



**KLASIFIKASI KELUHAN PASIEN TERHADAP DATA REKAM MEDIS
PASIEN DENGAN MENGGUNAKAN METODE K MEANS**

TESIS

Diajukan untuk Memenuhi Salah Satu Syarat
Memperoleh Gelar Magister Komputer

M. AGUNG VAFKY IDEAL
202321010

**PROGRAM MAGISTER (S2)
PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS PUTRA INDONESIA "YPTK" PADANG**

SEPTEMBER 2022

**PROGRAM MAGISTER (S2)
PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER**

Tanda Persetujuan Diberikan Kepada

**NAMA : M. AGUNG VAFKY IDEAL
NOBP : 202321010**

**KLASIFIKASI KELUHAN PASIEN TERHADAP DATA REKAM MEDIS
PASIEN DENGAN MENGGUNAKAN METODE K MEANS**

**Disetujui Untuk Diajukan Pada Ujian Akhir, Sidang Tertutup
Program Magister (S2)
Program Studi Teknik Informatika
Fakultas Ilmu Komputer
Universitas Putra Indonesia “YPTK” Padang**

MENYETUJUI

PEMBIMBING I

PEMBIMBING II

Prof. Dr. Sarjon Defit, S.Kom., M.Sc

NIDN : 1007087002

Dr. Yuhandri, S.Kom., M.Kom

NIDN : 1015057301

Telah dinyatakan lulus ujian tesis pada sidang tertutup. Program Magister (S2) Program Studi Teknik Informatika Fakultas Ilmu Komputer Universitas Putra Indonesia “YPTK” Padang pada Bulan September 2022 dengan Baik.

Padang, September 2022

Tim Penguji

Penguji I :

NIDN :

Penguji II :

NIDN :

Mengesahkan
Dekan Fakultas Ilmu Komputer
Universitas Putra Indonesia “YPTK” Padang

Dr. Yuhandri, S.Kom., M.Kom

NIDN : 1015057301

“Dengan ini saya menyatakan bahwa laporan penelitian ini adalah hasil pemikiran dan karya tulis saya sendiri, adapun bagian tertentu dalam laporan ini yang saya peroleh dari kutipan karya tulis narasumber, telah saya tuliskan sumbernya dengan jelas sesuai dengan kaedah penulisan karya ilmiah”

Tanda Tangan :

Nama Penulis : M. Agung Vafky Ideal

Tanggal : Oktober 2022

KATA PENGANTAR

Alhamdulillah, berkat rahmat dan nikmat Allah Yang Maha Mengetahui maka penulisan tesis yang berjudul **KLASIFIKASI KELUHAN PASIEN TERHADAP DATA REKAM MEDIS PASIEN DENGAN MENGGUNAKAN METODE K MEANS** dapat diselesaikan dengan baik. Shalawat dan salam kita sampaikan kepada teladan manusia, Rasulullah Shalallahu Alaihi Wassallam, karena Beliau adalah yang telah membawa kita dari zaman kejahiliah ke zaman ummat yang beradab dan berilmu pengerahuan.

Selanjutnya, penulis ucapkan terima kasih semua pihak-pihak yang telah membantu sehingga penelitian ini dapat diselesaikan. Ucapan terima kasih khususnya penulis ucapkan kepada :

1. Ibu **Dr. Hj. Zerni Melmusni, M.M., Ak.** selaku Ketua Yayasan Perguruan Tinggi Komputer Padang.
2. Bapak **Prof. Dr. Sarjon Defit, S.Kom., M.Sc.** selaku Rektor Universitas Putra Indonesia “YPTK” Padang.
3. Bapak **Dr. Yuhandri, S.Kom., M.Kom.** selaku Dekan Fakultas Ilmu Komputer Universitas Putra Indonesia “YPTK” Padang dan juga selaku Pembimbing II penulis yang telah memberikan motivasi, masukan dan saran sehingga tesis ini dapat diselesaikan dengan baik.
4. Bapak **Ir. Gunadi Widi Nurcahyo, M.Sc., PhD.**, selaku Ketua Program Studi Magister Teknik Informatika Fakultas Ilmu Komputer Universitas Putra Indonesia “YPTK” Padang.
5. Seluruh Dosen dan Staf Administrasi Program Studi Magister Teknik Informatika Fakultas Ilmu Komputer Universitas Putra Indonesia “YPTK” Padang yang telah membantu baik secara langsung maupun tidak langsung demi kelancaran penyusunan tesis ini.

6. Kepada seluruh pihak yang tidak dapat disebutkan satu persatu, yang telah membantu penulis dalam penyelesaian tesis ini.

Penulis menyadari bahwa dalam penulisan tesis ini masih terdapat banyak kekurangan dan kelemahannya, untuk itu penulis sangat mengharapkan masukan berupa kritik dan saran yang bersifat membangun. Penulis berharap semoga tesis ini dapat bermanfaat bagi pembaca dan semua pihak yang berkepentingan dengan tesis ini. Sebagai penutup, penulis mengajak agar kita semua selalu berserah diri kepada Allah dan meminta pertolongan hanya pada-Nya. Semoga apa yang kita usahakan di dunia bernilai ibadah hendaknya.

Padang, September 2022

Penulis,

M. Agung Vafky Ideal

ABSTRAK

Keluhan pasien merupakan efek dari tubuh yang sedang terjangkit oleh suatu penyakit. Keluhan yang dialami pasien dapat disebabkan oleh beberapa faktor, Faktor penyebab keluhan penyakit pasien seperti faktor internal dan faktor eksternal. Setiap harinya puskesmas padusunan menerima pasien yang cukup banyak. Penyakit yang banyak dan dengan gejala yang berbeda-beda dari masyarakat menyebabkan pihak puskesmas kesulitan dalam menentukan penyuluhan atau sosialisasi yang tepat dalam suatu penyakit yang ditimbulkan. Penyuluhan yang tepat akan meminimalisir kemungkinan masyarakat terjangkit suatu penyakit. Keluhan yang diderita pasien dapat menjadi acuan dalam mengetahui penyebab dari suatu penyakit. Penelitian ini bertujuan untuk menghasilkan kelompok data keluhan pasien dan Mengetahui parameter pada setiap keluhan pasien. Data yang digunakan pada penelitian ini adalah data keluhan pasien yang berada pada data rekam medis pasien. Data rekam medis adalah berkas yang berisikan pencatatan dari biodata dan rincian pelayanan yang telah diterima pasien. Pada penelitian ini data yang digunakan adalah data keluhan pasien dalam jangka waktu 6 bulan yang terdiri dari 72 kategori keluhan pasien . Data diolah dengan menggunakan metode *k-means*. Metode k-means mengelompokkan data dengan system partisi. Hasil yang diperoleh yaitu keluhan pasien sebanyak 3 cluster dimana cluster 1 adalah keluhan yang terjadi dengan intensitas tinggi sebanyak 9 kategori keluhan, cluster 2 adalah keluhan yang terjadi dengan intensitas sedang sebanyak 15 kategori keluhan, dan cluster 3 adalah keluhan yang terjadi dengan intensitas rendah sebanyak 48 kategori keluhan . penelitian ini menghasilkan klasifikasi berdasarkan keluhan pasien sehingga pihak puskesmas dapat menemukan pola informasi yang ada pada data dan dapat membantu pihak puskesmas dalam menentukan penyuluhan atau sosialisasi yang tepat dalam suatu penyakit yang ditimbulkan.

Kata kunci : *k-means*, rekam medis, data mining, klasifikasi, Keluhan

ABSTRACT

Patient complaints are the effects of the body being infected by a disease. Complaints experienced by patients can be caused by several factors, factors that cause patient complaints such as internal factors and external factors. Every day the Padusan Community Health Center receives quite a lot of patients. Diseases that are many and with different symptoms from the community cause the health center to find it difficult to determine the right counseling or socialization in a disease that is caused. Proper counseling will minimize the possibility of people contracting a disease. Complaints suffered by patients can be a reference in knowing the cause of a disease. This study aims to produce groups of patient complaint data and determine the parameters for each patient complaint. The data used in this study is data on patient complaints that are in the patient's medical record data. Medical record data is a file containing records of biodata and details of services that have been received by patients. In this study, the data used were patient complaints data for a period of 6 months which consisted of 72 categories of patient complaints. The data is processed using the k-means method. The k-means method groups data with system partitions. The results obtained are patient complaints as many as 3 clusters where cluster 1 is complaints that occur with high intensity as many as 9 categories of complaints, cluster 2 is complaints that occur with moderate intensity as many as 15 categories of complaints, and cluster 3 is complaints that occur with low intensity as many as 48 complaint category. This study produces a classification based on patient complaints so that the puskesmas can find patterns of information in the data and can help the puskesmas in determining the right counseling or socialization in a disease caused.

Keywords: *k-means, medical records, data mining, classification, complaints*

DAFTAR ISI

BAB	JUDUL	HALAMAN
	HALAMAN JUDUL	i
	HALAMAN PERSETUJUAN	ii
	HALAMAN PENGESAHAN	iii
	HALAMAN PENGAKUAN	iv
	KATA PENGANTAR	v
	ABSTRAK	vii
	ABSTRACT	viii
	DAFTAR ISI	ix
	DAFTAR TABEL	xii
	DAFTAR GAMBAR	xiii
	DAFTAR LAMPIRAN	xv
I	PENDAHULUAN	
	1.1 Latar Belakang	1
	1.2 Perumusan Masalah	6
	1.3 Batasan Masalah	6

1.4	Tujuan Penelitian	7
1.5	Manfaat Penelitian	7
1.6	Sistematika Penulisan	8
II	LANDASAN TEORI	
2.1	Rekam Medis	10
2.2	<i>Knowledge Discovery in Database</i>	11
2.3	<i>Data Mining</i>	13
	2.3.1 Fungsi <i>Data Mining</i>	14
	2.3.2 Tujuan <i>Data Mining</i>	14
	2.3.3 Tahapan dalam Data Mining	15
2.4	<i>Clustering</i>	19
2.5	Algoritma <i>K-Means Clustering</i>	20
2.6	Penelitian Terdahulu	23
III	METODOLOGI PENELITIAN	
3.1	Pendahuluan	30
3.2	Kerangka Kerja Penelitian	31
IV	ANALIS DAN PERANCANGAN SISTEM	

4.1	Tahapan Analisa dan Perancangan	34
4.2	Data Keluhan Pasien	35
4.2.1	Melakukan Data <i>Selection</i>	36
4.2.2	Proses Data <i>Transformation</i>	40
4.3	Menganalisa Sistem	41
4.3.1	Input Data Keluhan Pasien	43
4.3.2	Tentukan Jumlah <i>Cluster</i>	44
4.3.3	Proses Iterasi Pertama	44
4.3.4	Proses Iterasi Kedua	50
4.3.5	Perancangan Sistem	55
4.4	Perancangan Sistem	61
V	IMPLEMENTASI DAN HASIL	
5.1	Implementasi	63
5.2	Pengujian <i>K-Means</i> dengan menggunakan Weka	64
5.3	Pengujian <i>K-means</i> dengan Menggunakan Web	73
VI	KESIMPULAN DAN SARAN	
6.1	Kesimpulan	78
6.2	Saran	79

DAFTAR PUSTAKA

LAMPIRAN

DAFTAR TABEL

TABEL	JUDUL	HALAMAN
2.1	Penelitian Terdahulu	24
4.1	Keluhan Pasien	36
4.2	Hasil Transformation Data Keluhan Pasien	37
4.3	Hasil Transformation Data Keluhan Pasien	40
4.4	Keluhan Pasien Perbulan	43
4.5	Centroid Awal	46
4.6	Jarak Setiap Data Keluhan pada Iterasi 1	47
4.7	Hasil Pengelompokan Iterasi 1	49
4.8	Hasil Penentuan Pusat <i>Centroid</i> Baru	51
4.9	Hasil Iterasi 2	52
4.10	Hasil Pengelompokan Iterasi 2	53
4.11	Tabel Keluhan	58
4.12	Tabel Rekap	59
4.13	Tabel Jarak	60

4.14	Tabel Tampil	60
4.15	Tabel User	60
4.16	Hasil Clustering Data Keluhan Pasien	61

DAFTAR GAMBAR

GAMBAR	JUDUL	HALAMAN
2.1	Proses <i>Knowledge Discovery in Database</i>	12
3.1	Kerangka Kerja Penelitian	33
4.1	Bagan Alir Anlisa dan Perancangan	34
4.2	Bagan Alir Anlisa Pra Proses Data Keluhan	35
4.3	Flowchart Proses pada <i>K-Means Clustering</i>	42
4.4	Halaman Iterasi K-Means	56
4.5	Halaman <i>Outpur Cluster</i>	57
4.6	Halaman Input Data Keluhan	59
5.1	Tampilan Awal Menu Home	65
5.2	Tampilan Explorer	67
5.3	Input Data	68
5.4	Tampilan <i>Cluster</i>	66
5.5	Tampilan Pilihan Metode Cluster	69
5.6	Tampilan <i>Generic Object Editor</i>	70

5.7	Proses Algoritma <i>K-means Clustering</i>	71
5.8	Simpan Hasil <i>Cluster</i>	72
5.9	Menampilkan Hasil <i>Clustering</i>	73
5.10	Halaman Login	74
5.11	Halaman Dashboard	75
5.12	Halaman Input Data	73
5.13	Halaman Proses <i>K-means</i>	76
5.14	Tampilan Hasil <i>K-Means</i>	77

BAB I

PENDAHULUAN

1.1 Latar Belakang

Dunia saat ini mengalami perkembangan teknologi yang sangat pesat. Perkembangan ini menghasilkan teknologi yang semakin canggih dan data semakin banyak. Data yang semakin banyak akan mengakibatkan pertanyaan yang semakin banyak juga, maka untuk menagani hal tersebut, dapat diterapkan sebuah teknologi yang dikenal dengan *Knowledge Discovery in Database In Database* atau data mining.

Knowledge Discovery in Database In Database adalah suatu metode atau model yang digunakan untuk memperoleh pengetahuan yang tersembunyi dari database yang ada, pengetahuan yang didapatkan dapat menjadi acuan dalam mengambil suatu keputusan (Nugraha, 2021). *Database* yang umumnya hanya dipakai untuk menyimpan suatu data namun dengan KDD kita dapat menemukan suatu pengetahuan tersembunyi didalamnya, dari pengetahuan tersebut dapat digunakan sebagai patokan atau pedoman dalam menganalisa sesuatu atau melakukan suatu tindakan. Secara garis besar KDD terdiri dari beberapa langkah yaitu *data set cleaning*, *data integration*, *data selection*, *transformation of data*, dan *data mining*.

Data mining dapat digunakan untuk menggali informasi dan nilai-nilai tersembunyi dari suatu kumpulan data, dimana informasi ini tidak dapat ditemukan apabila dilakukan secara manual (Ordila Rian, 2020). Data mining merupakan salah satu langkah yang penting pada proses *Knowledge Discovery in Database*. Tahapan ini juga

bisa disebut proses inti dari KDD. 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 database yang besar. Istilah data mining memiliki hakikat sebagai disiplin ilmu yang tujuan utamanya adalah untuk menemukan, menggali, atau menambang pengetahuan dari data atau informasi yang kita miliki. Proses menggali informasi dalam data mining melibatkan integrasi teknik dari berbagai disiplin ilmu, seperti teknologi *database* dan data *warehouse*, statistik, *machine learning*, komputasi dengan kinerja tinggi, *pattern recognition*, *neural network*, visualisasi data dan sebagainya.

Teknik *clustering* adalah salah satu Teknik data *mining* yang dapat digunakan. Teknik *clustering* bersifat tanpa arahan (*unsupervised*), dalam *data mining* terdapat dua jenis metode *clustering* yang digunakan dalam pengelompokan data, yaitu *hierarchical clustering* dan *non-hierarchical clustering*. *Clustering* bisa dikatakan sebagai identifikasi kelas objek yang memiliki kemiripan. Dengan menggunakan teknik *clustering* kita bisa lebih lanjut mengidentifikasi kepadatan dan jarak daerah dalam objek ruang dan dapat menemukan secara keseluruhan pola distribusi dan korelasi antara atribut. Pendekatan klasifikasi secara efektif juga dapat digunakan untuk membedakan kelompok atau kelas objek. Tujuan utama dari metode *clustering* adalah pengelompokan sejumlah data atau objek ke dalam *cluster* atau grup sehingga dalam setiap cluster akan berisi data yang semirip mungkin. *Clustering* melakukan pengelompokan data yang didasarkan pada kesamaan antar objek, oleh karena itu klasterisasi digolongkan sebagai metode *unsupervised learning*.

K-means adalah metode partisi yang sering digunakan dalam penyesuaian pada *machine study* dan analisa suatu pola data, algoritma ini sangat terpengaruh pada *centroid* awal, tapi tidak dapat menjamin sampai menemukan solusi akhir hal ini dikarenakan sentroid awal di cari secara acak untuk cluster yang diberikan (Manochandar, 2020). *K-means clustering* sebagai salah satu metode data *clustering* non-hirarki yang mempartisi data yang tersedia ke dalam bentuk satu atau beberapa *cluster*, sehingga dapat menghasilkan kelompok data yang memiliki karakteristik atau *cluster* yang sama dan data yang mempunyai karakteristik yang berbeda dikelompokkan ke dalam kelompok yang lain. *K-means clustering* merupakan salah satu metode cluster analysis non hirarki yang berusaha untuk mempartisi objek yang

ada kedalam satu atau lebih cluster atau kelompok objek berdasarkan karakteristiknya, sehingga objek yang mempunyai karakteristik yang sama dikelompokkan dalam satu cluster yang sama dan objek yang mempunyai karakteristik yang berbeda dikelompokkan kedalam cluster yang lain

Penelitian tentang penerapan algoritma *k-means clustering* sebelumnya telah dilakukan oleh Rian Ordila (2020), dalam mengelompokkan data rekam medis berdasarkan jenis penyakit di Poli Klinik PT.Inecda dengan menggunakan metode *k-means clustering*. Data yang dipakai yaitu data rekam medis 600 pasien, Sehingga mendapatkan klasifikasi data berdasarkan wilayah, jenis penyakit, dan umur. Penelitian yang dilakukan oleh Amjad Rehman Khan (2020), melakukan penelitian tentang segmentasi tumor otak dengan menggunakan metode *k-means*, penelitian ini dilakukan pada data MRI yang telah dianalisis, menghasilkan tingkat ke akurasian yang tinggi dalam mendiagnosa tumor. Penelitian dilakukan oleh Jaja (2021), melakukan penelitian tentang kinerja agen WFH (*work from home*) dan WFO (*work from office*) studi kasus PT. infomedia telkom dengan menggunakan metode *k-means*. Data yang digunakan yaitu 80% data agen WFH dan 20% data agen WFO, dihasilkan data yang menunjukkan perbedaan kinerja dari karyawan yang bekerja dari rumah dan di kantor. Penelitian yang dilakukan oleh Ismail Virgo, Sarjon Defit, Yuhandri Yunus (2020), melakukan penelitian tentang klasterisasi tingkat kehadiran dosen dengan menggunakan metode *k-means clustering*, data yang dipakai adalah data pertemuan yang dilakukan selama proses pembelajaran oleh dosen non PNS yang mampu matakuliah. Penelitian ini menghasilkan pengelompokan dosen yaitu 72 matakuliah jarang melakukan pertemuan (4.7650%), 69 matakuliah sedang dalam melakukan pertemuan (4.5665%), dan 1370 matakuliah rajin melakukan pertemuan (90.6684%). kesimpulannya pada tahun akademik 2017/2018 semester gasal dan genap dosen non pns rajin masuk dengan tingkat kehadiran 12-16 kali pertemuan per semester. Penelitian yang dilakukan oleh Sri Dewi, Sarjon Defit, Yuhandri Yunus (2021), melakukan penelitian tentang akurasi pemetaan kelompok belajar siswa menuju prestasi dengan menggunakan metode *k-means*, data yang diteliti yaitu nilai rapor siswa Kelas IX.C di SMP Pembangunan Laboratorium UNP. Hasil penelitian ini memperoleh 3 *Cluster* siswa yaitu siswa Sangat Berprestasi, Berprestasi dan Kurang Berprestasi. Penelitian ini bisa dijadikan sebagai pedoman bagi guru dalam pengambilan keputusan di Kelas IX.C.

