18	26 Tahun	63 Kg	148 Cm	14 Minggu	80 mm	50 Hg	Anemia
...	...	...	...	...	...	...	...
274	30 Tahun	68 Kg	152 Cm	20 Minggu	148 mm	97 Hg	Hipertensi

## KODE PROGRAM

```
# Import libraries

import numpy as np

import pandas as pd

import seaborn as sns

import matplotlib.pyplot as plt

from sklearn.preprocessing import StandardScaler

from sklearn.decomposition import PCA

from sklearn.cluster import KMeans

from sklearn.metrics import silhouette_score

from google.colab import drive

drive.mount('/content/drive')

#Baca Dataset penyakit ibu hamil

dfPasien = pd.read_excel('/content/drive/MyDrive/SKRIPSI/DATA IBU HAMIL.xlsx')

dfPasien.head()

# Menghapus satuan pada kolom Umur

dfPasien['Usia Ibu Hamil'] = dfPasien['Usia Ibu Hamil'].str.replace(' Tahun',
 '').astype(int)

dfPasien

# Menghapus satuan pada kolom Berat_Badan

dfPasien['Berat Badan'] = dfPasien['Berat Badan'].str.replace(' Kg',
 '').astype(float)

dfPasien

# Menghapus satuan pada kolom Tinggi_Badan

dfPasien['Tinggi Badan'] = dfPasien['Tinggi Badan'].str.replace(' Cm',
 '').astype(float)

dfPasien

# Menghapus satuan pada kolom Usia Kehamilan

dfPasien['Usia Kehamilan'] = dfPasien['Usia Kehamilan'].str.replace(' Minggu',
 '').astype(float)

dfPasien

# Menghapus satuan pada kolom Sistol

dfPasien['Sistol'] = dfPasien['Sistol'].str.replace(' mm', ' ').astype(float)

dfPasien

# Menghapus satuan pada kolom Diastole
```

```

dfPasien['Diastole'] = dfPasien['Diastole'].str.replace(' Hg', '').astype(int)

dfPasien

# Mengonversi kolom Penyakit ke angka secara otomatis
dfPasien['Penyakit'], unique = pd.factorize(dfPasien['Penyakit'])

dfPasien

dfPasien.info()

output_file_path = '/content/drive/MyDrive/SKRIPSI/data_ibuhamil_fix.xlsx'
dfPasien.to_excel(output_file_path, index=False)

output_file_path = '/content/drive/MyDrive/SKRIPSI/data_ibuhamil_fix.csv'
dfPasien.to_csv(output_file_path, index=False)

#Baca Dataset Pasien Ibu Hamil hasil preprocessing

dfPasien = pd.read_csv('/content/drive/MyDrive/SKRIPSI/data_ibuhamil_fix.csv')

dfPasien.head()

#Menampilkan Deskripsi statistik

dfPasien.describe()

#Standarisasi data sebelum klastering

scaler = StandardScaler()

data_scaled = scaler.fit_transform(dfPasien.drop(columns=['No'])) # Drop kolom
non-numerik

data_scaled

# Buat DataFrame dari data yang telah distandarisasi

df_scaled = pd.DataFrame(data_scaled,
columns=dfPasien.drop(columns=['No']).columns)

# Menambahkan kembali kolom 'No' ke DataFrame yang telah distandarisasi

df_scaled['No'] = dfPasien['No'].values

# Menyimpan DataFrame yang telah distandarisasi ke file Excel

df_scaled.to_excel('/content/drive/MyDrive/SKRIPSI/data_scaled.xlsx', index=False)

print("Data yang telah distandarisasi berhasil disimpan ke data_scaled.xlsx")

# Menggunakan metode Elbow untuk menemukan jumlah klaster yang optimal

wcss = []

for i in range(1, 11):

    kmeans = KMeans(n_clusters=i, random_state=42)

    kmeans.fit(data_scaled)

    wcss.append(kmeans.inertia_)

# Plot hasil metode Elbow

```

```

plt.figure(figsize=(10, 8))

plt.plot(range(1, 11), wcss, marker='o')

plt.title('Metode Elbow')

plt.xlabel('Jumlah Klaster')

plt.ylabel('WCSS')

plt.show()

# Melakukan klastering K-Means

kmeans = KMeans(n_clusters=2, random_state=42)

dfPasien['Cluster'] = kmeans.fit_predict(data_scaled)

labels = kmeans.fit_predict(data_scaled)

# Melakukan PCA untuk visualisasi

pca = PCA(n_components=2)

data_pca = pca.fit_transform(data_scaled)

# Plot hasil klastering

plt.figure(figsize=(10, 8))

plt.scatter(data_pca[:, 0], data_pca[:, 1], c=dfPasien['Cluster'], cmap='viridis')

plt.xlabel('PCA Component 1')

plt.ylabel('PCA Component 2')

plt.title('PCA of Clustered Data')

plt.colorbar(label='Cluster')

plt.show()

# Melakukan klastering K-Means

kmeans = KMeans(n_clusters=3, random_state=42)

dfPasien['Cluster'] = kmeans.fit_predict(data_scaled)

labels = kmeans.fit_predict(data_scaled)

# Melakukan PCA untuk visualisasi

pca = PCA(n_components=2)

data_pca = pca.fit_transform(data_scaled)

# Plot hasil klastering

plt.figure(figsize=(10, 8))

plt.scatter(data_pca[:, 0], data_pca[:, 1], c=dfPasien['Cluster'], cmap='viridis')

plt.xlabel('PCA Component 1')

plt.ylabel('PCA Component 2')

plt.title('PCA of Clustered Data')