Puskesmas merupakan salah satu lembaga pemerintah yang bergerak dibidang pelayanan kesehatan masyarakat pada tingkat kecamatan. Salah satu peran puskesmas yaitu dalam menopang kinerja dari Lembaga Kesehatan di atasnya yaitu rumah sakit, sebagai upaya pencegahan dan penanggulangan masalah kesehatan masyarakat. Dalam upaya peningkatan kualitas pelayanan kesehatan yang lebih baik ditingkat puskesmas pada khususnya. Dalam Peraturan Menteri Kesehatan Republik Indonesia Nomor 75 Tahun 2014 Tentang Puskesmas, puskesmas memiliki fungsi berupa penyelenggaraan Upaya Kesehatan Perorangan (UKP) tingkat pertama di wilayah kerjanya dan Upaya kesehatan masyarakat (UKM) tingkat pertama di wilayah kerjanya. Salah satu fungsi puskesmas yaitu melaksanakan komunikasi, informasi, edukasi, dan pemberdayaan masyarakat dalam bidang Kesehatan. Maka diperlukan sebuah konsep atau sistem yang baik yang akan digunakan, sehingga nantinya dapat terwujud suatu pelayanan kesehatan yang bermutu, efektif dan efisien serta dapat meningkatkan kinerja dari puskesmas itu sendiri.

Rekam medis atau bisa disebut dengan ICD (*internasional Classification Diseases*) adalah rekaman atau catatan dari pasien yang telah melakukan pengobatan di rumah sakit atau puskesmas. Bahasa medis yang biasa dilakukan oleh dokter dalam melakukan identifikasi suatu penyakit lalu memberikan tindakan atas penyakit tersebut berupa bahasa kedokteran (rekam medis) yang selanjutnya dikodekan menjadi kode-kode ICD. Kode ini adalah bahasa standar yang biasa digunakan oleh dokter pada umumnya walaupun bukan dokter spesialis, dalam membaca kode ini sesuai dengan aturan yang telah ditetapkan (Rian Ordila, 2020).

Puskesmas Padusunan merupakan tempat pelayanan Kesehatan masyarakat di Desa Padusunan kota Pariaman. Setiap harinya Puskesmas Padusunan melayani masyarakat yang berbagai macam. Puskesmas juga berperan penting dalam pencegahan penyakit pada suatu daerah, hal tersebut dapat dilakukan dengan cara penyuluhan, sosialisasi, dan bentuk pencegahan lainnya. Setiap data dari pasien juga dicatat baik itu dari keluhan, biodata diri, penanganan, penyakit yang diderita dan lain sebagainya.

Setiap harinya puskesmas padusunan menerima pasien yang cukup banyak. Penyakit yang banyak dan dengan gejala yang berbeda-beda dari masyarakat

menyebabkan pihak puskesmas kesulitan dalam menentukan penyuluhan atau sosialisasi yang tepat dalam suatu penyakit yang ditimbulkan. Penyuluhan yang tepat akan meminimalisir kemungkinan masyarakat terjangkit suatu penyakit. Keluhan yang diderita pasien dapat menjadi acuan dalam mengetahui penyebab dari suatu penyakit. Apakah penyakit tersebut disebabkan oleh faktor internal, atau faktor eksternal. Faktor internal seperti suatu penyakit yang disebabkan oleh suatu kondisi yang abnormal pada pasien itu sendiri. Faktor eksternal seperti penyakit yang disebabkan oleh lingkungan tinggal pasien, ditularkan dari orang lain dan lain sebagainya.

Salah satu solusi yang penulis lakukan dalam mengatasi permasalahan di atas adalah dengan menggunakan Teknik *data maining*. *Data maining* yang digunakan yaitu *clustering* dengan menggunakan algoritma *k-means clustering*. Hasil dari penerapan algoritma *clustering* adalah berupa cara menentukan kelompok gejala penyakit yang terjadi pada masyarakat sehingga dari clustering tersebut pihak puskesmas dapat menentukan penyuluhan yang efektif dan efisien agar dapat mencegah penyebab dari suatu gejala penyakit. Berdasarkan latar belakang yang telah penulis jelaskan, penulis melakukan penelitian yang berjudul “**klasifikasi keluhan pasien terhadap data rekam medis pasien dengan menggunakan metode *k-means*”**”.

1.2 Perumusan Masalah

Rumusan masalah dirumuskan berdasarkan penjabaran dari latar belakang penelitian sebagai acuan peneliti dalam mencari jawaban pertanyaan yang dikemukakan. Rumusan masalah yang dirumuskan dapat diteliti secara ilmiah. Dalam penelitian ini ditetapkan beberapa rumusan masalah yang dapat dirumuskan sebagai berikut :

1. Bagaimana menghasilkan parameter gejala penyakit masyarakat pada tiap bulannya?
2. Bagaimana menghasilkan kelompok atau *cluster* keluhan penyakit masyarakat dengan menggunakan algoritma *K-Mean Clustering*?
3. Bagaimana analisa dan pertimbangan terhadap hasil *cluster* gejala penyakit masyarakat yang dilakukan dengan menggunakan algoritma *K-Mean Clustering*?

1.3 Batasan Masalah

Batasan masalah berguna dalam membantu mengidentifikasi masalah, membatasi ruang lingkup. Batasan masalah memberikan gambaran terkait hal yang akan diteliti supaya penelitian lebih terarah. Maka berdasarkan perumusan masalah di atas diberikan batasan ruang lingkup penelitian sebagai berikut:

1. Teknik *data mining* yang digunakan adalah *Clustering* menggunakan algoritma *K-means Clustering*.
2. Data yang digunakan dalam penelitian ini adalah data rekam medik masyarakat padusunan selama 1 tahun yaitu tahun.
3. Penganalisaan data gejala penyakit masyarakat menggunakan.

1.4 Tujuan Penelitian

Hasil akhir dalam suatu penelitian adalah tujuan dari sebuah *research*. Diperlukan suatu perencanaan yang baik dan pelaksanaan penelitian dengan matang agar tujuan penelitian dapat dicapai dengan baik. Dalam penelitian ini ada beberapa tujuan yang hendak dicapai dan diringkas menjadi beberapa poin, yaitu:

1. Mengetahui parameter keluhan penyakit masyarakat pada tiap bulannya.

2. Menghasilkan knowledge dalam bentuk *cluster* menggunakan algoritma *k-means clustering*.
3. Menghasilkan model dan pertimbangan untuk menentukan penanganan atau penanggulangan yang tepat sasaran terhadap gejala penyakit masyarakat menggunakan Teknik data *mining* dengan algoritma *k-means clustering*

1.5 Manfaat Penelitian

Suatu penelitian yang dilaksanakan diharapkan dapat memberikan manfaat baik bagi kepentingan peneliti, organisasi, maupun ilmu pengetahuan bagi peneliti sesudahnya. Manfaat penelitian ini adalah jawaban dari tujuan penelitian yang dilaksanakan. Berikut manfaat penelitian yang akan didapatkan dalam penelitian ini sebagai berikut :

1. Manfaat bagi instansi yaitu untuk memberikan teknik dan pertimbangan terhadap data gejala penyakit masyarakat, sehingga memudahkan pihak puskesmas dalam menentukan penanganan dan pencegahan penyakit secara efektif.
2. Manfaat bagi peneliti yaitu dapat mengimplementasikan ilmu yang telah didapat di bangku perkuliahan.
3. Manfaat bagi pembaca pada umumnya yaitu sebagai referensi khususnya dalam memahami *data mining* dengan algoritma *K-means clustering*.

1.6 Sistematika Penulisan

Pada suatu penelitian dilakukan pendokumentasian di tiap tahapan dalam bentuk tertulis. Sistematika penulisan merupakan kerangka penelitian agar lebih terstruktur. Terdapat enam BAB untuk mekanisme penelitian ini dengan gambaran umum sebagai berikut :

Bab I : Pendahuluan

Bab ini berisikan pendahuluan dan latar belakang masalah dalam melakukan penelitian, gambaran umum tentang permasalahan yang akan dibahas pada bab-bab berikutnya, yaitu perumusan masalah, Batasan masalah, tujuan penelitian, manfaat penelitian yang diakhiri dengan sistematika penulisan.

Bab II : Landasan Teori

Pada bab ini dijelaskan tentang beberapa hal yang mendasar tentang masalah yang akan dibahas, yang mana nantinya akan digunakan sebagai landasan dasar dalam melakukan pemecahan masalah pada penelitian yang dilakukan.

Bab III : Metodologi Penelitian

Pada bab ini berisikan tentang langkah kerja dan prosedur serta metode yang akan dilakukan dalam penelitian.

Bab IV : Kesimpulan dan Saran

Pada bab ini dijelaskan tentang kesimpulan dan saran-saran dalam melakukan penelitian dan pengembangan untuk selanjutnya. *k-means clustering*. Selain itu bab ini juga memuat hasil *clustering* gejala penyakit dan analisa hasil *clustering* yang dapat dijadikan sebagai pertimbangan oleh pihak Puskesmas Padusunan.

Bab V : Implementasi dan Pengujian

Pada bab ini berisi penerapan teknik *clustering* dengan algoritma *k-means clustering* menggunakan aplikasi WEKA dan pengujian terhadap data sampel dan data seluruh mahasiswa yang mengajukan beasiswa kurang mampu.

Bab VI : Kesimpulan dan Saran

Pada bab ini dijelaskan tentang kesimpulan dan saran-saran dalam melakukan penelitian dan pengembangan untuk selanjutnya.

BAB II

LANDASAN TEORI

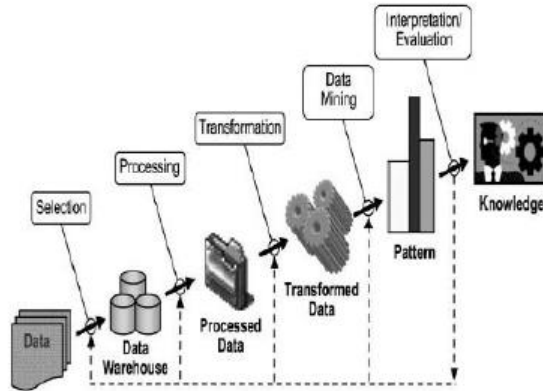
2.1 Rekam Medis

Menurut Pasal 46 ayat (1) UU Rekam medis adalah berkas atau file yang berisi catatan dan dokumen mengenai identitas pasien, pemeriksaan, pengobatan, tindakan dan pelayanan lain yang sudah diberikan untuk pasien. Peraturan Menteri Kesehatan Nomor 749a/Menkes/Per/XII/1989 mengenai Rekam Medis menjelaskan tentang rekam medis adalah berkas yang berisikan catatan dan dokumen mengenai identitas pasien, yaitu berupa pemeriksaan, pengobatan, tindakan dan pelayanan lain kepada pasien pada sarana pelayanan kesehatan yang diperbaharui dengan Permenkes Nomor 69 /MenKes /Per/III/2008 tentang Rekam Medis mengatakan rekam medis adalah berkas berisi catatan dan dokumen tentang pasien yang berisi identitas, pemeriksaan, pengobatan, tindakan medis lain pada sarana pelayanan kesehatan untuk rawat jalan, rawat inap *baik* dikelola pemerintah ataupun swasta. Kedua pengertian rekam medis tersebut kita menemukan perbedaan dimana Permenkes hanya menitik beratkan pada sarana pelayanan kesehatan, sedangkan dalam UU Praktik Kedokteran tidak. Hal ini menunjukkan pengaturan rekam medis pada UU Praktik Kedokteran lebih luas, berlaku baik untuk sarana kesehatan maupun di luar sarana kesehatan. Secara umum rekam medis berguna untuk: (Rachmad Abduh, 2021)

1. Sebagai alat interaksi antara dokter atau tenaga kesehatan
2. Merupakan dasar *planing* pengobatan atau perawatan yang harus diberikan kepada pasien.
3. Sebagai alat bukti fisik atas *service* dan pengobatan pada pasien.
4. Sebagai dasar analitis studi. penilaian mutu pelayanan pada pasien
5. Melindungi kepentingan hukum bagi pasien.rumah sakit.maupun dokter dan tenaga kesehatan lainnya.
6. Menyediakan data khusus yang sangat bermanfaat untuk keperluan penelitian
7. Sebagai dasar perhitungan biaya pelayanan medis pasien
8. Menjadi sumber penyimpanan dan bahan pertanggungjawaban

2.2 Knowledge Discovery in Database (KDD)

Knowledge Discovery in Database In Database adalah suatu metode atau model yang digunakan untuk memperoleh pengetahuan yang tersembunyi dari database yang ada, pengetahuan yang didapatkan dapat menjadi acuan dalam mengambil suatu keputusan (Nugraha, 2021). Penemuan pengetahuan di dalam suatu database ini dalam tahapannya secara sistematis dan terstruktur dalam mengidentifikasi suatu data sehingga menemukan pola-pola tertentu, dengan demikian dapat menghasilkan suatu informasi atau pengetahuan yang sebelumnya tidak terlihat di karenakan data yang besar dan kompleks. Data mining merupakan suatu proses yang sangat penting didalam tahapan KDD, dimana dalam proses ini terjadi pencarian suatu pola atau informasi yang unik dari data yang ada yang diseleksi dengan memakai suatu cara atau metode tertentu, pemilihan suatu algoritma atau suatu metode tergantung pada kebutuhan suatu tujuan KDD. Berikut adalah proses KDD : (Molina-Coronado, 2020)



Gambar 2.1 Proses *Knowledge Discovery in Database*

1. *Data Collection* adalah proses pengumpulan dan memilih data mentah dari sumbernya. Data ini harus memberikan informasi yang diperlukan untuk memecahkan masalah yang dihadapi.
2. *Data prepirocessing* adalah melibatkan ekstraksi seperangkat fitur yang valid dan terstruktur yang diperlukan untuk penggalian pengetahuan. Penerapan metode pemilihan untuk menghindari faktor-faktor seperti ketidak samaan, nilai yang tidak sesuai atau nilai yang hilang, dan derivasi fitur baru dari yang lain untuk menyusun data dengan cara yang sesuai untuk penerapan algoritma.
3. *Data Reduction* adalah langkah yang diperlukan untuk mengoptimalkan penerapan algoritme data mining, mengurangi waktu komputasinya dan/atau meningkatkan hasilnya. Ini melibatkan pemilihan subset fitur (penghapusan fitur yang tidak berguna atau berlebihan), pengurangan dimensi (proyeksi sekumpulan fitur ke dalam ruang yang lebih kecil) atau modifikasi set sampel (menambah atau menghapus sampel)
4. *Data Mining* adalah proses ini adalah inti dari KDD, tetapi langkah dapat berhasil hanya jika langkah-langkah sebelumnya dilakukan dengan benar. Terdiri dari memilih dan menerapkan teknik, termasuk algoritma *Machine Learning* (ML), untuk mengekstrak pengetahuan dari data dengan menemukan pola dan hubungan yang berguna antara fitur. Pilihan algoritma penambangan data tertentu tergantung pada masalah yang dihadapi dan sifat data yang tersedia.

5. *Interpretation and Evaluation*, menggunakan seperangkat metrik, mengukur kinerja dan efektivitas *algoritme* data mining, dan, akibatnya, tugas-tugas yang dilakukan pada langkah-langkah proses KDD sebelumnya. Ini juga terdiri dari visualisasi hasil dan pelaporan.

2.3 Data Mining

Data Mining adalah suatu Teknik *artificial Technology* yang modern dan terbaru digunakan untuk mencari pengetahuan atau knowledge yang tersembunyi dari suatu kumpulan data atau *data base* yang banyak (Mohammad, dkk, 2020). *Data mining* memiliki makna sebagai suatu metode yang bertujuan untuk mengidentifikasi, menggali, mencari suatu *knowledge* yang tersembunyi sehingga memiliki manfaat dalam mengambil suatu keputusan atau sebagai patokan dalam mengambil keputusan. Proses menambang suatu pengetahuan dalam data mining memerlukan kolaborasi dari beberapa disiplin ilmu seperti teknologi *database*, *machine learning*, *statistic*, visualisasi data dan sebagainya (Darma, dkk, 2020).

Data yaitu kumpulan fakta yang ada atau sebuah entiti yang tidak memiliki arti. Sedangkan mining yaitu proses penambangan atau pencarian. Data Mining itu dapat diartikan yaitu proses penambangan data yang menghasilkan pengetahuan (Syahril, M., Erwansyah, K., dan Yetri, M. 2020). Data Mining adalah Serangkaian langkah untuk menggali nilai tambah berupa informasi yang selama ini tersembunyi secara manual dari suatu *data base* dengan melakukan *mining* pola pola dari data dengan tujuan untuk memanipulasi data menjadi informasi yang lebih berharga yang diperoleh dengan caramengekstraksi dan mengenali pola yang penting atau menarik dari data yang terdapat dalam basis data.

2.3.1 Fungsi *Data Mining*

Fungsi *Data mining* yaitu mengidentifikasi fakta-fakta atau kesimpulan-kesimpulan yang di sarankan berdasarkan penyaringan melalui data untuk menjelajahi pola-pola atau anomali-anomali data. *Data Mining* mempunyai 5 fungsi: (Nasution, Halim, & Marsono, 2020)

1. *Classification*, yaitu memberikan kesimpulan definisi karakteristik sebuah kelompok data.
2. *Clustering*, yaitu mengidentifikasikan grup barang atau produk yang memiliki karakteristik *unic*.
3. *Association*, yaitu mengidentifikasikan relasi antara kejadian yang terlaksana pada suatu periode.
4. *Sequencing*, mengidentifikasikan relasi yang tidak sama pada suatu satuan waktu tertentu.
5. *Forecasting*, memprediksi nilai yang akan terjadi pada masa yang akan datang bersis pola data.

2.3.2 Tujuan *Data Mining*

Data Mining bertujuan untuk menggali atau menemukan informasi yang ada pada suata *database*, yang mana ketika dilakukan dengan cara manual informasi yang tersembunyi tersebut tidak dapat ditemukan. Berikut beberapa tujuan dari *data mining* : (Nasution, Halim, & Marsono, 2020)

1. *Explanatory* adalah untuk menjelaskan kondisi suatu *research*.
2. *Confirmatory* adalah untuk mempertegas suatu hipotesis.
3. *Exploratory* adalah untuk menganalisis relasi data yang baru

2.3.3 Tahapan dalam Data Mining

Sebelum mengetahui teknik-teknik yang dapat digunakan dalam data mining operasi yang dapat dihubungkan dengan data mining sebagai berikut. Operasi itu biasa disebut *Predictive Modeling*. Berikut adalah proses *predictive modeling* : (Nasution, Halim, & Marsono, 2020)

1. Estimasi

Digunakan untuk melakukan estimasi terhadap sebuah data baru yang tidak memiliki keputusan berdasarkan histori data yang telah ada. Contohnya ketika *melakukan* Estimasi Pembiayaan pada saat pembangunan sebuah Hotel baru pada Kota yang berbeda.

2. Asosiasi

Digunakan untuk mengenali kelakuan dari kejadian- kejadian khusus atau proses dimana hubungan asosiasi muncul pada setiap kejadian. Adapun metode pemecahan masalah yang sering digunakan seperti Algoritma Apriori. Contoh pemanfaatan Algoritma Asosiasi yaitu pada Bidang Marketing ketika sebuah Minimarket melakukan Tata letak produk yang dijual berdasarkan Produk-produk mana yang paling sering dibeli konsumen, selain itu seperti tata letak buku yang dilakukan pustakawan di perpustakaan

3. Klasifikasi

Suatu teknik dengan melihat pada kelakuan dan atribut dari kelompok yang telah didefinisikan. Teknik ini dapat memberikan klasifikasi pada data baru dengan memanipulasi data yang ada yang telah diklasifikasi dan dengan menggunakan hasilnya untuk memberikan sejumlah aturan. Salah satu contoh yang mudah dan populer adalah dengan *Decision tree* yaitu salah satu metode klasifikasi yang paling populer karena mudah untuk interpretasi seperti Algoritma C4.5, ID3 dan lain-lain. Contoh pemanfaatannya adalah pada bidang Akademik yaitu Klasifikasi siswa yang layak masuk kedalam kelas unggulan atau akselerasi di sekolah tertentu.

4. Klustering

Digunakan untuk menganalisis pengelompokan berbeda terhadap data, mirip dengan klasifikasi, namun pengelompokan belum didefinisikan sebelum dijalankannya tool data mining. Menggunakan metode neural *network* atau statistik, analitikal hierarki cluster. Clustering membagi item menjadi kelompok- kelompok berdasarkan yang ditemukan *tool* data mining.

5. Prediksi

Algoritma prediksi biasanya digunakan untuk memperkirakan atau *forecasting* suatu kejadian sebelum kejadian atau peristiwa tertentu terjadi. Contohnya pada bidang Klimatologi dan Geofisika, yaitu bagaimana Badan Meterologi Dan Geofisika (BMKG) memperkirakan tanggal tertentu bagaimana Cuacanya, apakah Hujan, Panas dan lain sebagainya. Ada beberapa metode yang sering digunakan salah satunya adalah *Metode Rough Set*. *Data mining* juga sama halnya dengan konsep *Neural Network* mengandung dua pengelompokan. Berikut adalah pengelompokan yang ada pada *data Mining* yaitu:

1. *Supervised Learning* yaitu pembelajaran menggunakan guru dan biasanya ditandai dengan adanya Class/Label/Target pada himpunan data. Adapun metode- metode

yang digunakan yang bersifat supervised learning seperti Metode Prediksi dan Klasifikasi seperti Algoritma C4.5, Metode Rough Set dan Lain-lain.