```

```

plt.colorbar(label='Cluster')

plt.show()

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

#Cetak Pusat Cluster
print(kmeans.cluster_centers_)

# Menghitung Silhouette Score
silhouette_avg = silhouette_score(data_scaled, dfPasien['Cluster'])
print(f'Silhouette Score: {silhouette_avg}')

# Menyimpan data yang telah dibersihkan dan diklaster ke file Excel baru
output_file_path = '/content/drive/MyDrive/SKRIPSI/data_ibuhamil_clustered.xlsx'

dfPasien.to_excel(output_file_path, index=False)

print(f"\nData yang telah dibersihkan dan diklaster disimpan ke file:
{output_file_path}")

print("\nMapping kategori ke angka:")

mapping_dict = dict(enumerate(unique))

for key, value in mapping_dict.items():
    print(f"{key}: {value}")

# Calculate descriptive statistics for each cluster
cluster_summary = dfPasien.groupby('Cluster').agg({
    'Usia Ibu Hamil': ['mean', 'std', 'min', 'max'],
    'Tinggi Badan': ['mean', 'std', 'min', 'max'],
    'Berat Badan': ['mean', 'std', 'min', 'max'],
    'Usia Kehamilan': ['mean', 'std', 'min', 'max'],
    'Sistole': ['mean', 'std', 'min', 'max'],
    'Diastole': ['mean', 'std', 'min', 'max'],
    'Penyakit': ['mean', 'std']
}).reset_index()

# Print the summary

```



KEMENTERIAN PENDIDIKAN, KEBUDAYAAN, RISET, DAN TEKNOLOGI  
UNIVERSITAS ICHSAN GORONTALO  
LEMBAGA PENELITIAN

Kampus Unisan Gorontalo Lt.3 - Jln. Achmad Nadjamuddin No. 17 Kota Gorontalo  
Telp: (0435) 8724466, 829975 E-Mail: [lembagapenelitian@unisan.ac.id](mailto:lembagapenelitian@unisan.ac.id)

Nomor : 5101/PIP/LEMLIT-UNISAN/GTO/IV/2024

Lampiran : -

Hal : Permohonan Izin Penelitian

Kepada Yth,

Kepala Puskesmas Boliyohuto Kabupaten Gorontalo

di,-

Tempat

Yang bertanda tangan di bawah ini :

Nama : Dr. Rahmisyari, ST.,SE.,MM  
NIDN : 0929117202  
Jabatan : Ketua Lembaga Penelitian

Meminta kesediannya untuk memberikan izin pengambilan data dalam rangka penyusunan **Proposal / Skripsi**, kepada :

Nama Mahasiswa : Nazly Ibrahim  
NIM : T3122223  
Fakultas : Fakultas Ilmu Komputer  
Program Studi : Teknik Informatika  
Lokasi Penelitian : PUSKESMAS BOLIYOHUTO KABUPATEN GORONTALO  
Judul Penelitian : SISTEM PAKAR DIAGNOSA PENYAKIT MASA KEHAMILAN MENGGUNAKAN METODE NAIVE BAYES

Atas kebijakan dan kerja samanya diucapkan banyak terima kasih.



Gorontalo, 23 April 2024

Ketua,

Dr. Rahmisyari, ST.,SE.,MM  
NIDN 0929117202

+



PEMERINTAH KABUPATEN GORONTALO  
DINAS KESEHATAN  
PUSKESMAS BOLIYOHUTO

Jln. Raja Tolangohula Desa Sidomulyo Kode pos : 96235



SURAT PERNYATAAN

Nomor : 440 / Plen - Bol / 464 / VI / 2024  
Perihal : Balasan permohonan Izin Penelitian

Kepada Yth,  
Ketua Lembaga Penelitian  
Di  
Tempat

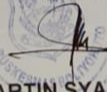
Dengan Hormat,

Sehubungan dengan surat Permohonan Izin Penelitian Sdra. Nazly Ibrahim dalam rangka penyusunan Proposal/Skripsi, maka dengan kami sampaikan bahwa kami mengizinkan yang bersangkutan untuk melakukan kegiatannya tersebut di Puskesmas Boliyohuto.

Demikian surat ini kami buat, Kami Ucapkan terimakasih.