2. *Unsupervised Learning* yaitu pembelajaran tanpa menggunakan guru dan biasanya ditandai pada himpunan datanya tidak memiliki atribut keputusan atau Class/Label/Target. Adapun metode-metode yang bersifat Unsupervised Learning yaitu Metode *Estimasi*, *Clustering*, Dan Asosiasi seperti *Regresi Linier*, *Analytical Hierarchy Clustering* dan lain-lain.

Suatu rangkaian proses, *data mining* dapat dibagi menjadi beberapa tahap. Tahap-tahap tersebut bersifat interaktif, pemakai terlibat langsung atau dengan perantara *knowledge base*. Tahap-tahap *data mining* sebagai berikut : (Oulhiq, 2022)

1. *Formatting Data*

Data mentah tidak selalu disimpan dalam format yang dapat dieksploitasi. Dengan demikian, diperlukan langkah pemformatan data sebelum mengerjakan pra-pemrosesan data. Pemformatan data digunakan untuk menyiapkan data dengan cara yang paling sesuai dengan kebutuhan langkah selanjutnya. Jika data dikumpulkan dari sumber yang berbeda, atau diperbarui oleh bagian yang berbeda, kemungkinan akan ada anomali dalam cara data diformat. Oleh karena itu, menyeragamkan nilai dalam kolom akan memastikan data akan dikumpulkan dengan benar.

2. *Data integration*

Integrasi data adalah proses penggabungan dan konsolidasi data dari sumber yang berbeda. Dalam industri proses, integrasi data mungkin menyangkut kombinasi data otomatis dan manual. Dari laporan operator hingga data yang disimpan di server sejarawan. Kemungkinan penggunaan integrasi data lainnya

adalah ketika data direkam dari sensor yang berlebihan. Misalnya, dalam studi kasus, dua sensor digunakan untuk mengukur nilai laju aliran. Sensor digunakan satu per satu. Dengan demikian, nilai maksimum kedua sensor digunakan sebagai nilai sebenarnya dari parameter yang diukur.

3. *Data Selection*

Pemilihan data, dalam konteks ini, adalah proses pengambilan data yang relevan dengan tujuan identifikasi sistem. Dalam makalah ini, karena fokus identifikasi sistem adalah pada data di sekitar operasi, rasio Di titik sistem, pemilihan data diterapkan untuk mengekstrak data di sekitar titik operasi yang diinginkan. Untuk mengekstrak data ini, Campuran Gaussian Model (GMM) digunakan. GMM adalah teknik pengelompokan tanpa pengawasan yang mengelompokkan berdasarkan estimasi kepadatan probabilitas menggunakan algoritma *Expectation Maximization* (EM).

4. *Data pre-filtering*

Filter anti-aliasing digunakan untuk tidak mendistorsi bagian data yang menarik. Frekuensi pemutusan filter yang digunakan dipilih sama dengan frekuensi. Misalkan frekuensi pengambilan sampel yang(perlu ditambah)

5. *Data re-sampling*

Pilihan waktu sampling adalah salah satu pilihan penting dari identifikasi sistem. Meskipun benar bahwa waktu pengambilan sampel yang lebih rendah akan menghasilkan lebih sedikit kehilangan informasi, pengambilan sampel yang sangat cepat menyebabkan masalah numerik.

6. *Data cleaning*

Pencilan. Deteksi *outlier*, sebagai bagian dari pra-proses data sing, tetap sangat penting untuk identifikasi sistem. Pencilan yang tidak terdeteksi menurunkan kinerja model perkiraan. Banyak pendekatan telah dikembangkan untuk deteksi outlier .

7. *Data normalization*

Dalam konteks industri, parameter tidak memiliki urutan besarnya yang sama. Akibatnya, variabel dengan nilai numerik tertinggi secara otomatis akan mendapatkan bobot tertinggi dalam fungsi biaya minimum yang digunakan untuk pemodelan. Dengan demikian, normalisasi digunakan untuk menghilangkan satuan pengukuran dalam data.

2.4 Clustering

Analisis Cluster adalah suatu metode mempelajari data berdasarkan ciri-ciri dari sesuatu itu sendiri, dengan tujuan untuk mengklasifikasikan hal-hal yang serupa (Zou, H. 2020). Prinsipnya adalah bahwa data-data dalam kategori yang sama memiliki kesamaan yang lebih besar, dan data-data dalam kategori yang berbeda memiliki kesamaan terkecil (yaitu, perbedaannya lebih besar). Analisis pengelompokan memiliki karakteristik sebagai berikut: Sangat cocok untuk klasifikasi tanpa pengetahuan apriori; itu dapat menangani klasifikasi yang ditentukan oleh banyak variabel; analisis pengelompokan adalah metode analisis eksplorasi; pengelompokan terutama berfokus pada analisis

pengelompokan berbasis jarak. Teknik clustering telah ada sejak zaman Aristoteles. Adapun jenis *clustering* sebagai berikut : (Singh, 2020)

1. *Hierarchical Clustering*

Umumnya dikenal sebagai struktur seperti pohon dan pengelompokan dihasilkan oleh urutan partisi bersarang. Dalam hierarki pengelompokan pohon disebut dendrogram dan kumpulan data dikenal sebagai simpul akar. Berdasarkan ukuran kedekatan dan iteratif penambahan atau pembagian tinggi setiap *cluster* yang ditetapkan dalam dendrogram dibagi. Setiap cluster set seperti node daun dengan ketinggian cluster yang berbeda menunjukkan seberapa baik cluster dipisahkan. Pengelompokan hierarkis dapat dilakukan dalam dua pola, ketika kita menggabungkan klaster sambil bergerak dari bawah ke atas dikenal sebagai pengelompokan aglomerasi dan ketika kita melanjutkan dengan memecah himpunan klaster dari atas ke bawah disebut pengelompokan divisi.

2. *Partitional Clustering*

Pengelompokan partial lebih disukai oleh para peneliti karena kemampuannya untuk melakukan pengelompokan mau tidak mau bahkan dengan kumpulan data yang besar karena persyaratan komputasi yang lebih sedikit. Tetapi dalam hal menjadi logis dan konsisten, pengelompokan hierarkis jauh lebih unggul daripada pengelompokan Partisi. Pengelompokan partisi umumnya mengkategorikan data dalam cluster K, dan setiap data milik satu *cluster*. Ada empat kategori besar untuk mendapatkan clustering partisi seperti Centroid based, Density based, Model based dan *Graph based*. Selain clustering partisi umum (*hard clustering*) clustering juga dapat dilakukan secara overlapping yaitu soft clustering, dimana suatu data atau objek tertentu dapat menjadi bagian dari lebih dari satu cluster.

2.5 Algoritma K-Means Clustering

Algoritma *K-means* adalah salah satu algoritma dengan *partitional*, karena *K-Means* basisnya pada penentuan jumlah awal kelompok atau *cluster* dengan mengidentifikasi nilai *centroid* awalnya. Algoritma *K-means* memakai proses secara berulang-ulang untuk mendapatkan basis data *cluster*. Dibutuhkan jumlah *cluster* awal yang diharapkan sebagai input dan menghasilkan titik *centroid* akhir sebagai output atau keluaran. Metode *K-means* memilih pola *k* sebagai titik awal *centroid* secara acak atau random. Jumlah iterasi untuk mencapai *cluster centroid* akan dipengaruhi oleh calon *cluster centroid* awal secara random. Sehingga didapat cara dalam pengembangan algoritma dengan menentukan *centroid cluster* yang dilihat dari kepadatan data awal yang tinggi agar mendapatkan kinerja yang lebih tinggi. Pada intinya algoritma *k-means* mengacu pada *centroid* awal dan disesuaikan dengan jumlah *cluster* yang diinginkan, dan dilakukan iterasi sehingga menghasilkan keluaran yang diinginkan (Indraputra dan Fitriana, 2020).

Algoritma *K-Means* akan menghasilkan titik *centroid* yang dijadikan tujuan dari algoritma *K-Means*. Setelah iterasi *K-Means* berhenti, setiap objek dalam data set menjadi anggota dari suatu *cluster*. Nilai *cluster* ditentukan dengan mencari seluruh objek untuk menemukan *cluster* dengan jarak terdekat ke objek. Algoritma *K-means* mengelompokkan item data dalam suatu dataset ke suatu *cluster* berdasarkan jarak terdekat. Pada penelitian mereka, Bangoria et.al. melakukan survei terhadap efisiensi berbagai jenis algoritma clustering *K-means* yang berbeda terhadap efisiensinya dalam menentukan nilai pusat. Nilai *centroid* awal yang dipilih secara acak yang menjadi titik pusat awal, akan dihitung jarak dengan semua data menggunakan rumus *Euclidean Distance*. Data yang memiliki jarak pendek terhadap *centroid* akan membuat sebuah *cluster*. Proses ini berkelanjutan sampai tidak terjadi perubahan pada setiap kelompok.

K-Means Clustering merupakan salah satu metode clustering non-hirarki yang mengelompokkan data dalam bentuk satu atau lebih *cluster* atau kelompok. Data-

data yang memiliki karakteristik yang sama dikelompokkan dalam satu *cluste* atau kelompok yang lain sehingga data yang berada dalam satu *cluster* kelompok memiliki tingkat variasi yang kecil.

K-means adalah metode partisi yang sering digunakan dalam penyesuaian pada *macine study* dan analisa suatu pola data, algoritma ini sangat terpengaruh pada *centroid* awal, tapi tidak dapat menjamin sampai menemukan solusi akhir hal ini dikarenakan sentroid awal di cari secara acak untuk cluster yang diberikan (Manochandar, 2020). *K-means* biasanya dipengaruhi oleh inisialisasi dan perlu diberikan sejumlah *cluster* secara apriori. Secara umum jumlah *cluster* tidak diketahui. Dalam hal ini, indeks validitas dapat digunakan untuk menemukan nomor *cluster* yang seharusnya independen dari algoritma *clustering* (Sinaga, 2020). Menurut Virgo, Defit, & Yuhandri (2020) Langkah algoritma *k-means* sebagai berikut :

1. Memasukkan data yang diklaster
2. Menentukan jumlah klaster, sesuai dengan kebutuhan penelitian.
3. Menentukan sentroid (pusat klaster) dengan cara mengambil data secara acak. Pilih K *centroid* (titik pusat *cluster*) awal secara random. Dalam menentukan n buah pusat cluster awal dilakukan pemilihan bilangan random yang merepresentasikan urutan data input. Pusat awal cluster didapatkan dari data sendiri bukan dengan menentukan titik baru, yaitu dengan merandom pusat awal dari data
4. Menghitung pusat kelompok (*centroid* atau rata-rata) dari data yang ada di masing-masing kelompok. Lokasi *centroid* setiap kelompok diambil dari rata-rata (*mean*) semua nilai data pada setiap fiturnya. Jika M menyatakan jumlah data dalam sebuah kelompok, i menyatakan fitur ke- i dalam sebuah kelompok, dan p menyatakan dimensi data, maka persamaan untuk menghitung *centroid* fitur ke- i digunakan persamaan
5. Menghitung jarak antara data dengan pusat klaster, Alokasikan masing-masing data ke *centroid* atau rata-rata terdekat. Ada beberapa cara yang dapat dilakukan

untuk mengukur jarak data ke pusat kelompok, diantaranya adalah Euclidean berikut persamaan yang digunakan :

$$D(a, b) = \sqrt{(x_a - y_b)^2 + (x_a - y_b)^2 + (x_n - y_n)^2} \dots\dots\dots(2.1)$$

Keterangan :

$D(a, b)$ = jarak data ke i ke pusat klaster j

X_{ki} = data ke i pada atribut ke k

X_{kj} = titik pusat ke j pada atribut ke k

6. Hitung Kembali pusat klaster dengan keanggotaan klaster yang baru
7. apabila masih ada data yang berpindah kelompok atau apabila ada perubahan nilai centroid di atas nilai ambang yang ditentukan, atau apabila perubahan nilai pada fungsi objektif yang digunakan masih di atas nilai ambang yang ditentukan. Jika pusat klaster tidak berubah maka proses klaster telah selesai, jika belum maka ulangi langkah ke (4) sampai pusat klaster tidak berubah lagi.
8. Uji coba terhadap hasil penelitian yang telah dilakukan, pengujian ini bertujuan untuk mendapatkan keakuratan dari sistem yang telah dirancang dan apakah sistem tersebut telah berjalan dengan baik
9. Menarik kesimpulan dari hasil pengujian, pada tahapan ini didapatkan hasil pengujian algoritma *k-means clustering* dengan menggunakan software seperti excel, weka, dan rapid miner

Adapun karakteristik dari algoritma *K-Means* salah satunya adalah sangat sensitif dalam penentuan titik pusat awal cluster karena *K-Means* membangkitkan titik pusat cluster awal secara random. Pada saat pembangkitan awal titik pusat yang random tersebut mendekati solusi akhir pusat cluster, *K-Means* mempunyai kemungkinan yang tinggi untuk menemukan titik pusat cluster yang tepat. Sebaliknya, jika awal titik pusat tersebut jauh dari solusi akhir pusat cluster, maka besar kemungkinan ini menyebabkan hasil pengklasteran yang tidak tepat. Akibatnya *K-Means* tidak menjamin hasil pengklasteran yang unik. Inilah yang menyebabkan metode *K-Means* sulit untuk mencapai optimum global, akan tetapi hanya minimum lokal. Selain itu, algoritma *K-Means* hanya bisa digunakan untuk data yang atributnya bernilai numeric (Rahmadani, N, dan Kurniawan, 2020)..

2.6 Penelitian Terdahulu

Beberapa peneliti terdahulu telah melakukan penelitian terkait dengan penelitian yang telah dilakukan oleh penulis. Penelitian terdahulu yang telah dilakukan menjadi acuan bagi penulis agar dapat meminimalisir kesalahan dalam penelitian ini. Ada beberapa peneliti terdahulu yang telah melakukan penelitian terkait dengan penelitian ini diantaranya :

Tabel 2.1 Penelitian Terdahulu

No	Nama Peneliti	Judul	Hasil Penelitian
1	Ismail Virgo, Sarjon Defit, Yuhandri Yunus (2020)	Klasterisasi Tingkat Kehadiran Dosen Menggunakan Algoritma <i>K-Means Clustering</i>	Hasil penelitian ini adalah 3 kelompok yaitu 72 matakuliah diampu dosen non pns kelompok jarang melakukan pertemuan (4.7650%), 69 matakuliah yang diampu dosen non pns pada kelompok sedang dalam melakukan pertemuan (4.5665%), dan 1370 matakuliah yang diampu dosen non pns pada kelompok rajin melakukan pertemuan (90.6684%).
2	Sri Dewi,Sarjon Defit, Yuhandri Yunus (2021)	Akurasi Pemetaan Kelompok Belajar Siswa Menuju Prestasi Menggunakan Metode <i>K-Means</i>	Hasil dalam penelitian ini mendapatkan 3 cluster siswa, yaitu siswa Sangat Berprestsi, Berprestsi dan Kurang Berprestsi. Penelitian ini dapat dijadikan sebagai pedoman bagi guru pengajardalam pengambilan keputusan terhadap pembentukan kelompok belajar siswa di Kelas IX.C
3	Yusma Elda, Sarjon Defit, Yuhandri Yunus, Raemon Syaljumairi(2021)	Klasterisasi Penempatan Siswa yang Optimal untuk Meningkatkan Nilai Rata-Rata Kelas Menggunakan <i>K-Means</i>	Perhitungan klasterisasi siswa dengan <i>K-Means</i> berhasil mengelompokkan 90 siswa menjadi 3 klaster, cluster 1 berjumlah 47 siswa, cluster 2 berjumlah 10 siswa, dan cluster 3 berjumlah 33 siswa. Anggota klaster akan dibagi merata kedalam 3 kelompok.

Tabel 2.1 Penelitian Terdahulu (lanjutan)

No	Nama Peneliti	Judul	Hasil Penelitian
4	Rio Andika Malik, Sarjon Defit, Yuhandri (2018)	Perbandingan Algoritma <i>K-Means</i> Clustering dengan Fuzzy C-Means dalam Mengukur Tingkat Kepuasan Terhadap Televisi Dakwahsurau Tv	Hasil penelitian menunjukkan nilai RMSE clustering adalah sebesar 2.09879 dan besaran nilai penelusuran dari RMSE clustering menggunakan algoritma dengan pemodelan Fuzzy <i>K-Means</i> adalah sebesar 2.07911 dan kecepatan pemodelan Fuzzy C-Means lebih cepat dalam melakukan proses.
5	Kuntal Chowdhury, Debasis Chaudhuri, Arup Kumar Pal (2020)	<i>An entropy-based initialization method of K-means clustering on the optimal number of clusters</i>	Merancang sebuah algoritma untuk menghitung jumlah <i>cluster</i> dari kumpulan data menggunakan beberapa indeks validitas <i>cluster</i> . Algoritma inialisasi berbasis entropi dan diterapkan pada kumpulan data 2D dan 3D. Perbandingan dengan metode inialisasi yang ada telah direpresentasikan dalam makalah ini.
6	N. Gokilavani, B. Bharathi. (2021)	<i>Test case prioritization to examine software for fault detection using PCA extraction and K-means clustering with ranking</i>	Kinerja pekerjaan ini dianalisis dan diverifikasi silang dengan menganalisis peringkat prioritas untuk cluster, jumlah kesalahan berbeda yang terdeteksi berdasarkan jumlah kasus uji, matriks kedekatan antara kasus uji dan tingkat deteksi kesalahan.

Tabel 2.1 Penelitian Terdahulu (lanjutan)

No	Nama Peneliti	Judul	Hasil Penelitian
7	Bansal, M., Kumar, M., Kumar, M., & Kumar, K. (2021).	<i>An efficient technique for object recognition using Shi-Tomasi corner detection algorithm</i>	Mencapai akurasi pengenalan 85,9%, 80,8% dan 74,8%, pohon keputusan dan pengklasifikasi K-NN, masing-masing. Penulis telah menerapkan metodologi peningkatan adaptif untuk meningkatkan akurasi. Penulis telah mendapatkan peningkatan akurasi pengenalan sebesar 86,4% menggunakan peningkatan adaptif dengan pengklasifikasi.
8	Tuti Hartati ¹ , Odi Nurdiawan, Eko Wiyandi. (2021)	Analisis dan Penerapan Algoritma <i>k-means</i> dalam strategi promosi kampus akademi maritim suaka bahari	Menghasilkan jumlah cluster 30 (k=30). Dari hasil penelitian berdasarkan uji davies bouldin pada algoritma K-means menghasilkan nilai paling mendekati 0 adalah k = 29 dengan Davies bouldin : 0.070, dengan persebaran anggota cluster terbanyak ada pada cluster 16 berisi anggota cluster 115 items
9	Rian Ordila (2020)	Penerapan Data Mining untuk Penglompokan Data Rekam Medis Pasien berdasarkan Jenis Penyakit dengan Algoritma <i>Clustering</i>	Mendapatkan klasifikasi dengan menggunakan metode <i>K-means clustering</i> dan Rapid Miner yang dapat membantu pengelompokan berdasarkan wilayah, jenis penyakit, dan umur

Tabel 2.1 Penelitian Terdahulu (lanjutan)

No	Nama Peneliti	Judul	Hasil Penelitian
10	Amjad Rehman Khan (2020)	<i>Brain tumor segmentation using K-means clustering and deep learning with synthetic data augmentation for classification</i>	penelitian ini dilakukan pada data MRI yang telah dianalisis, menghasilkan tingkat ke akurasian yang tinggi dalam mendiagnosa tumor.
11	Latifa Suryani Nasution, Widiarti Rista Maya, Jufri Halim, Marsono. (2020)	Data Mining Untuk Menganalisa Pola Pembelian Perak Dengan Menggunakan Algoritma Fp-Growth Pada Toko Emas Dan Perak Adi Saputra Tanjung	Didapatkan support sebesar 9% dan nilai <i>confidence</i> sebesar 30% dengan jenis perak yaitu cincin putar, kalung, cincin rantai pilin dan anting. Hasilnya dapat membantu pemilik toko untuk mengambil keputusan dalam penentuan stok perak yang perlu diperbanyak sehingga meningkatkan keuntungan.
12	Muhammad Syahril, Kamil Erwansyah, Milfa Yetri. (2020)	Penerapan Data Mining Untuk Menentukan Pola Penjualan Peralatan Sekolah Pada Brand Wigglo Dengan Menggunakan Algoritma Apriori	Hasil perancangan aplikasi data mining dalam menentukan pola penjualan peralatan sekolah maka diperoleh. Membentuk pola kombinasi itemset maka dilakukan perhitungan terhadap data yang ada, dilakukan pencarian nilai support dan nilai confidence setelah itu dilakukan membentuk pola kombinasi itemset.

Tabel 2.1 Penelitian Terdahulu (lanjutan)

No	Nama Peneliti	Judul	Hasil Penelitian
13	Rachmad Abduh (2021)	Kajian Hukum Rekam Medis Sebagai Alat Bukti Malapraktik Medis	Catatan-catatan diri pasien mulia identitas.riwayat penyakit yang diderita pasien. Rekam medis bukan saja berguna bagi pasien tetapi juga berguna bagi dokter.karena dapat digunakan sebagai alat bukti jika terjadinya dugaan malapraktek medis ini menunjukkan begitu penting dan berharganya rekam medis.
14	Ridouane Oulhiq, KhalidBenjelloun, Yassine Kali d and Maarouf Saad. (2021)	<i>A data mining based approach for process identification using historical data</i>	Berdasarkan 20 langkah prediksi ke depan dan dibandingkan dengan metode yang ada. Dalam konteks ini, pendekatan yang diusulkan memberikan hasil terbaik dengan R2 masing-masing sebesar 98,11% dan 62,70% untuk prediksi 1 dan 20 langkah.
15	Jaja, Nandi Priatna, Tazkia Salsabila Ardan (2021)	<i>Implementation of Data Mining Technique for Performance of WFH and WFO Agents Using the K-Means Method Case Study Study of PT. Infomedia Telkom Consumer Profiling Services</i>	Hasil pembahasan ini menunjukkan bahwa terdapat perbedaan yang signifikan antara kinerja WFH dan Agen WFO. Dimana ada perbedaan yang cukup banyak dalam segi performa dalam pekerjaan antra karyawan yang kerja di kantor dengan karyawan yang bekerja di rumah

Tabel 2.1 Penelitian Terdahulu (lanjutan)

16	R. A. Indraputra, R. Fitriana (2020)	<i>K-Means Clustering</i> Data COVID-19	Dari hasil pengolahan data, diperoleh dua cluster data, dimana cluster2 memiliki jumlah terjangkit dan meninggal yang lebih tinggi dibandingkan dengan cluster1, maka daerah-daerah <i>cluster</i> tersebut perlu diprioritaskan penanganannya.
17	Surya Darma1, Sarjon Defit, Dedy Hartama1, (2020)	Penerapan Metode <i>K-Means</i> Dalam Pengelompokan Jumlah Wisatawan Asing Di Indonesia	Hasil penelitian ini diperoleh bahwa bulan Juni tahun 2017-2019 merupakan bulan yang paling banyak dikunjungi wisatawan asing di Indonesia dan bulan lainnya merupakan bulan yang rendah dikunjungi wisatawan asing di Indonesia
18	Nurul Rahmadani, Edi Kurniawa (2020)	Implementasi Metode <i>K-Means Clustering</i> Tunggakan Rekening Listrik pada PT. PLN (Persero) Gardu Induk Kisaran	Data jumlah penduduk menunggak, tarif daya, tunggakan. Hasil penelitian ini menunjukkan algoritma <i>k-means</i> mengelompokkan data tunggakan terbesar, sedang, dan terendah. dalam pembuatan penjualan tiket pesawat dan memberikan keamanannya.

BAB III

METODOLOGI PENELITIAN

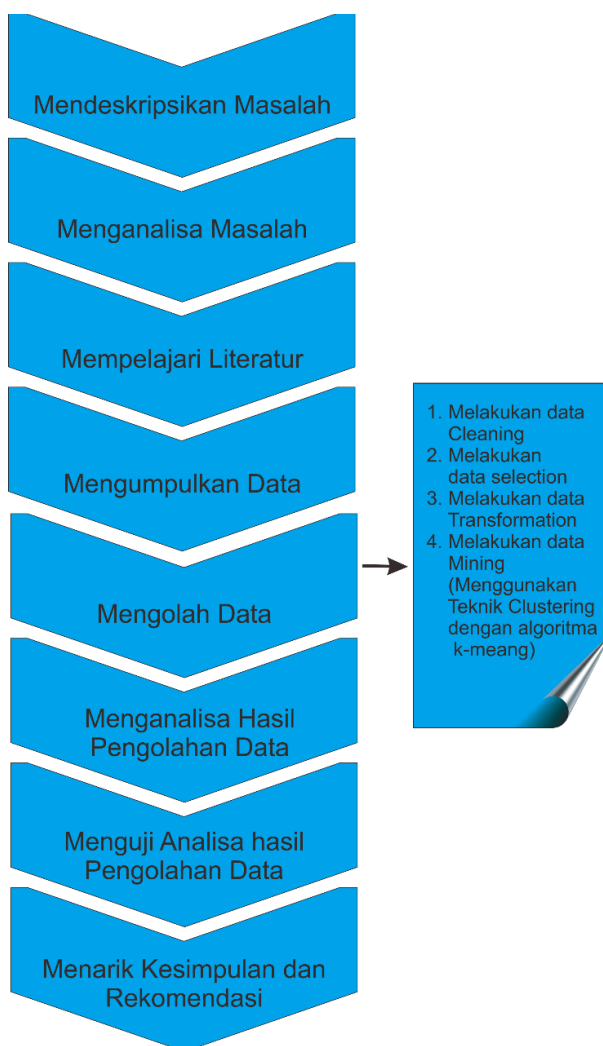
3.1 Pendahuluan

Metodologi penelitian merupakan sekumpulan peraturan, kegiatan, dan prosedur yang digunakan oleh pelaku suatu disiplin ilmu. Metodologi juga merupakan analisis teoritis mengenai suatu cara atau metode. Metodologi penelitian adalah suatu cabang ilmu pengetahuan yang menjelaskan mengenai cara-cara melaksanakan penelitian yang dimulai dari kegiatan mencari, mencatat, merumuskan, menganalisis, hingga menyusun laporan berdasarkan fakta-fakta atau gejala-gejala secara ilmiah.

Metodologi penelitian ini dilakukan secara sistematis, sehingga dapat digunakan sebagai pedoman bagi peneliti dalam melakukan penelitian ini agar hasil yang dicapai sesuai dengan tujuan yang ditetapkan sebelumnya dan dapat terlaksana dengan baik. Bab ini akan diuraikan kerangka kerja penelitian. Kerangka kerja dalam penelitian ini diawali dengan mendeskripsikan masalah, menganalisa masalah sampai pada tahap menarik kesimpulan dan rekomendasi.

3.2 Kerangka Kerja Penelitian

Kerangka kerja penelitian ini merupakan langkah-langkah yang akan dilakukan dalam menyelesaikan masalah yang akan dibahas. Kerangka kerja dalam penelitian ini adalah sebagai berikut :



Gambar 3.1 Kerangka Kerja Penelitian

Berdasarkan kerangka kerja penelitian pada Gambar 3.1, maka dapat menjelaskan beberapa kerangka kerja atau tahapan yang akan dilakukan dalam penelitian ini. Ada delapan kerangka kerja yang terdapat pada penelitian ini yaitu:

1. Mendeskripsikan masalah

Mendeskripsikan masalah adalah menjelaskan tentang masalah dalam penelitian secara terstruktur dan sistematis untuk mendukung dalam pengambilan keputusan yang lebih baik dan tepat. Masalah yang dideskripsikan dalam penelitian ini adalah klasifikasi keluhan pasien.

2. Menganalisa Masalah

Menganalisa masalah adalah tahap di mana peneliti menganalisa permasalahan yang ada sebagai langkah untuk dapat memahami masalah yang ada. Belum adanya klasifikasi keluhan pasien selama ini. Penerapan data mining teknik *clustering* menggunakan algoritma *K-means clustering* membantu dalam menentukan kelompok dari keluhan pasien yang ada.

3. Mempelajari Literatur

Tahap ini akan dicari dan dikumpulkan serta dipelajari sejumlah literatur mengenai teori dan konsep yang mendukung penyelesaian permasalahan dalam penelitian. Literatur yang digunakan berupa buku referensi atau buku penunjang, jurnal internasional dan nasional serta konsep-konsep yang mendukung dalam menyelesaikan penelitian ini. Dilakukan studi kepustakaan yaitu dengan membaca buku-buku yang menunjang dalam melakukan penganalisisan terhadap data dan informasi yang didapat. Literatur-literatur yang dipelajari tersebut diseleksi untuk dapat ditentukan literatur-literatur mana yang akan digunakan dalam penelitian.

4. Mengumpulkan Data

Pengumpulan data dilakukan observasi yaitu pengamatan secara langsung di tempat penelitian sehingga permasalahan yang ada dapat diketahui secara jelas. Kemudian dilakukan *interview* atau wawancara dengan staf atau tenaga kesehatan yang ada di Puskesmas Padusunan

5. Mengolah Data

Data yang diperlukan dalam penelitian selesai dikumpulkan, langkah selanjutnya adalah melakukan pengolahan data dengan beberapa tahap diantaranya adalah :

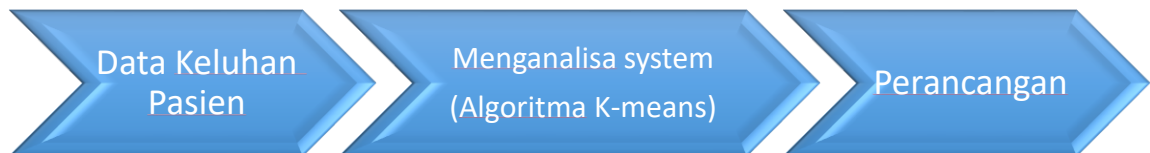
- a. *Data cleaning*, dilakukan untuk menghilangkan data yang tidak konsisten atau tidak relevan.
 - b. *Data selection*, dilakukan untuk mengambil data yang sesuai untuk dianalisa. Pada penelitian ini atribut yang digunakan berupa data keluhan pasien perperiode (perbulan).
 - c. *Data transformation*, agar data dapat diolah menggunakan algoritma *k-means*
 - d. *Data mining*, setelah data ditransformasikan maka data tersebut dapat diproses menggunakan algoritma *k-means clustering*.
6. Menganalisa hasil pengolahan data
- Tahap ini dilakukan analisa terhadap hasil pengolahan data berupa *cluster* calon penerima beasiswa kurang mampu yang telah diolah dengan menggunakan algoritma *k-means clustering*
7. Menguji analisa hasil pengolahan data
- Hasil analisa pengolahan data yang dilakukan dengan teknik clustering menggunakan algoritma *k-means clustering* selanjutnya dilakukan pengujian. Pengujian dilakukan terhadap data yang diolah secara manual berdasarkan langkah-langkah algoritma *k-means clustering*, kemudian dibandingkan dengan hasil perhitungan menggunakan WEKA. Pengujian dilakukan terhadap data sampel dan seluruh data yang meliputi pengujian iterasi dan hasil *cluster*.
8. Menarik kesimpulan dan rekomendasi
- Langkah terakhir setelah semua proses selesai, maka penulis dapat menarik suatu kesimpulan terhadap masalah yang diteliti. Selanjutnya diberikan rekomendasi dengan tujuan kelemahan-kelemahan dapat dihilangkan dan solusi kedepannya dapat diimplementasikan sehingga tujuan yang diharapkan dapat dicapai.

BAB IV

ANALISA DAN PERANCANGAN

4.1 Tahapan Analisa dan Perancangan

Pada bagian ini akan digambarkan tahapan analisa dan perancangan yang akan dibuat pada penelitian ini, pada penelitian ini sistem yang dirancang dengan menggunakan algoritma *K-means* sehingga dapat membantu dan menyelesaikan permasalahan. Sistem yang akan dibangun pada penelitian ini adalah sistem berbasis *web*.

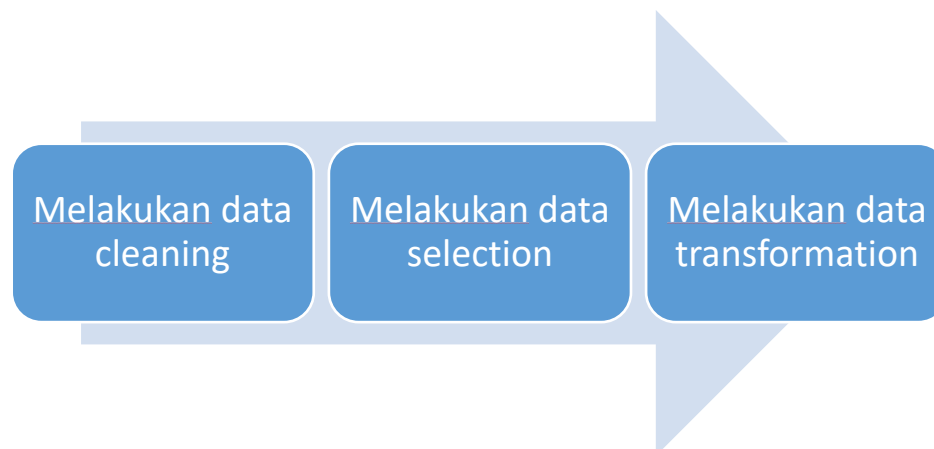


Gambar 4.1 Bagan Alir Anlisa dan Perancangan

Tahapan analisa dan perancangan ini berdasarkan kerangka kerja penelitian yang terdapat pada metodologi penelitian di bab 3. Tahapan kerja penelitian terdiri dari Mendeskripsikan masalah, menganalisa masalah, mempelajari literatur, mengumpulkan data, mengolah data, menganalisa hasil pengolahan data, menguji hasil analisa Pengolahan data, serta menarik kesimpulan dan rekomendasi. Agar mempermudah dalam melakukan analisa dan perancangan sistem maka di buat bagan alir dari proses analisa dan perancangan seperti pada Gambar 4. 1.

4.2 Data Keluhan Pasien

Data yang diambil pada penelitian yang dilakukan adalah data keluhan pasien. Data keluhan pasien ini terdapat pada rekam medis atau *medical record* pasien. Data ini dicatat oleh petugas puskesmas pada saat pasien melakukan pengobatan. Keluhan yang dirasakan pasien merupakan salah satu data yang ada didalam *medical record*. Setiap pasien memiliki *medical record* masing-masing. Setiap pasien melakukan pengobatan, maka data *record* pengobatan akan bertambah pada *medical record* pasien di setiap waktunya, dan semua data tersebut disimpan secara manual oleh pihak puskesmas. Sebelum melakukan *data mining*, maka terlebih dahulu perlu melakukan pra-proses data dengan menggunakan *software microsoft excel*. Berikut adalah bagan alir tahapan dalam melakukan pra-proses data :



Gambar 4.2 Bagan Alir Anlisa Pra Proses Data Keluhan

Berdasarkan Gambar 4.2 diatas Pra-proses data meliputi melakukan data *cleaning*, melakukan data *selection*, melakukan data *transformation*. Melakukan data *cleaning* dengan memilih data yang diperlukan dalam proses *data mining*. Melakukan data

selection dengan cara membersihkan data-data yang ada dan menentukan atribut data yang diperlukan dalam proses *data mining*. Melakukan data *transformation* yaitu menjadikan data yang siap diolah dengan menggunakan algoritama *K-means*.

4.2.1 Melakukan Data Cleaning

Data yang diperlukan dalam penelitian ini diperoleh melalui bagian pelayanan pasien Puskesmas Kampung Baru Padusunan. Data yang berjumlah 72 kategori keluhan pasien Puskesmas Kampung Baru Padusunan selama 6 bulan mulai dari Februari-Juli 2022. Data ini diperoleh pada data rekam medis dari masing-masing pasien, atribut yang tidak diperlukan pada penelitian ini diabaikan. Data yang telah diproses pada data cleaning ini selanjutnya diproses pada *data selection*.

4.2.2 Melakukan Data Selection

Adapun data keluhan yang dibersihkan berjumlah 72 kategori keluhan pasien Puskesmas Kampung Baru Padusunan selama 6 bulan mulai dari Februari-Juli 2022. Semua data keluhan pasien perbulan yang ada pada *medical record* masing-masing pasien dikumpulkan. Data keluhan yang ada seperti panas dalam, diare, nyeri ulu hati, sakit tenggorokan, pusing, pilek, sesak nafas, sakit gigi, dan lain sebagainya. Data yang dipakai dalam penelitian ini berjumlah Berikut adalah data keluhan pasien yang telah dikumpulkan :

Tabel 4.1 Keluhan Pasien

Penyakit	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
Amandel											1		1				2		1			1						
Asam lambung		1					1																1	2				
Asam urat								1		1				2														
BAB berdarah		1																					1					
BAB tidak lancar					1							1																
Badan dingin			1				1	1			2								1									
Badan Pegal		1																										
BAK tidak lancar								2				2		1									1					
Batuk		1						1			1											1			1	1		
Batuk berdahak		1		1			1		1					1									1			1		
Batuk berdarah											1								1									
Batuk kering		1						1	1	1		1		1	1					1		1					3	
BB turun											2																	
Bersin-bersin												1								2								
Bibir bengkak								2																				
Bintik merah di kulit																												
Bisul																												
Demam		2		1				1	1		1			1	2				2			3			3	1	3	
Diare																3												
Gatal-gatal		2	1				1				1			1	2	2						2		2	1	2		
Insomnia										1		3								1			3					

Sumber : Puskesmas Kampung Baru Padusunan

Tabel 4.1 Keluhan Pasien(Lanjutan)

Keluhan	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
Kaki bengkak			1					1																				
Kaki gatal		1	1											2								2						
Keputihan										1					2	1	1											
Kesemutan			1																1						1			
Lemas							1		1			1				1												
Lidah pahit																		1										
Mata Bengkak											1												2					
Mata berair																												
Mata kabur															1										1			
Mata Merah											1								1						1			
Mata Pedih									2								1											
Mencret												2									2					1		
Mual		2		1	1		2				2			1		2		2			1		1	1		2		
Mudah lelah			1	1	2					1					1													
Muntah									1		1			1					4		3				1			
Nafsu makan berkurang				1	1						1			1							1	3	1	1				
Nyeri sendi									3																			
Nyeri ulu hati		1	2	1	2							1				1	3				1					1		
Panas dalam				1																								
Pegal									2										2									

Sumber : Puskesmas Kampung Baru Padusunan

Tabel 4.1 Keluhan Pasien(Lanjutan)

Keluhan	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
Perut kambung				1					1																1			
Perut perih											1					2												
Pilek				1				1	1	1												2		1				
Pusing		3	1				1					3						1				1	5		1			
Sakit badan		1			2				2					2											1		1	
Sakit bahu																1												
Sakit BAK					1		1	1			1						2	1							2			
Sakit gigi							1																					
Sakit jari					1				2																			
Sakit kaki			1																		2						1	
Sakit Kepala			1	1			2	1				1		1	1							3		1	1		1	
Sakit kerongkongan		1		1	1										2									2				
Sakit leher																1												
Sakit lutut							1								1			3				2						
Sakit paha				1					1	4							2					2			1			
Sakit persendian							2					1													2			
Sakit perut					1				1									5	1		2		1			2		
Sakit pinggang			1	1						3							1		1									
Sakit pinggul								1							1	1										1		
Sakit pundak				1												2						1		2				
Sakit punggung		1	1							2	1																	

Sumber : Puskesmas Kampung Baru Padusunan

4.2.3 Proses Data Transformation

Agar data keluhan dapat diolah dengan menggunakan data algoritma *K-means* dan sesuai dengan tujuan dari penelitian ini, maka dilakukanlah *data transformation*. Tahapan transformation ini seperti merangkap semua data keluhan perhari menjadi data perbulan. Hasil dari *data transformation* pada Tabel 4.2 berikut ini :

Tabel 4.2 Hasil Transformation Data Keluhan Pasien

No	Keluhan	Februari	Maret	April	Mei	Juni	Juli
1	Amandel	6	8	4	2	3	3
2	asam lambung	5	7	8	6	10	9
3	Asam urat	4	6	4	2	3	5
4	BAB berdarah	2	9	6	5	1	4
5	BAB tidak lancar	2	7	6	6	1	2
6	badan dingin	6	9	6	7	14	7
7	Badan Pegal	1	7	9	6	7	10
8	BAK tidak lancar	6	11	2	4	5	2
9	batuk	6	16	5	6	17	9
10	batuk berdahak	7	5	8	7	7	7
11	Batuk berdarah	2	12	2	8	4	2
12	batuk kering	12	3	4	20	9	14
13	BB turun	2	14	3	4	3	2
14	Bersin-bersin	3	6	5	6	8	8
15	Bibir bengkak	2	8	1	5	4	7
16	Bintik merah di kulit	0	1	5	3	4	12
17	Bisul	0	25	2	8	6	4
18	demam	21	4	17	17	21	29
19	Diare	3	17	1	6	6	3
20	gatal-gatal	17	7	8	9	15	22
21	Insomnia	8	12	3	5	3	0
22	kaki bengkak	2	2	6	11	3	2

1.1 Menganalisa Sistem

Berdasarkan alir analisa dan perancangan yang telah digambarkan sebelumnya, dalam penelitian ini menggunakan algoritma *K-means clustering* dalam melakukan analisa sistem. Algoritma *K-means* adalah salah satu algoritma *clustering*. Secara garis besar Algoritma *K-means* menghasilkan *cluster* dari suatu *big data* dengan cara terlebih dahulu menentukan jumlah *cluster*, menentukan centroid awal lalu melakukan iterasi secara berulang, dan berhenti jika hasil iterasi yang dihasilkan sama dengan hasil iterasi sebelumnya.