Dikeluarkan di : Boliyohuto  
Tanggal : 11 Juni 2024

KEPALA PUSKESMAS BOLIYOHUTO  
KABUPATEN GORONTALO

  
SUDARTIN SYAMSUDIN, SKM  
NIP. 19641006 198803 2 010

## PAPER NAME

SKRIPSI\_T3122223\_NAZLY IBRAHIM.pdf

## AUTHOR

NAZLY IBRAHIM nazlyibrahim1116@gmail.com

## WORD COUNT

7855 Words

## CHARACTER COUNT

45503 Characters

## PAGE COUNT

44 Pages

## FILE SIZE

2.1MB

## SUBMISSION DATE

Jun 12, 2024 12:56 PM GMT+8

## REPORT DATE

Jun 12, 2024 12:56 PM GMT+8

**18% Overall Similarity**

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

- 18% Internet database
- 10% Publications database
- Crossref database
- Crossref Posted Content database
- 0% Submitted Works database

**Excluded from Similarity Report**

- Bibliographic material
- Quoted material
- Cited material
- Small Matches (Less than 30 words)

● **18% Overall Similarity**

Top sources found in the following databases:

- 18% Internet database
- 10% Publications database
- Crossref database
- Crossref Posted Content database
- 0% 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	4%
2	<b>ejurnal.teknokrat.ac.id</b> Internet	2%
3	<b>researchgate.net</b> Internet	2%
4	<b>teknsi.fti.unand.ac.id</b> Internet	2%
5	<b>jurnal.umk.ac.id</b> Internet	1%
6	<b>ijins.umsida.ac.id</b> Internet	<1%
7	<b>ibtpi.pelitaIndonesia.ac.id</b> Internet	<1%
8	<b>pemrogramanmatlab.com</b> Internet	<1%

9	<a href="https://kompasiana.com">kompasiana.com</a> Internet	<1%
10	<a href="https://eprints.bsi.ac.id">eprints.bsi.ac.id</a> Internet	<1%
11	<a href="https://es.scribd.com">es.scribd.com</a> Internet	<1%
12	<a href="https://pustaka.ut.ac.id">pustaka.ut.ac.id</a> Internet	<1%
13	<a href="https://repositori.buddhidharma.ac.id">repositori.buddhidharma.ac.id</a> Internet	<1%
14	Yuhendri Putra, Ega Sharifa Siregar. "PENGARUH TERAPI RENDAM AIR..." Crossref	<1%
15	<a href="https://etheses.uin-malang.ac.id">etheses.uin-malang.ac.id</a> Internet	<1%
16	<a href="https://cdn.gob.pe">cdn.gob.pe</a> Internet	<1%
17	<a href="https://loftmego.com">loftmego.com</a> Internet	<1%



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 : 040/Perpustakaan-Fikom/VI/2024

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

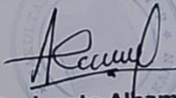
Nama Anggota : Nazly Ibrahim  
No. Induk : T3122223  
No. Anggota : M202460

Terhitung mulai hari, tanggal : Jumat, 21 Juni 2024, dinyatakan telah bebas pinjam buku dan koleksi perpustakaan lainnya.

Demikian keterangan ini di buat untuk di gunakan sebagaimana mestinya.



Gorontalo, 21 Juni 2024  
Mengetahui,  
Kepala Perpustakaan

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



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. 162/FIKOM-UIG/R/VI/2024

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 : Nazly Ibrahim  
NIM : T3122223  
Program Studi : Teknik Informatika (S1)  
Fakultas : Fakultas Ilmu Komputer  
Judul Skripsi : Clustering Penyakit Kehamilan Menggunakan Metode K-Means Pada Puskesmas Boliyohutuo

Sesuai hasil pengecekan tingkat kemiripan skripsi melalui aplikasi Turnitin untuk judul skripsi di atas diperoleh hasil *Similarity* sebesar **18%**, 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, 19 Juni 2024  
Tim Verifikasi,

  
Zulfrianto Y. Lamasigi, M.Kom  
NIDN. 0914089101

Tersampir :  
Hasil Pengecekan Turnitin

## BIODATA PENULIS



**Penulis**, lahir di Gorontalo tanggal 14 Mei 2001, anak pertama dari dua bersaudara dari ayah Anis Ibrahim dan Ibu Isna Nur. Berikut ini addalah biodata tentang penulis :

Nama : Nazly Ibrahim  
Alamat : Desa Sukamaju, Kec. Mootilango Kab. Gorontalo  
Agama : Islam  
E-mail : [nazlyibrahim1116@gmail.com](mailto:nazlyibrahim1116@gmail.com)

Riwayat pendidikan formal yang pernah ditempuh :

- SDN 11 Mootilango, lulus tahun 2013
- SMPN 01 Mootilango, lulus tahun 2016
- SMKN 01 Mootilango, lulus tahun 2019
- Politeknik Gorontalo, lulus tahun 2022 dengan gelar Ahli Madya Komputer

Pada bulan Juni 2024 mengikuti sidang skripsi sebagai salah satu syarat untuk mendapatkan gelar Sarjana di Universitas Ichsan Gorontalo