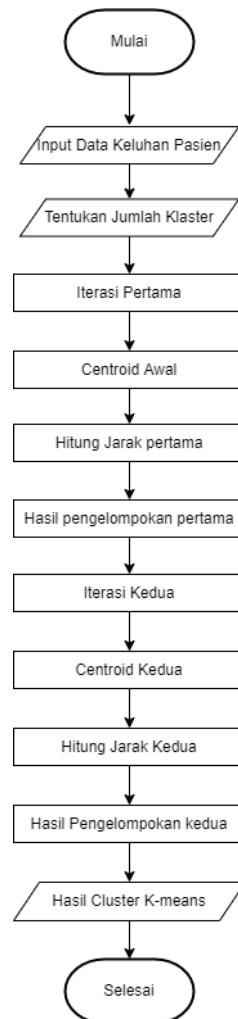
Algoritma *K-means* adalah salah satu algoritma dengan *partitional*, karena *K-Means* basisnya pada penentuan jumlah awal kelompok atau *cluster* dengan mengidentifikasi nilai *centroid* awalnya. Tahapan dari algoritma *Kmeans Clustering* ini adalah dengan menggambarkan langkah-langkah pemodelan *K-means clustering* dari langkah awal algoritma dimulai yaitu menentukan jumlah *cluster* yang akan dibentuk, menentukan *centroid* awal, dan melakukan iterasi sampai menghasilkan *cluster*. Berikut adalah langkah-langkah dari algoritma *K-means* secara detail dalam penelitian ini :

Algoritma Proses *K-Means*

1. Input data keluhan pasien
2. Tentukan jumlah *cluster*
3. Proses iterasi pertama
 - a. Tentukan pusat klaster (*Centroid*) awal
 - b. Menghitung jarak iterasi pertama
 - c. Hasil pengelompokkan data iterasi pertama
4. Proses iterasi kedua

- a. Tentukan pusat kluster kedua
 - b. Menghitung jarak iterasi kedua
 - c. Hasil pengelompokan data iterasi kedua
5. Hasil *K-Means Clustering*

Berdasarkan algoritma *K-means* yang telah dibuat, maka dapat digambarkan flowchart mengenai gambaran proses yang terjadi seperti Gambar 4.2. Algoritma dan *flowchart* tersebut akan diimplementasikan kedalam sistem berbasis *web* dengan menggunakan bahasa pemrograman PHP. Berikut adalah flowchart dari sistem yang akan dibuat :



Gambar 4.3 Flowchart Proses pada *K-Means Clustering*

Algoritma dan *flowchart* diatas telah diimplementasikan ke dalam sistem berbasis web dengan menggunakan bahasa pemrograman PHP. Berdasarkan Gambar 4.3 maka langkah-langkah pengolahan data pada Tabel 4.2 keluhan pasien di Puskesmas Kampung Baru Padusunan dengan menggunakan algoritma *K-means clustering* adalah sebagai berikut :

4.3.1 Input Data Keluhan Pasien

Data input yang akan diinputkan adalah data keluhan pasien yang dihasilkan pada tahapan pra-proses pembersihan dan transformasi yang sebelumnya dilakukan. Tahapan yang dilakukan dengan menggunakan *software microsoft excel*. Berikut adalah data keluhan pasien yang akan diinputkan :

Tabel 4.3 Keluhan Pasien Perbulan

No	Keluhan	Februari	Maret	April	Mei	Juni	Juli
1	Amandel	6	8	4	2	3	3
2	asam lambung	5	7	8	6	10	9
3	Asam urat	4	6	4	2	3	5
4	BAB berdarah	2	9	6	5	1	4
5	BAB tidak lancar	2	7	6	6	1	2
6	badan dingin	6	9	6	7	14	7
7	Badan Pegal	1	7	9	6	7	10
8	BAK tidak lancar	6	11	2	4	5	2
9	batuk	6	16	5	6	17	9
10	batuk berdahak	7	5	8	7	7	7
11	Batuk berdarah	2	12	2	8	4	2
12	batuk kering	12	3	4	20	9	14
13	BB turun	2	14	3	4	3	2
14	Bersin-bersin	3	6	5	6	8	8
15	Bibir bengkak	2	8	1	5	4	7
16	Bintik merah di kulit	0	1	5	3	4	12
17	Bisul	0	25	2	8	6	4
18	demam	21	4	17	17	21	29

Tabel 4.3 Keluhan Pasien Perbulan (lanjutan)

No	Keluhan	Februari	Maret	April	Mei	Juni	Juli
19	Diare	3	17	1	6	6	3
20	gatal-gatal	17	7	8	9	15	22
21	Insomnia	8	12	3	5	3	0
22	kaki bengkak	2	2	6	11	3	2
23	kaki gatal	6	7	3	3	8	6

4.3.2 Tentukan Jumlah Cluster

Tahapan berikutnya yang akan dilakukan sebelum pengolahan data adalah menentukan jumlah *cluster*. Jumlah *cluster* adalah jumlah kelompok data yang akan dihasilkan pada penelitian ini. Pada penelitian ini penulis akan mengelompokkan data keluhan pasien menjadi 3 *cluster*.

4.3.3 Proses Iterasi Pertama

Proses iterasi pertama adalah tahapan yang dilakukan untuk mendapatkan kelompok data keluhan pasien berdasarkan jarak terdekat data dengan pusat klaster yang telah ditentukan. Pada tahapan iterasi pertama terdiri dari beberapa tahapan yaitu menentukan *centroid* awal, menghitung jarak, dan mengelompokkan data.

1. Tentukan pusat klaster (*Centroid*) awal

Pada tahapan ini, dalam menentukan *centroid* awal diambil secara acak dari data keluhan pasien yang akan diolah. Data yang akan dijadikan *centroid* awal dapat

dilihat pada tabel 4.4. Data *centroid* awal terdiri dari C1, C2, dan C3, berikut adalah tabel data *centroid* awal :

Tabel 4.4 Centroid Awal

Titik Awal Pusat	Nama	Februari	Maret	April	Mei	Juni	Juli
Cluster 1	Mual	18	5	15	7	8	19
Cluster 2	Sakit persendian	5	8	14	3	3	2
Cluster 3	Kaki gatal	7	7	3	3	8	6

2. Menghitung jarak iterasi pertama

Proses iterasi pertama akan menghasilkan kelompok data pada pengolahan data pertama. Penulis menggunakan rumus *euclidean* untuk menghitung jarak data terhadap *centroid* awal. Berikut adalah tahapan perhitungan iterasi pertama :

$$D(a, b) = \sqrt{(x_a - y_b)^2 + (x_a - y_b)^2 + (x_n - y_n)^2}$$

Perhitungan jarak keluhan pertama dengan pusat *cluster* pertama adalah :

$$\begin{aligned} D(1, 1) &= \sqrt{(6 - 18)^2 + (8 - 5)^2 + (4 - 15)^2 + (2 - 7)^2 + (3 - 8)^2 + (3 - 19)^2} \\ &= 24,08318916 \end{aligned}$$

Perhitungan jarak keluhan pertama dengan pusat *cluster* kedua adalah :

$$\begin{aligned} D(1, 2) &= \sqrt{(6 - 5)^2 + (8 - 8)^2 + (4 - 14)^2 + (2 - 3)^2 + (3 - 3)^2 + (3 - 2)^2} \\ &= 10,14889157 \end{aligned}$$

Perhitungan jarak keluhan pertama dengan pusat *cluster* ketiga adalah :

$$D(1, 3) = \sqrt{(6 - 7)^2 + (8 - 7)^2 + (4 - 3)^2 + (2 - 3)^2 + (3 - 8)^2 + (3 - 6)^2}$$

$$= 6,164414003$$

Perhitungan jarak keluhan kedua dengan pusat *cluster* pertama adalah :

$$D(2, 1) = \sqrt{(5 - 18)^2 + (7 - 5)^2 + (8 - 15)^2 + (6 - 7)^2 + (10 - 8)^2 + (9 - 19)^2}$$

$$= 18,08314132$$

Perhitungan jarak keluhan kedua dengan pusat *cluster* kedua adalah :

$$D(2, 1) = \sqrt{(5 - 5)^2 + (7 - 8)^2 + (8 - 14)^2 + (6 - 3)^2 + (10 - 3)^2 + (9 - 2)^2}$$

$$= 12$$

Perhitungan jarak keluhan kedua dengan pusat *cluster* ketiga adalah :

$$D(1, 3) = \sqrt{(5 - 7)^2 + (7 - 7)^2 + (8 - 3)^2 + (6 - 3)^2 + (10 - 8)^2 + (9 - 6)^2}$$

$$= 7,141428429$$

Perhitungan ini dilakukan juga pada data keluhan pasien ke 3 sampai data ke 72. Hasil perhitungan jarak keluhan pasien dengan ketiga pusat *cluster* awal. Berikut adalah hasil iterasi 1 pada Tabel 4.4 :

Tabel 4.5 Jarak Setiap Data Keluhan pada Iterasi 1

Keluhan ke	<i>Cluster</i>		
	C1	C2	C3
1	24,08318916	10,14889157	6,164414003
2	18,08314132	12	7,141428429
3	23,74868417	10,72380529	6,164414003
4	25,11971337	9,273618495	9,746794345
5	26,07680962	9,327379053	10,39230485
6	20,51828453	15,09966887	8,185352772
7	20,29778313	11,44552314	9,899494937
8	25,61249695	12,60952021	6,633249581
9	23,38803113	20	13,60147051
10	17,74823935	10,29563014	6,8556546
11	27,92848009	13,96424004	10,39230485
12	18,86796226	25,35744467	19,89974874
13	28,35489376	12,92284798	10,77032961
14	21,16601049	12,60952021	5,830951895
15	25	14,4222051	7,141428429
16	22,82542442	16	11,87434209
17	33,51119216	22,27105745	20,174241
18	19,57038579	39,11521443	35,98610843
19	28,74021573	16,52271164	11,91637529
20	10,77032961	27,58622845	21,58703314
21	26,13426869	12,40967365	9,539392014
22	26	13,15294644	12,80624847
23	21,84032967	12,80624847	1

3. Hasil Pengelompokan Iterasi Pertama

Clustering objek dengan memasukkan setiap objek ke dalam *cluster* berdasarkan pada jarak minimumnya. Suatu data akan menjadi anggota dari suatu *cluster* (C1, C2 atau C3) yang memiliki jarak atau nilai terkecil dari pusat *cluster*-nya, misalnya untuk data pertama, jarak terkecil ada pada *cluster* yang ke-3 yaitu 6,164414003, maka keluhan pertama akan menjadi anggota pada *cluster* 2. Berikut adalah hasil iterasi 1 :

Tabel 4.6 Hasil Pengelompokan Iterasi 1

Keluhan ke	Cluster			Anggota Cluster
	C1	C2	C3	
1	24,08318916	10,14889157	6,164414003	3
2	18,08314132	12	7,141428429	3
3	23,74868417	10,72380529	6,164414003	3
4	25,11971337	9,273618495	9,746794345	2
5	26,07680962	9,327379053	10,39230485	2
6	20,51828453	15,09966887	8,185352772	3
7	20,29778313	11,44552314	9,899494937	3
8	25,61249695	12,60952021	6,633249581	3
9	23,38803113	20	13,60147051	3
10	17,74823935	10,29563014	6,8556546	3
11	27,92848009	13,96424004	10,39230485	3
12	18,86796226	25,35744467	19,89974874	1
13	28,35489376	12,92284798	10,77032961	3
14	21,16601049	12,60952021	5,830951895	3
15	25	14,4222051	7,141428429	3
16	22,82542442	16	11,87434209	3
17	33,51119216	22,27105745	20,174241	3
18	19,57038579	39,11521443	35,98610843	1

Tabel 4.6 Hasil Pengelompokan Iterasi 1 (lanjutan)

Keluhan Ke	Cluster			Anggota Cluster
	C1	C2	C3	
19	28,74021573	16,52271164	11,91637529	3
20	10,77032961	27,58622845	21,58703314	1
21	26,13426869	12,40967365	9,539392014	3
22	26	13,15294644	12,80624847	3
23	21,84032967	12,80624847	1	3

4.3.4 Proses Iterasi Kedua

Proses iterasi kedua adalah tahapan yang dilakukan untuk mendapatkan kelompok data keluhan pasien berdasarkan jarak terdekat data dengan pusat klaster yang baru. Pada tahapan iterasi kedua terdiri dari beberapa tahapan yaitu menentukan *centroid* kedua, menghitung jarak, dan mengelompokkan data.

1. Tentukan Pusat Cluster Kedua

Menentukan pusat *cluster* baru adalah tahapan berikutnya dalam iterasi kedua. Pusat *cluster* yang baru ditentukan berdasarkan pengelompokan anggota masing-masing *cluster*. rata-rata yang didapat dari semua kelompok *cluster* pada iterasi pertama yang akan menjadi pusat *centroid* yang baru Berdasarkan Tabel 4.6 maka prosesnya sebagai berikut :

C1 beranggotakan 8 data maka pusat cluster baru menjadi :

$$C1 = (12+21+17+18+10+13+16+14)/8 = 15,125$$

$$= (3+4+7+5+2+2+2+14)/8 = 4,875$$

$$= (4+17+8+15+16+11+11+17)/8 = 12,375$$

$$= (20+17+9+7+7+12+11+19)/8 = 12,75$$

$$= (9+21+15+8+7+10+8+13)/8 = 11,375$$

$$= (14+29+22+19+10+8+13+17)/8 = 16,5$$

Pusat C1 yang baru adalah (15,125; 4,875; 12,375; 12,75; 11,375; 16,5). Perhitungan ini dilakukan juga pada penentuan pusat C2 dan C3 yang baru. Berikut adalah hasil penentuan pusat baru pada Tabel 4.7 :

Tabel 4.7 Hasil Penentuan Pusat *Centroid* Baru

Titik Pusat Baru	Februari	Maret	April	Mei	Juni	Juli
Cluster 1	15,125	4,875	12,375	12,75	11,375	16,5
Cluster 2	4,8	5,1	10,2	4,3	4,2	3,6
Cluster 3	4,40740	6,9629	4,1666	5,1296	5,4814	6,24074

2. Menghitung Jarak Iterasi Kedua

Melakukan penentuan pusat *centroid* yang baru, maka proses selanjutnya adalah melakukan iterasi 2. Iterasi dua langkahnya sama dengan iterasi 1. Iterasi dua tentunya dihitung dengan menggunakan hasil dari penentuan *centroid* yang baru. Hasil dari iterasi 2 dapat dilihat pada Tabel 4.7 :

Tabel 4.8 Hasil Iterasi 2

Keluhan ke	<i>Cluster</i>		
	C1	C2	C3
1	23,04614935	7,441773982	5,485826869
2	15,1616292	8,6127812	6,620803307
3	22,65226258	6,969935437	4,314257344
4	23,53454907	7,181921748	6,199002858
5	24,10134851	6,69178601	6,926988988
6	16,4658738	12,21392648	9,310580886
7	17,95480437	8,436824047	7,222767921
8	23,57700151	10,33344086	6,558982881
9	19,87146195	18,52511808	15,03150812
10	15,08724627	6,03158354	5,625816957
11	24,85709154	11,7889779	8,017451991
12	12,16038651	21,73430468	19,15970398
13	26,464599	11,95742447	9,060593603
14	18,2174367	8,232860985	3,714568136
15	22,71838463	10,62920505	4,43699633
16	21,6304184	11,79745735	9,776286434
17	30,80787237	22,43613157	19,05212157
18	17,98958032	37,39759351	36,72194979
19	26,07920628	15,36164054	11,14807501
20	9,199184746	25,19087136	22,9609533
21	24,32848125	11,16154111	9,206572458
22	22,46385986	9,164060236	9,612000647
23	19,80845779	8,875809822	3,85163036

3. Hasil Pengelompokan Iterasi Kedua

Pada tahapan ini caranya sama dengan pengelompokan iterasi pertama. *Clustering* objek dengan memasukkan setiap objek ke dalam *cluster* berdasarkan pada jarak minimumnya. Suatu data akan menjadi anggota dari suatu *cluster* (C1, C2 atau C3) yang memiliki jarak atau nilai terkecil dari pusat *cluster*-nya. Berikut adalah hasil pengelompokan iterasi kedua :

Tabel 4.9 Hasil Pengelompokan Iterasi 2

Keluhan ke	Cluster			Anggota Cluster
	C1	C2	C3	
1	23,04614935	7,441773982	5,485826869	3
2	15,1616292	8,6127812	6,620803307	3
3	22,65226258	6,969935437	4,314257344	3
4	23,53454907	7,181921748	6,199002858	3
5	24,10134851	6,69178601	6,926988988	2
6	16,4658738	12,21392648	9,310580886	3
7	17,95480437	8,436824047	7,222767921	3
8	23,57700151	10,33344086	6,558982881	3
9	19,87146195	18,52511808	15,03150812	3
10	15,08724627	6,03158354	5,625816957	3
11	24,85709154	11,7889779	8,017451991	3
12	12,16038651	21,73430468	19,15970398	1
13	26,464599	11,95742447	9,060593603	3
14	18,2174367	8,232860985	3,714568136	3
15	22,71838463	10,62920505	4,43699633	3
16	21,6304184	11,79745735	9,776286434	3

Tabel 4.9 Hasil Pengelompokan Iterasi 2 (lanjutan)

Keluhan Ke	Cluster			Anggota Cluster
	C1	C2	C3	
17	30,80787237	22,43613157	19,05212157	3
18	17,98958032	37,39759351	36,72194979	1
19	26,07920628	15,36164054	11,14807501	3
20	9,199184746	25,19087136	22,9609533	1
21	24,32848125	11,16154111	9,206572458	3
22	22,46385986	9,164060236	9,612000647	2
23	19,80845779	8,875809822	3,85163036	3

4.3.5 Hasil K-Means Clustering

Proses iterasi *K-means* berhenti ketika hasil iterasi tidak berubah dengan hasil iterasi sebelumnya. Proses iterasi akan terus dilakukan sampai iterasi tidak berubah dari nilai iterasi sebelumnya. Berikut adalah hasil dari proses algoritma *K-means* yang telah selesai pada Tabel 4.10 :

Tabel 4.10 Hasil Clustering Data Keluhan Pasien

No	Keluhan	Februari	Maret	April	Mei	Juni	Juli	Hasil
1	Amandel	6	8	4	2	3	3	C2
2	asam lambung	5	7	8	6	10	9	C1
3	Asam urat	4	6	4	2	3	5	C3
4	BAB berdarah	2	9	6	5	1	4	C3
5	BAB tidak lancar	2	7	6	6	1	2	C3
6	badan dingin	6	9	6	7	14	7	C3
7	Badan Pegal	1	7	9	6	7	10	C3
8	BAK tidak lancar	6	11	2	4	5	2	C2
9	Batuk	6	16	5	6	17	9	C2
10	batuk berdahak	7	5	8	7	7	7	C3
11	Batuk berdarah	2	12	2	8	4	2	C2

Tabel 4.10 Hasil *Clustering* Data Keluhan Pasien (lanjutan)

No	Keluhan	Februari	Maret	April	Mei	Juni	Juli	Hasil
12	batuk kering	12	3	4	20	9	14	C1
13	BB turun	2	14	3	4	3	2	C2
14	Bersin-bersin	3	6	5	6	8	8	C3
15	Bibir bengkak	2	8	1	5	4	7	C2
16	Bintik merah di kulit	0	1	5	3	4	12	C3
17	Bisul	0	25	2	8	6	4	C2
18	demam	21	4	17	17	21	29	C1
19	Diare	3	17	1	6	6	3	C2
20	Gatal-gatal	17	7	8	9	15	22	C1
21	Insomnia	8	12	3	5	3	0	C2
22	kaki bengkak	2	2	6	11	3	2	C3
23	kaki gatal	6	7	3	3	8	6	C3

Berdasarkan hasil *cluster* dari data keluhan pasien kita dapat menyimpulkan bahwa, C1 merupakan keluhan penyakit yang sering terjadi dengan intensitas tinggi sehingga pihak puskesmas memberikan perhatian khusus pada keluhan tersebut yaitu dengan cara memberikan penyuluhan luar ruangan, penyuluhan dalam ruangan. C2 merupakan keluhan yang terjadi dengan intensitas sedang sehingga tindakan yang dapat dilakukan puskesmas yaitu dengan cara penyuluhan dalam ruangan, C3 merupakan keluhan yang terjadi dengan intensitas rendah sehingga puskesmas hanya melakukan arahan atau peringatan pada saat pasien melakukan pengobatan.

4.3 Perancangan

Perancangan pada penelitian ini untuk memperlihatkan sistem berbasis web dengan menggunakan bahasa pemrograman PHP yang akan dibangun. Sistem yang akan dibangun berdasarakan hasil analisa sistem yang telah dilakukan sebelumnya. Pada

perancangan sistem ini meliputi perancangan tatap muka dan perancangan tabel pada *data base*

4.4.1 Desain Output

Desain *output* merupakan tampilan hasil pengolahan data dari data keluhan. Hasil dari output sistem ini didapatkan dari algoritma *K-means* yang telah dijelaskan sebelumnya. Perancangan desain output terdiri dari halaman proses *cluster K-means* dan Halaman *Output Cluster*.

1. Perancangan Halaman Proses *cluster K-means*

Halaman ini berfungsi untuk user dalam melakukan proses *clustering*. Proses *cluster* berjalan sesuai dengan algoritma *K-means* yang telah dijelaskan sebelumnya. Pengguna dapat memulai proses dengan cara menekan tombol mulai iterasi, tombol terus ditekan sehingga jika proses iterasi telah selesai maka hasil *cluster* akan di tampilkan. Berikut adalah tampilan dari perancangan halaman proses *cluster K-means* :







Puskesmas Kampung Baru Padusunan									
Admin > Iterasi K-Means									
Iterasi									
Keluhan Pasien									
Keluhan	1	2	3	4	5	6	C1	C2	C3
xxxxx Z xxxxx	99 Z 99	99 Z 99	99 Z 99	99 Z 99	99 Z 99	99 Z 99	999999 Z 999999	999999 Z 999999	999999 Z 999999

Gambar 4.4 Halaman Iterasi K-Means

Pada Gambar 4.4 dapat kita lihat, halaman ini berfungsi untuk melakukan proses iterasi *K-Means* dari data yang ada. Pada halaman ini lah inti algoritma *K-Means* di jalankan. Pengguna dapat melakukan proses iterasi dengan cara menekan tombol iterasi, tombol iterasi dapat ditekan terus menerus sampai proses iterasi selesai dan mengeluarkan hasil *cluster*.

2. Perancangan Halaman *Output Cluster*

Halaman ini berisikan hasil dari iterasi yang telah dilakukan dan menghasilkan *cluster* dari data keluhan pasien. Halaman ini akan muncul ketika proses iterasi telah selesai. Berikut adalah tampilan halaman *output cluster*

Puskesmas Kampung Baru Padusunan					
 Main Navigation <ul style="list-style-type: none"> Dashboard Data Iterasi K-Means 	Admin > Iterasi K-Means				
	Hasil				
	Keluhan	C1	C2	C3	Keterangan
xxxxxxxx  xxxxxxxx	999999  999999	999999  999999	999999  999999	xxxxxxxx  xxxxxxxx	

Gambar 4.5 Halaman *Output Cluster*

Pada Gambar 4.6 dapat kita lihat, halaman ini berisikan hasil dari algoritma *K-means* yang telah diterapkan. Masing-masing *cluster* mempunyai keterangan masing-masing sesuai dengan *cluster* yang dihasilkan. Keterangan berisikan tindakan atau pelayanan yang seharusnya diberikan oleh pihak puskesmas sesuai dengan hasil *cluster*.

4.4.2 Perancangan Halaman input Data Keluhan Puskesmas

Halaman input data puskesmas merupakan halaman yang berisikan form yang dapat digunakan bagi pengguna untuk menginputkan data yang telah ditentukan pada form yang tersedia. Data yang akan diinputkan seperti data nama keluhan pasien dan jumlah data keluhan pasien perbulannya. Berikut adalah tampilan dari perancangan halaman input data puskesmas :

Puskesmas Kampung Baru Padusunan	
<div style="text-align: center;">○</div> <p>Main Navigation</p> <ul style="list-style-type: none"> Dashboard Data Iterasi K-Means 	<p>Admin > Data > Add</p> <p><input type="text" value="Backspace"/></p> <p>Keluhan <input type="text"/></p> <p>Bulan <input type="text" value="▼"/></p> <p>Jumlah <input type="text"/></p> <p><input type="button" value="Save"/></p>

Gambar 4.6 Halaman Input Data Keluhan

Pada Gambar 4.6 dapat kita lihat pengguna dapat menginputkan data keluhan pasien pada form yang telah disediakan. Data yang diinputkan seperti keluhan, bulan, dan jumlah, lalu pengguna dapat menyimpan data dengan menekan tombol save. Data pada form ini adalah sebagai data yang akan diolah pada saat proses iterasi *K-means*.

4.4.3 Perancangan File

Perancangan tabel mendeskripsikan dan menjelaskan tentang tabel yang ada pada *database* dari sistem yang akan dibuat. Adapun pada sistem ini *database* diberi nama *K-means*. Adapun tabel yang dibutuhkan dalam membangun sistem yaitu :

1. Tabel Input Data Keluhan

Tabel 4.11 Tabel Data Input

Nama Tabel : Data

Primary Key : Id_keluhan

Deskripsi : Berisi data keluhan pasien yang akan diolah

No	Nama Atribut	Type	Size
1	ID_keluhan	Int	11
2	Keluhan	Varchar	30
3	Bulan	Int	11
4	Tahun	Int	11
5	Jumlah	Int	11

2. Tabel Data Titik Pusat *Cluster***Tabel 4.12 Tabel Data Titik Pusat *Cluster***

Nama Tabel : Centroid

Primary Key : Id_Centroid

Deskripsi : Berisi data centroid awal dan *centroid* untuk iterasi berikutnya

No	Nama Atribut	Type	Size
1	ID_Centroid	Int	11
2	Iterasi	Int	11
3	Cluster	Int	11
4	Centroid 1	Int	11
5	Centroid 2	Int	11

6	Centroid 3	Int	11
---	------------	-----	----

3. Tabel Data Hasil Iterasi

Tabel 4.13 Tabel Data Hasil Iterasi

Nama Tabel : Hasil

Primary Key : Id_Hasil

Deskripsi : Berisi data hasil *K-means*

No	Nama Atribut	Type	Size
1	ID_hasil	Int	11
2	Id_keluhan	Int	11
3	iterasi	Int	11
4	Ketrangan	Varchar	100

BAB V

IMPLEMENTASI DAN HASIL

5.1 Implementasi

Pada bab ini menjelaskan pengujian dari rancangan sebelumnya, dalam melakukan pengujian algoritma *k-means* dari hasil manual ke komputerisasi maka dibutuhkan suatu sistem aplikasi. Pada penelitian ini dibentuk sebuah sistem aplikasi berbasis web dengan menggunakan bahasa pemrograman php *mysql*, dimana sistem ini digunakan untuk menganalisa hasil dari proses algoritma *k-means*. Sebelum membangun sebuah sistem aplikasi berbasis web maka terlebih dahulu dilakukan pengujian algoritma *k-means* dengan menggunakan *software* weka untuk mendapatkan jumlah *cluster* yang optimal

Berdasarkan tujuan diatas oleh karena itu pada bab ini akan dijelaskan tahapan-tahapan dalam melakukan analisis *k-means* dengan menggunakan *software* weka, dan bagaimana cara kerja dari sebuah sistem aplikasi berbasis web yang dibangun sehingga menghasilkan cluster yang sama dengan hasil pencarian secara manual dan hasil pencarian dengan menggunakan *software* weka yang sebelumnya telah dilakukan sehingga pengujian dapat tercapai. Tahapan pengujian untuk menentukan hasil pengelompokan *k-means* membutuhkan perangkat lunak sebagai berikut :

- a. *Operation System* Windows 10
- b. *Microsoft office excel* 2016 sebagai pendukung pengolahan data penelitian
- c. *Weka* merupakan tool untuk pengujian metode *k-means*

5.2 Pengujian *K-Means* dengan menggunakan Weka

Weka adalah salah satu platform perangkat lunak data *mining* dimana aplikasi ini bersifat *open source*. Weka merupakan sebuah *software* yang menerapkan berbagai algoritma *machine learning* untuk melakukan beberapa proses yang berkaitan dengan sistem temu kembali informasi atau *data mining*. *Software* ini bisa digunakan dalam pengembangan aplikasi, penelitian, pendidikan, pelatihan, pembuatan prototipe. *Software* ini juga bisa membantu proses *machine learning* termasuk data pra-proses, visualisasi hasil, validasi model, dan optimalisasi.

Salah satu fitur yang ada pada *software* ini adalah pengelompokan dengan menggunakan algoritma *k-means*. Kemudahan penggunaannya menjadi alasan peneliti untuk menggunakan weka. Selain itu *software* ini fleksibilitas dan skalabilitas membuat weka dapat diakses oleh peneliti dengan semua tingkat keterampilan dan proyek berbagai ukuran dan kompleksitas untuk membantu peneliti menemukan peluang baru, meningkatkan efisiensi, dan meminimalkan resiko.

5.2.1 Tahapan *K-Means* pada Weka

Secara umum tahapan pengelompokan data dengan menggunakan metode *k-means* pada *software* weka dimana langkah awal membuka *software* dari *icon* weka yang tampil pada *desktop windows*, lalu menginput data keluhan pasien yang telah di praproses sebelumnya dan akan di analisis, dan terakhir adalah melakukan proses *data mining* dalam hal ini adalah proses *k-means*. *Software* ini dapat mengenali berbagai macam ekstensi file yang akan menjadi input data, adapun format yang dapat dikenal oleh weka yaitu CSV file dan arff. Pada penelitian ini ekstensi file data yang digunakan dalam bentuk file excel dengan ekstensi *.csv*.

1. Langkah 1 – Membuka *Software* Weka

Langkah awal untuk membuka *software* weka dapat dilakukan dengan klik *icon* weka yang telah terinstal sebelumnya. Kemudian akan muncul tampilan awal menu home dari *software* weka seperti Gambar 5.1 :

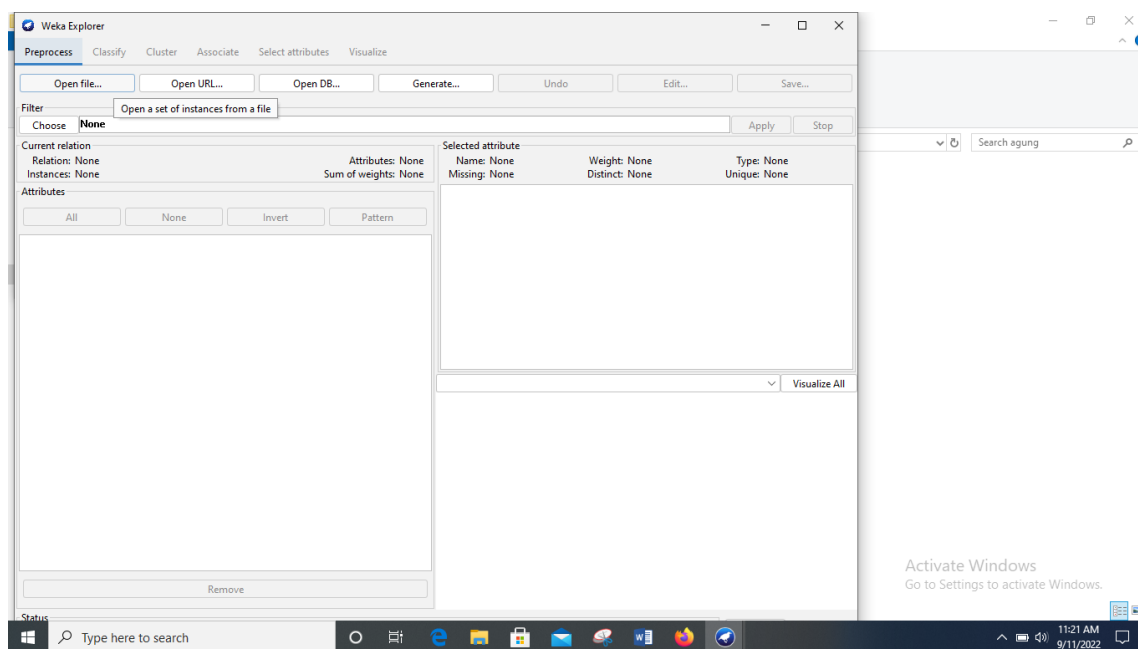


Gambar 5.1 Tampilan Awal Menu Home

Pada tampilan awal yang terdapat pada Gambar 5.1 dapat kita lihat berisikan menu bar dan beberapa tombol pada bagian *applications*. Masing-masing tombol dapat digunakan oleh *user* sesuai dengan kebutuhan. Pada penelitian ini tombol yang digunakan yaitu tombol *explores* untuk membuat proses baru dan *tools* yang digunakan untuk menampilkan file yang berekstensi *csv* dan *arff*.

2. Langkah 2 - Membuat proses baru

Langkah selanjutnya yaitu membuat proses baru dengan cara klik tombol explorer yang ada pada tampilan awal pada Gambar 5.1. kemudian akan tampil tampilan explorer dari *software* weka, seperti pada Gambar 5.2 :

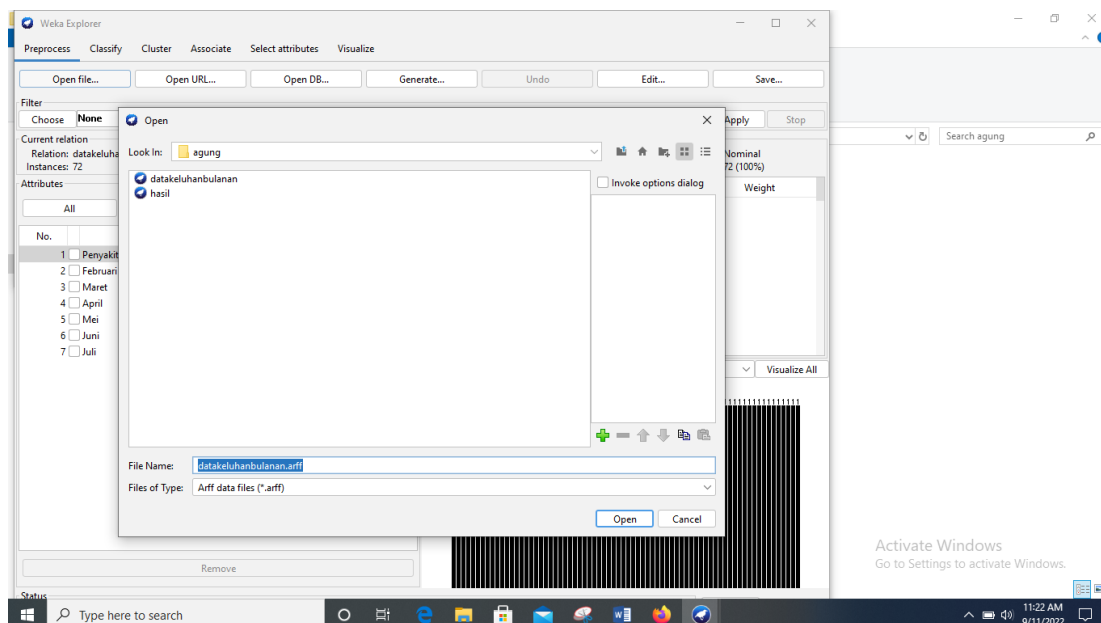


Gambar 5.2 Tampilan Explorer

Pada tampilan explorer yang terdapat pada Gambar 5.2 merupakan tampilan dari langkah selanjutnya dalam menggunakan weka. Pada tampilan ini berisikan *tools* yang digunakan untuk melakuakn analisa data *mining*. Dimana pada langkah berikutnya harus menginputkan data yang akan dianalisa.

3. Langkah 3 – Menginput Data Keluhan

Pada proses analisa data mining langkah terpenting yang dilakukan sebelumnya adalah menginputkan data yang akan dianalisa. Input data pada penelitian ini adalah data keluhan pasien yang telah dibuat pada *microsoft excel* yang berekstensi csv. Pada *software* weka input data dapat dilihat pada Gambar 5.3 :

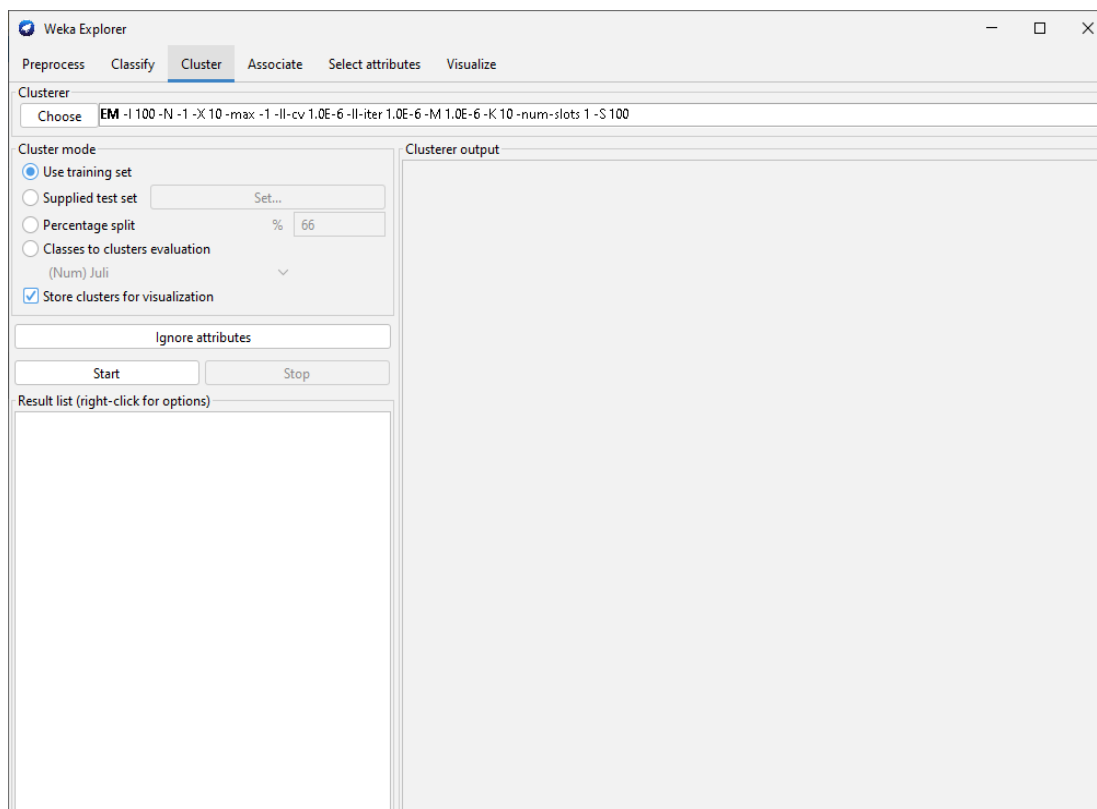


Gambar 5.3 Input Data

Pada tampilan input data Gambar 5.3 setelah menekan tombol open file setelah itu tampil jendela file tempat data yang akan diinputkan. Atur *files of type* menjadi csv agar file csv yang telah dibuat dapat terbaca. Klik open setelah memilih file yang akan diinputkan.

4. Langkah 4 – Memlih Bar Cluster

Pada langkah selanjutnya setelah melakukan input data yaitu memilih menu bar *cluster*. tampilan cluster ini merupakan tampilan yang akan digunakan untuk melakukan analisa *k-means clustering*. Tampilan cluster dapat dilihat pada gamabar 5.4 :

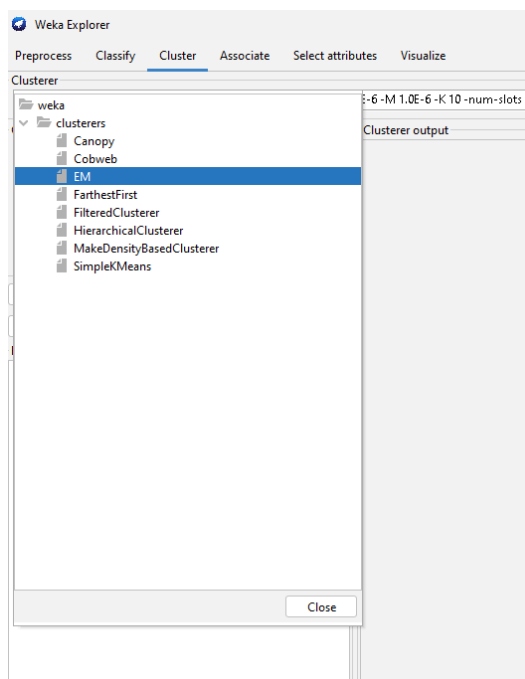


Gambar 5.4 Tampilan Cluster

Pada Gambar 5.4 merupakan tampilan *cluster* dimana pada halaman ini tempat proses dari *clustering* dari data yang telah diinputkan. Pada tampilan ini berisikan tombol *konfigurasi* dari clustering yang akan dilakukan sesuai dengan kebutuhan penelitian yang dilakukan. Proses selanjutnya terjadi pada tampilan ini.

5. Langkah 5 – Memilih Metode *K-means Clustering*

Langkah selanjutnya yang dilaksanakan yaitu memilih metode algoritma yang akan dipakaia. Pada penelitian ini kita akan menggunakan metode *k-means clustering*. Klik tombol *chose* untuk memilih algoritma yang dipakai. Setelah itu akan tampil pilihan yang dapat kita lihat pada Gambar 5.5 :

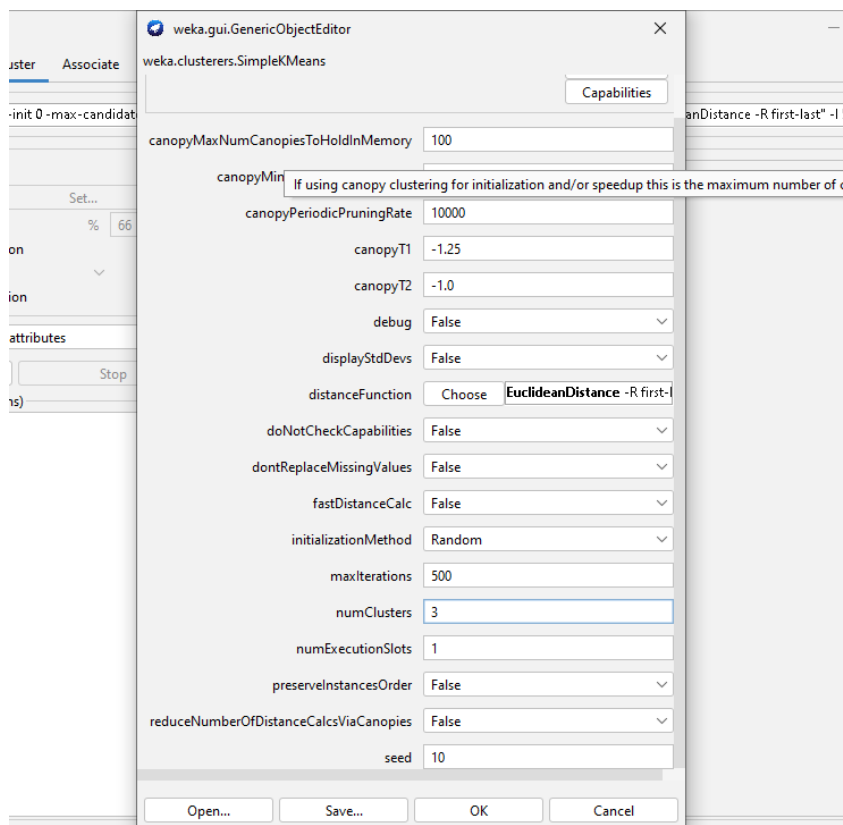


Gambar 5.5 Tampilan Pilihan Metode Cluster

pada Gambar 5.5 merupakan tampilan dari metode-metode yang dapat digunakan. Pada penelitian ini memakai metode *k-means clustering* oleh karena itu pilihan yang dipilih adalah *simplekmeans*. Setelah itu *klik* close untuk menutup jendela pilihan metode *cluster*.

6. Langkah 6 – Menentukan Jumlah *cluster*

Setelah melakukan pemilihan metode *cluster* langkah selanjutnya yang dilakukan yaitu memilih jumlah cluster, dengan cara klik kolom yang berada di sebelah tombol *chose*. Mekan akan tampil jendela *generic object editor* seperti pada Gambar 5.6 :

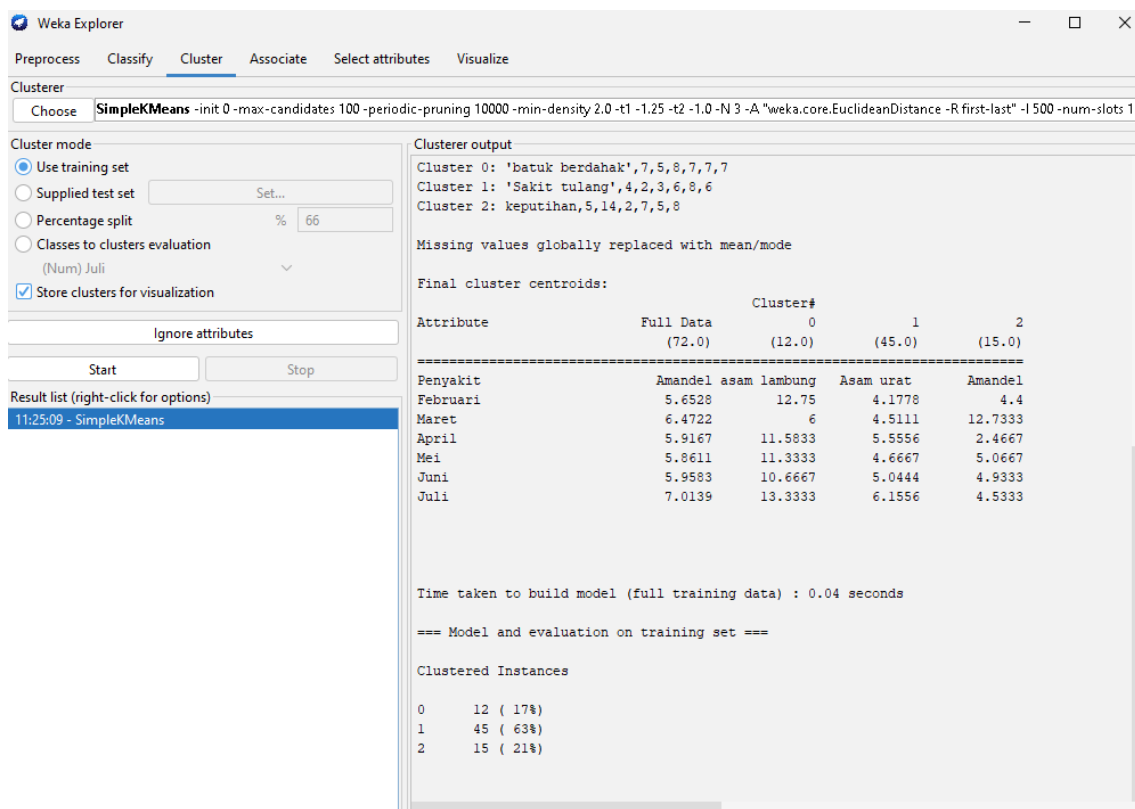


Gambar 5.6 Tampilan *Generic Object Editor*

Pada Gambar 5.6 dapat kita lihat dimana berisikan form yang berisikan setingan yang dapat disesuaikan dengan kebutuhan pengguna dalam melakukan suatu analisis data. Pada penelitian ini bertujuan untuk membentuk 3 *cluster*, maka bagian yang akan diset adalah *numclusters* dirubah menjadi 3, lalu klik ok untuk menyimpan.

7. Langkah 7 – Proses Algoritma *K-Means*

Tahapan ini adalah tahapan inti dalam melakukan analisis data dengan menggunakan *software* weka. Setelah melakukan setingan jumlah *cluster* maka langkah selanjutnya kita dapat memulai proses *k-means*. Memulai proses *k-means* dengan cara klik tombol start seperti pada Gambar 5.7 :

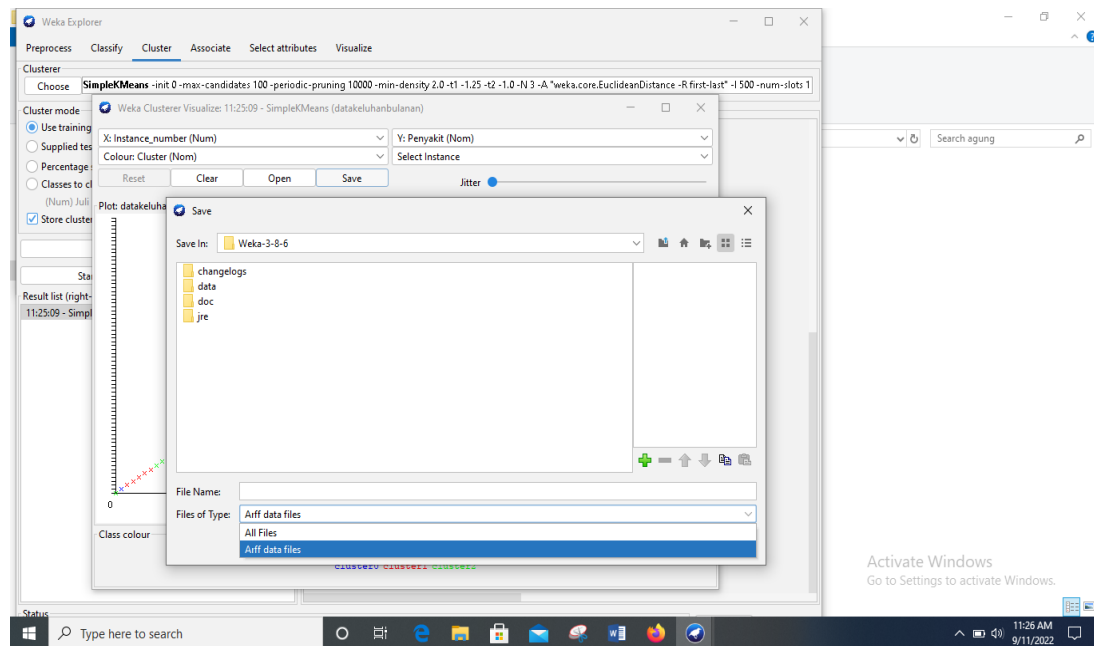


Gambar 5.7 Proses Algoritma *K-means Clustering*

Berdasarkan Gambar 5.7 dapat kita lihat, setelah memulai algoritma *clustering* dengan cara klik tombol start, maka setelah itu sistem melakukan proses *clustering*, dan menampilkan hasil dari proses *clustering* tersebut. Pada tampilan ini akan menghasilkan data yang telah diklaster yang dapat kita lihat pada bagian result list. Tampilan ini juga berisikan rincian seperti persentase dari masing-masing *cluster*, centroid awal yang digunakan, dan hasil jarak objek terhadap pusat data.

8. Langkah – 8 Simpan Hasil *Cluster*

Proses dari algoritma *k-means* telah selesai, setelah itu maka kita perlu menyimpan hasil analisa yang telah dilakukan. Pada tahapan ini kita dapat menyimpan hasil analisa dengan cara klik kanan pada hasil analisa yang terdapat pada bagian *list result* dan pilih *Visualize cluster assignments*, setelah itu akan tampil jendela simpan pada Gambar 5.8 :



Gambar 5.8 Simpan Hasil *Cluster*

Berdasarkan Gambar 5.8 dimana pada tampilan ini merupakan jendela simpan dari hasil *clustering* yang telah dilakukan. Pengguna dapat menentukan tempat penyimpanan yang diinginkan pada bagian *save to*, memilih ekstensi arff untuk format file, dan klik ok untuk menyimpan.

9. Langkah 9 - Menampilkan hasil Anlisa

Pada tahapan sebelumnya file analisa telah disimpan dengan menggunakan ekstensi arff. Pada tahapan ini kita dapat menampilkan hasil analisa dengan cara silahkan tutup tampilan yang sedang berjalan sampai pada tampilan seperti pada Gambar 5.1, lalu pilih tools yang terdapat pada menu bar, pilih arff viewer, dan silahkan buka file yang telah disimpan sebelumnya maka hasil dari clustering akan tampil seperti pada Gambar 5.9 :

No.	1: Instance_number Numeric	2: Februari Numeric	3: Maret Numeric	4: April Numeric	5: Mei Numeric	6: Juni Numeric	7: Juli Numeric	8: Cluster Nominal
1	0.0	6.0	8.0	4.0	2.0	3.0	3.0	cluster1
2	1.0	5.0	7.0	8.0	6.0	10.0	9.0	cluster1
3	2.0	4.0	6.0	4.0	2.0	3.0	5.0	cluster1
4	3.0	2.0	9.0	6.0	5.0	1.0	4.0	cluster1
5	4.0	2.0	7.0	6.0	6.0	1.0	2.0	cluster1
6	5.0	6.0	9.0	6.0	7.0	14.0	7.0	cluster1
7	6.0	1.0	7.0	9.0	6.0	7.0	10.0	cluster1
8	7.0	6.0	11.0	2.0	4.0	5.0	2.0	cluster2
9	8.0	6.0	16.0	5.0	6.0	17.0	9.0	cluster2
10	9.0	7.0	5.0	8.0	7.0	7.0	7.0	cluster1
11	10.0	2.0	12.0	2.0	8.0	4.0	2.0	cluster2
12	11.0	12.0	3.0	4.0	20.0	9.0	14.0	cluster0
13	12.0	2.0	14.0	3.0	4.0	3.0	2.0	cluster2
14	13.0	3.0	6.0	5.0	6.0	8.0	8.0	cluster1
15	14.0	2.0	8.0	1.0	5.0	4.0	7.0	cluster2
16	15.0	0.0	1.0	5.0	3.0	4.0	12.0	cluster1
17	16.0	0.0	25.0	2.0	8.0	6.0	4.0	cluster2
18	17.0	21.0	4.0	17.0	17.0	21.0	29.0	cluster0
19	18.0	3.0	17.0	1.0	6.0	6.0	3.0	cluster2
20	19.0	17.0	7.0	8.0	9.0	15.0	22.0	cluster0
21	20.0	8.0	12.0	3.0	5.0	3.0	0.0	cluster2
22	21.0	2.0	2.0	6.0	11.0	3.0	2.0	cluster1
23	22.0	6.0	7.0	3.0	3.0	8.0	6.0	cluster1
24	23.0	5.0	14.0	2.0	7.0	5.0	8.0	cluster2
25	24.0	3.0	2.0	5.0	5.0	11.0	7.0	cluster1
26	25.0	4.0	7.0	8.0	13.0	6.0	8.0	cluster1
27	26.0	1.0	5.0	2.0	1.0	2.0	6.0	cluster1
28	27.0	3.0	3.0	4.0	6.0	6.0	4.0	cluster1
29	28.0	0.0	4.0	3.0	8.0	7.0	11.0	cluster1
30	29.0	2.0	6.0	4.0	2.0	5.0	9.0	cluster1
31	30.0	3.0	3.0	6.0	2.0	7.0	5.0	cluster1

Gambar 5.9 Menampilkan Hasil Clustering

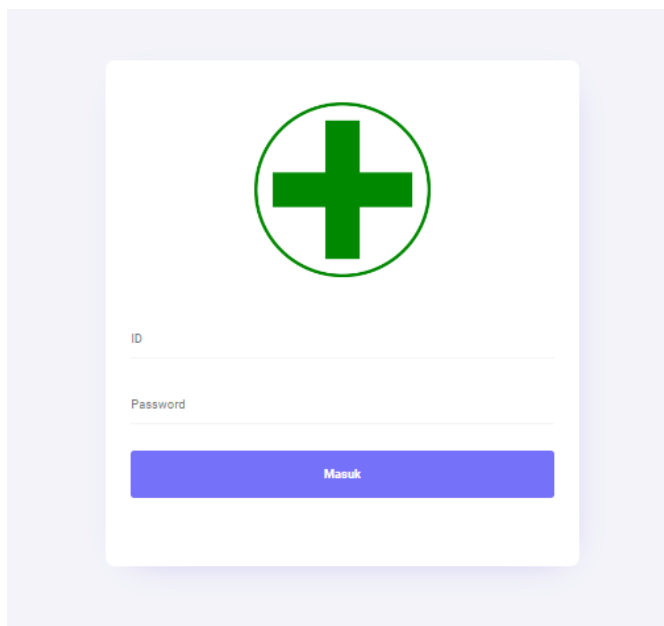
Berdasarkan Gambar 5.9 merupakan tampilan dari hasil clustering. Dimana file yang dibuka adalah yang berekstensi arff. Pada tampilan ini dapat kita lihat dimana berisikan rincian *cluster* dari masing-masing data. Dimana berisikan 3 *cluster* yang dimulai dari *cluster 0* sampai *cluster 2*. Urutan data masih sama dengan urutan data pada saat diinputkan sebelumnya.

5.3 Pengujian *K-means* dengan Menggunakan Web

Pada bagian ini menjelaskan tentang implementasi dari algoritma *k-means* dengan menggunakan sistem berbasis web yang dibangun. Dimana pada sistem ini dibuat dengan menggunakan *frontend css bootstrap* dan untuk backand menggunakan *php native*. Pada sistem ini telah diterapkan logika yang sesuai dengan algoritma dari metode *k-means* yang digunakan pada penelitian ini. Berikut adalah langkah-langkah dari implementasi *k-means* dengan menggunakan sistem berbasis web :

1. Langkah 1 – Akses Halaman Login Sistem

Pada tahapan awal hal yang dilakukan adalah mengakses halaman sistem login yang ada pada serve localhost dengan menggunakan browser. Pada sistem ini menggunakan *server* xampp dengan. Berikut adalah tampilan login dari sistem pada Gambar 5.10 :

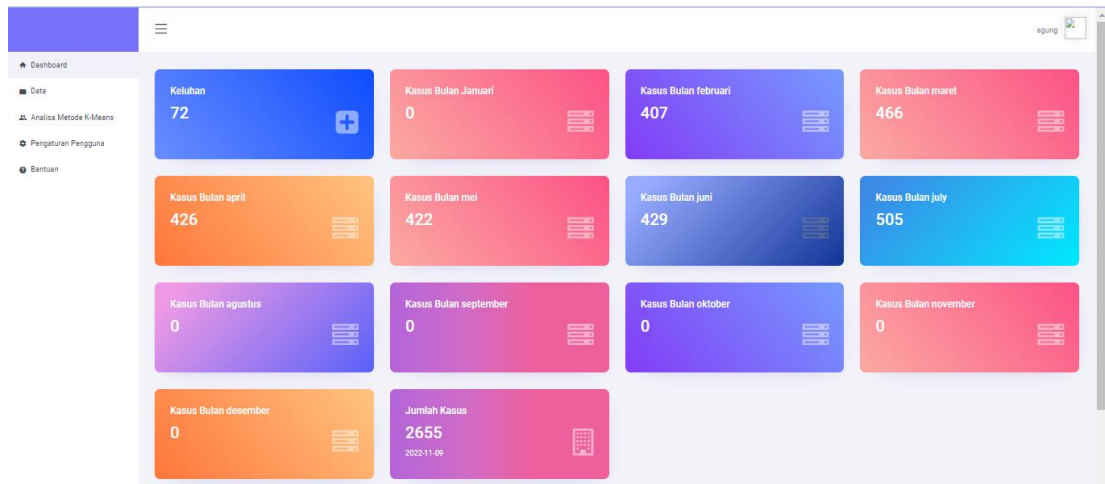


Gambar 5.10 Halaman Login

Pada Gambar 5.10 merupakan halaman login dari sistem web. Penggunan dapat login dengan cara menginputkan *id* dan *password*, dan menekan tombol masuk untuk mengeksekusi. Jika id dan password tidak sesuai maka pengguna tidak dapat masuk ke dalam sistem.

2. Langkah 2 – Masuk Menu Dashboard

Pada tahapan selanjutnya setelah melakukan login sistem akan mengarahkan langsung kehalaman dashboard. Pada halaman ini adalah halaman awal dari sistem. Halaman dashboard dapat dilihat pada Gambar 5.11 :



Gambar 5.11 Halaman Dashboard

Pada Gambar 5.11 dapat kita lihat pada halaman dashboard ini berisikan *card* yang berisikan data rekap dari keluhan yang ada. Pada halaman ini juga terdapa sidebar yang berisikan menu pada sistem ini. Card yang ada sesuai dengan jumlah kasus yang terjadi perbulannya.

3. Langkah 3 – Input Data

Langkah selanjutnya adalah menginputkan data yang akan dianalisa. Pada penelitian ini data yang digunakan adalah data keluhan pasien. Berikut halaman input data pada gambar 5.12 :

No	Keluhan	Persentase		Aksi
1	Amandel	<div style="width: 100%;"></div>	6.57102826911488%	<input type="text"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
2	Asam Lambung	<div style="width: 100%;"></div>	6.6846132642372%	<input type="text"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
3	Asam Urat	<div style="width: 100%;"></div>	6.9029548122289%	<input type="text"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
4	BAB Berdarah	<div style="width: 100%;"></div>	6.61664891325214%	<input type="text"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
5	BAB Tidak Lancar	<div style="width: 100%;"></div>	6.5690548122289%	<input type="text"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
6	Badan Dingin	<div style="width: 100%;"></div>	6.8403248122289%	<input type="text"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>

Gambar 5.12 Halaman Input Data

Berdasarkan gambar 5.12 dapat kita lihat pada halaman ini pengguna dapat melakukan input data. Data dapat diinput dengan cara klik tambah keluhan lalu isi form yang telah disediakan dan tekan tombol simpan untuk menyimpan, untuk jumlah kasus perbulan dapat diinputkan dengan cara klik tombol edit yang ada pada kolom aksi.

4. Langkah 4 – Melakukan Proses Metode K-means

Langkah selanjutnya setelah input data keluhan adalah melakukan pemrosesan analisa data dengan menggunakan metode k-means. Pada penelitian ini melakukan analisa dengan cara klik analisa metode *k-means*. Berikut adalah tampilan proses metode *k-means* pada Gambar 5.13 :

Klatisasi dengan K-Means
Jumlah Keluhan = 72

No	Keluhan	Januari	Februari	Maret	April	Mei	Juni	July	Agustus	September	Oktober	November	Desember
1	Amandel	0	6	8	4	2	3	3	0	0	0	0	0
2	Asam Lambung	0	5	7	8	6	10	9	0	0	0	0	0
3	Asam Urat	0	4	6	4	2	3	5	0	0	0	0	0
4	BAB Berdarah	0	2	9	6	5	1	4	0	0	0	0	0
5	BAB Tidak Lancar	0	2	7	6	6	1	2	0	0	0	0	0
6	Badan Dingin	0	6	9	6	7	14	7	0	0	0	0	0
7	Badan Pegal	0	1	7	9	6	7	10	0	0	0	0	0
8	BAK Tidak Lancar	0	6	11	2	4	5	2	0	0	0	0	0
9	Batuk	0	6	16	5	6	17	9	0	0	0	0	0
10	Batuk Berdahak	0	7	5	8	7	7	7	0	0	0	0	0
11	Batuk Berdarah	0	2	12	2	8	4	2	0	0	0	0	0
12	Batuk Kering	0	12	3	4	20	9	14	0	0	0	0	0
13	BB Turun	0	2	14	3	4	3	2	0	0	0	0	0
14	Bersin-bersin	0	3	6	5	6	8	8	0	0	0	0	0
15	Bibir Bengkak	0	2	8	1	5	4	7	0	0	0	0	0

Gambar 5.13 Halaman Proses K-means

Berdasarkan Gambar 5.13 merupakan halaman yang berfungsi sebagai tempat melakukan metode *k-means*. Berisikan data yang akan dianalisa dan tombol iterasi. Sesuai dengan langkah k-means, pengguna dapat memulai perhitungan dengan klik tombol iterasi. Tombol iterasi di klik terus menerus hingga hasil iterasi sama dengan hasil iterasi sebelumnya sama. Jika telah sama maka langsung mengarah pada halaman hasil *cluster*

5. Langkah 5 – Hasil *Clustering* dengan *K-means*

Pada tahapan sebelumnya yaitu proses metode *k-means*. Setelah melakukan proses tentunya menghasilkan sesuatu. Pada tahapan ini akan menghasilkan data-data yang telah di *cluster*. berikut adalah tampilan hasil *k-means* pada Gambar 5.14 :

Hasil dengan K-Means [Kembali](#)

Hasil C1(Keluhan dengan intensitas tinggi)

Jumlah Keluhan = 9
70.44444444444444

No	Keluhan	Kelompok	Tindakan
1	Batuk Kering	C 1	Penyuluhan dalam gedung, Penyuluhan luar gedung, Kunjungan rumah
2	Demam	C 1	Penyuluhan dalam gedung, Penyuluhan luar gedung, Kunjungan rumah
3	Gatal-gatal	C 1	Penyuluhan dalam gedung, Penyuluhan luar gedung, Kunjungan rumah
4	Mual	C 1	Penyuluhan dalam gedung, Penyuluhan luar gedung, Kunjungan rumah
5	Nafsu Makan Berkurang	C 1	Penyuluhan dalam gedung, Penyuluhan luar gedung, Kunjungan rumah
6	Nyeri Ulu Hati	C 1	Penyuluhan dalam gedung, Penyuluhan luar gedung, Kunjungan rumah
7	Pusing	C 1	Penyuluhan dalam gedung, Penyuluhan luar gedung, Kunjungan rumah
8	Sakit Kepala	C 1	Penyuluhan dalam gedung, Penyuluhan luar gedung, Kunjungan rumah
9	Sakit Perut	C 1	Penyuluhan dalam gedung, Penyuluhan luar gedung, Kunjungan rumah

Hasil C2(Keluhan intensitas sedang)

Jumlah Keluhan = 15
38.4

No	Keluhan	Kelompok	Tindakan
----	---------	----------	----------

Gambar 5.14 Tampilan Hasil *K-Means*

Berdasarkan Gambar 5.14 menampilkan hasil dari cluster *k-means* yang telah dilakukan. Dimana menghasilkan 3 *cluster*. *Cluster 1* adalah keluhan yang terjadi dengan intensitas paling tinggi sebanyak 9 data keluhan. *Cluster 2* dengan intensitas sedang sebanyak 15 data keluhan. *Cluster 3* merupakan keluhan dengan intensitas paling rendah sebanyak 48 data.

BAB VI

KESIMPULAN DAN SARAN

6.1 Kesimpulan

Berdasarkan penelitian yang telah dilakukan dan pembahasan yang telah dipaparkan pada bab-bab sebelumnya mengenai penelitian mengenai *clustering* data keluhan pasien dengan menggunakan metode *k-means*. Dapatlah hasil dari penelitian ini dan dapat ditarik kesimpulan sebagai berikut ini :

1. Pada penelitian ini telah menghasilkan parameter dari masing-masing keluhan masyarakat pada tiap bulannya.
2. Penelitian ini menghasilkan 3 *cluster* dimana *Cluster 1* adalah keluhan yang terjadi dengan intensitas paling tinggi sebanyak 9 data keluhan. *Cluster 2* dengan intensitas sedang sebanyak 15 data keluhan. *Cluster 3* merupakan keluhan dengan intensitas paling rendah sebanyak 48 data.
3. *K-means* dapat membantu pihak puskesmas dalam melakukan pencegahan dan pemberian pelayanan masyarakat dengan efektif karena pada penelitian ini telah menghasilkan model dan pertimbangan untuk menentukan penanganan atau penanggulangan yang tepat sasaran terhadap gejala penyakit masyarakat.

6.2 Saran

Berdasarkan penelitian yang telah dilakukan dalam pencarian metode yang paling tepat untuk melakukan evaluasi kinerja programmer, menemukan beberapa permasalahan yang dapat dijadikan sebagai bahan penelitian lebih lanjut agar dalam melakukan evaluasi pemrograman menjadi lebih baik, diantaranya:

1. Proses *k-means* diantaranya yaitu menentukan *centroid* awal, setelah melakukan penelitian, peneliti menyadari bahwa *centroid* awal yang berbeda menghasilkan sedikit perbedaan pada hasil clustering.
2. Penelitian ini dapat dikembangkan dengan menghubungkan hasil pengelompokan dengan data-data lainnya seperti jenis kelamin, umur, dan obat-obatan. Sehingga sistem dapat menemukan keterkaitan antara satu data terhadap data lainnya. Seperti apakah atribut-atribut tertentu memiliki dampak terhadap kinerja programmer.

DAFTAR PUSTAKA

- Abduh, R. (2021).** Kajian Hukum Rekam Medis Sebagai Alat Bukti Malapraktik Medis. De Lega Lata: Jurnal Ilmu Hukum, 6(1), 221-234. DOI : <http://dx.doi.org/10.30596%2Fdll.v6i1.4661>
- Bagaskara, M. B., Dewi, A. S. L., & Suryani, L. P. (2022).** Tanggung Jawab Rumah Sakit Terhadap Kerahasiaan Rekam Medis (Medic Record) di Masa Pandemi Covid-19. Jurnal Analogi Hukum, 4(1), 26-30. DOI : <https://doi.org/10.22225/ah.4.1.2022.26-30>
- Bansal, M., Kumar, M., Kumar, M., & Kumar, K. (2021).** *An efficient technique for object recognition using Shi-Tomasi corner detection algorithm.* Soft Computing, 25(6), 4423-4432. DOI : <https://doi.org/10.1007/s00500-020-05453-y>
- Chowdhury, K., Chaudhuri, D., & Pal, A. K. (2021).** An entropy-based initialization method of K-means clustering on the optimal number of clusters. Neural Computing and Applications, 33(12), 6965-6982. DOI : <https://doi.org/10.1007/s00521-020-05471-9>
- Darma, S., Defit, S., Hartama, D., Robiansyah, W., & Firzada, F. (2020, July).** Penerapan Metode K-Means Dalam Pengolompokan Jumlah Wisatawan Asing Di Indonesia. In Prosiding Seminar Nasional Riset Information Science (SENARIS) (Vol. 2, pp. 255-261). DOI : <http://dx.doi.org/10.30645/senaris.v2i0.169>
- Dewi, S., Defit, S., & Yuhandri, Y. (2021).** Akurasi Pemetaan Kelompok Belajar Siswa Menuju Prestasi Menggunakan Metode K-Means. Jurnal Sistim Informasi dan Teknologi, 28-33. DOI : <https://doi.org/10.37034/jsisfotek.v3i1.40>
- Elda, Y., Defit, S., Yunus, Y., & Syaljumairi, R. (2021).** Klasterisasi Penempatan Siswa yang Optimal untuk Meningkatkan Nilai Rata-Rata Kelas Menggunakan K-Means. Jurnal Informasi Dan Teknologi, 103-108. DOI : <https://doi.org/10.37034/jidt.v3i3.130>
- Gokilavani, N., & Bharathi, B. (2021).** *Test case prioritization to examine software for fault detection using PCA extraction and K-means clustering with ranking.* Soft Computing, 25(7), 5163-5172. DOI : <https://doi.org/10.1007/s00500-020-05517-z>
- Hartati, T., Nurdiawan, O., & Wiyandi, E. (2021).** Analisis Dan Penerapan Algoritma K-Means Dalam Strategi Promosi Kampus Akademi Maritim Suaka Bahari. Jurnal Sains Teknologi Transportasi Maritim, 3(1), 1-7. DOI : <https://doi.org/10.51578/j.sitektransmar.v3i1.31>
- Haval, B., Abdulrahman, K. J., & Abraham, A. R. (2021).** [Student Performance Predictions Using Knowledge Discovery Database and Data Mining, DPU Students Records as](#)

[Sample. Academic Journal of Nawroz University, 10\(3\), 121-127. DOI :
https://doi.org/10.25007/ajnu.v10n3a875](#)

Indraputra, R. A., & Fitriana, R. (2020). K-Means Clustering Data COVID-19. *Jurnal Teknik Industri*, 10(3), 275-282. DOI : <https://doi.org/10.25105/jti.v10i3.8428>

Jaja, J., Priatna, N., & Ardan, T. S. (2021). Implementation of Data Mining Technique for Performance of WFH and WFO Agents Using the K-Means Method Case Study Study of PT. Infomedia Telkom Consumer Profiling Services. *Budapest International Research in Exact Sciences (BirEx) Journal*, 3(2), 117-125. DOI : <https://doi.org/10.33258/birex.v3i2.1810>

Khan, A. R., Khan, S., Harouni, M., Abbasi, R., Iqbal, S., & Mehmood, Z. (2021). Brain tumor segmentation using K-means clustering and deep learning with synthetic data augmentation for classification. *Microscopy Research and Technique*, 84(7), 1389-1399. DOI : <https://doi.org/10.1002/jemt.23694>

Lu, W. (2020). Improved K-means clustering algorithm for big data mining under Hadoop parallel framework. *Journal of Grid Computing*, 18(2), 239-250. DOI : <https://doi.org/10.1007/s10723-019-09503-0>

Malik, R. A., Defit, S., & Yuhandri, Y. (2018). Comparison of K-Means Clustering Algorithm with Fuzzy C-Means In Measuring Satisfaction Level Of Television Da'wah Surau TV. *Rabit: Jurnal Teknologi dan Sistem Informasi Univrab*, 3(1), 10-21. DOI : <https://doi.org/10.36341/rabit.v3i1.387>

Manochandar, S., Punniyamoorthy, M., & Jeyachitra, R. K. (2020). Development of new seed with modified validity measures for k-means clustering. *Computers & Industrial Engineering*, 141, 106290. DOI : <https://doi.org/10.1016/j.cie.2020.106290>

Molina-Coronado, B., Mori, U., Mendiburu, A., & Miguel-Alonso, J. (2020). Survey of network intrusion detection methods from the perspective of the knowledge discovery in databases process. *IEEE Transactions on Network and Service Management*, 17(4), 2451-2479. DOI : <https://doi.org/10.1109/TNSM.2020.3016246>

Muhammad, L. J., Islam, M., Usman, S. S., & Ayon, S. I. (2020). Predictive data mining models for novel coronavirus (COVID-19) infected patients' recovery. *SN Computer Science*, 1(4), 1-7. DOI : <https://doi.org/10.1007/s42979-020-00216-w>

[Nasution, L. S., Maya, W. R., Halim, J., & Marsono, M. \(2020\). Data Mining Untuk Menganalisa Pola Pembelian Perak Dengan Menggunakan Algoritma Fp-Growth Pada Toko Emas Dan Perak Adi Saputra Tanjung. Jurnal Teknologi Sistem Informasi dan Sistem Komputer TGD, 3\(2\), 96-107. DOI : https://doi.org/10.53513/jsk.v3i2.2039](#)

- Nugraha, U. (2021).** Classify Event Participants in Universities and Industries Using Knowledge Discovery in Databases. *Review of International Geographical Education Online*, 11(1), 526-542. DOI : [10.48047/rigeo.11.1.36](https://doi.org/10.48047/rigeo.11.1.36)
- Ordila, R., Wahyuni, R., Irawan, Y., & Sari, M. Y. (2020).** Penerapan Data Mining Untuk Pengelompokan Data Rekam Medis Pasien Berdasarkan Jenis Penyakit Dengan Algoritma Clustering (Studi Kasus: Poli Klinik Pt. Inecda). *Jurnal Ilmu Komputer*, 9(2), 148-153. DOI : <https://doi.org/10.33060/JIK/2020/Vol9.Iss2.181>
- Oulhiq, R., Benjelloun, K., Kali, Y., & Saad, M. (2022).** A data mining based approach for process identification using historical data. *International Journal of Modelling and Simulation*, 42(2), 335-349. DOI : <https://doi.org/10.1080/02286203.2021.1905375>
- Rahmadani, N., & Kurniawan, E. (2020).** Implementasi Metode K-Means Clustering Tunggalan Rekening Listrik pada PT. PLN (Persero) Gardu Induk Kisaran. *Jurnal Teknologi Sistem Informasi dan Sistem Komputer TGD*, 3(1), 103-117. DOI : <https://doi.org/10.53513/jsk.v3i1.201>
- Sinaga, K. P., & Yang, M. S. (2020).** Unsupervised K-means clustering algorithm. *IEEE access*, 8, 80716-80727. DOI : <https://doi.org/10.1109/ACCESS.2020.2988796>
- Singh, S., & Srivastava, S. (2020).** Review of Clustering Techniques in Control System: Review of Clustering Techniques in Control System. *Procedia Computer Science*, 173, 272-280. DOI : <https://doi.org/10.1016/j.procs.2020.06.032>
- Syahril, M., Erwansyah, K., & Yetri, M. (2020).** Penerapan Data Mining untuk menentukan pola penjualan peralatan sekolah pada brand wigglo dengan menggunakan algoritma apriori. *Jurnal Teknologi Sistem Informasi Dan Sistem Komputer TGD*, 3(1), 118-136. DOI : <https://doi.org/10.53513/jsk.v3i1.202>
- Virgo, I., Defit, S., & Yuhandri, Y. (2020).** Klasterisasi Tingkat Kehadiran Dosen Menggunakan Algoritma K-Means Clustering. *Jurnal Sistim Informasi dan Teknologi*, 23-28. DOI : <https://doi.org/10.37034/jsisfotek.v2i1.17>
- Zou, H. (2020).** Clustering algorithm and its application in data mining. *Wireless Personal Communications*, 110(1), 21-30. DOI : <https://doi.org/10.1007/s11277-019-06709-z>

LAMPIRAN

**PEMERINTAH KOTA PARIAMAN**
DINAS PENANAMAN MODAL PELAYANAN TERPADU
SATU PINTU DAN TENAGA KERJA
Alamat : Jln Syekh Burhanuddin No.145 Pariaman No.Telp : 0751 91529 Call Center :08116606609
Website : pisp.pariamankota.go.id Email: dpm@tsp.pariamankota@gmail.com

SURAT KETERANGAN PENELITIAN
Nomor: 241/SKP/DPMPPTSP&NAKER/VI/2022

Dasar : 1. Peraturan Menteri Dalam Negeri Nomor 3 Tahun 2018 Tentang Penerbitan Surat Keterangan Penelitian.
2. Keputusan Walikota Pariaman Nomor 188/570/2021 Tentang Perubahan Ketiga Atas Keputusan Walikota Pariaman Nomor 35/570/2017 Tentang Pendelegasian Kewenangan Pelayanan Perizinan Pada Dinas Penanaman Modal Pelayanan Terpadu Satu dan Tenaga Kerja Kota Pariaman.

Menimbang : Surat dari Universitas Putra Indonesia "YPTK" Padang Nomor : 037/M.KOM/UPI "YPTK"/A.6/VI/2022 tanggal 5 Juni 2022 Perihal Mohon Izin Penelitian.

Kepala Dinas Penanaman Modal Pelayanan Terpadu Satu Pintu dan Tenaga Kerja Kota Pariaman, menerima dan tidak keberatan atas kegiatan penelitian yang dilakukan oleh :

Nama : M. Agung Vafky Ideal
NIK : 1305061208980001
Alamat : Perumahan Griya Taluk Permai Blok K No. 3 Desa Taluk Kec. Pariaman Selatan Kota Pariaman
Judul Penelitian : "Klasifikasi Keluhan Pasien Terhadap Data Rekam Medis Pasien Dengan Menggunakan Metode K Means."
Bidang Penelitian : Bidang Kesehatan
Tujuan Penelitian : Penyelesaian Skripsi/Tesis
Lokasi Penelitian : Puskesmas Kampung Baru Padusunan
Waktu Penelitian : 22 Juni 2022 s/d 22 September 2022
Nama Lembaga : Universitas Putra Indonesia "YPTK" Padang
Penanggung Jawab : M. Agung Vafky Ideal
Anggota : -
Status Penelitian : Baru

Dengan ketentuan-ketentuan sebagai berikut :

1. Dalam pelaksanaan penelitian tidak boleh menyimpang dari kerangka tujuan penelitian, serta mematuhi peraturan dan adat istiadat yang berlaku di daerah setempat;
2. Memberitahukan kedatangan serta tujuan penelitian kepada pejabat instansi yang dijadikan objek/lokasi penelitian, serta melaporkan diri sebelum meninggalkan daerah Penelitian;
3. Menyampaikan laporan hasil penelitian kepada Dinas Penanaman Modal PTSP dan Tenaga Kerja Kota Pariaman;
4. Apabila terjadi penyimpangan/pelanggaran terhadap ketentuan di atas, maka surat keterangan ini tidak dicabut kembali.

Dikeluarkan di Pariaman
Pada tanggal 27 Juni 2022
Kepala Dinas,


Ditandatangani secara elektronik oleh
KOTADINAS PENANAMAN MODAL, PELAYANAN TERPADU SATU PINTU
KOTA PARIAMAN
NOVARDI, SH
Pemilihan Utama Maba, No. 1
NIK: 5495204 198903 1 001

Tembusan disampaikan kepada :

1. Bapak Walikota Pariaman (Sebagai Laporan);
2. Kepala Kantor Kesbangpol dan Litmas Kota Pariaman;
3. Kepala Lembaga / Instansi Terkait dengan Lokasi Penelitian;
4. Perguruan Tinggi Ybs;
5. Ybs;
6. Arsip.

Dokumen ini dibenarkan secara elektronik menggunakan sertifikat digital yang diterbitkan oleh Data Signifikan Elektronik (DSNE), Badan Siber dan Sandi Negara (BSN)



Gambar A1 Surat Izin Pengambilan Data