

PENERAPAN *DATA MINING* ANALISA PENYAKIT

MENULAR PADA MANUSIA

SKRIPSI



Oleh:

Susi Susanti Tampubolon

170210152

PROGRAM STUDI TEKNIK INFORMATIKA

FAKULTAS TEKNIK DAN KOMPUTER

UNIVERSITAS PUTERA BATAM

2021

PENERAPAN *DATA MINING* ANALISA PENYAKIT

MENULAR PADA MANUSIA

SKRIPSI

**Untuk memenuhi salah satu syarat
memperoleh gelar sarjana**



Oleh:

**Susi Susanti Tampubolon
170210152**

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNIK DAN KOMPUTER
UNIVERSITAS PUTERA BATAM
2021**

SURAT PERNYATAAN ORSINILITAS

Yang bertanda tangan di bawah ini Penulis:

Nama : Susi Susanti Tampubolon

NPM : 170210152

Fakultas : Teknik Dan Komputer

Program Studi : Teknik Informatika

Menyatakan bahwa "Skripsi" yang penulis buat dengan judul:

PENERAPAN *DATA MINING* ANALISA PENYAKIT MENULAR PADA MANUSIA.

Ini adalah karya sendiri dan bukan "duplikasi" dari karya orang lain. Sejauh yang penulis tahu, dalam teks skripsi ini tidak ada karya ilmiah atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang disebutkan dalam teks ini dan disebutkan dalam sumber dan referensi kutipan. Jika ternyata dalam naskah ini, dapat ditunjukkan bahwa ada elemen PLAGIASI, Penulis siap untuk menghentikan naskah ini dan judul Penulis dibatalkan dan diproses sesuai dengan hukum dan peraturan yang berlaku. Jadi pernyataan yang Penulis buat ini sebenarnya tanpa paksaan dari siapa pun.

Batam, 24 Juli 2021



Susi Susanti Tampubolon
170210152

PENERAPAN *DATA MINING* ANALISA PENYAKIT

MENULAR PADA MANUSIA

SKRIPSI

**Untuk memenuhi salah satu syarat
memperoleh gelar sarjana**

**Oleh:
Susanti Susanti Tampubolon
170210152**

**Telah disetujui Pembimbing pada tanggal
seperti tertera di bawah ini**

Batam, 24 Juli 2021

A handwritten signature in blue ink, appearing to read 'Koko Handoko', is centered on the page.

**Koko Handoko, S.Kom., M.Kom.
Pembimbing**

ABSTRAK

Pemanfaatan *data mining* dalam teknologi kini semakin hari semakin beragam. Pada penelitian ini penulis membahas tentang penerapan *data mining* dibidang medis, yaitu analisa data penyakit menular pada manusia serta pemanfaatan data penyakit menular di UPT Puskesmas Sei Langkai. Penyakit menular pada manusia merupakan salah satu jenis penyakit yang memiliki jumlah data yang banyak dan bertumpuk karena pada dasarnya penyakit menular memiliki sebab akibat yang beragam sehingga pada penelitian ini penulis menerapkan *data mining* yang berguna untuk dapat mengelola data yang bertumpuk menjadi sebuah informasi yang bermanfaat. Tujuan dari penelitian ini adalah untuk mengidentifikasi kemudian menganalisis tingkat tertinggi sampai terendah dari 7 jenis data penyakit menular pada manusia dengan total 1.212 pasien ditahun 2019 dan 2020 dengan menggunakan metode *Algoritma K-Means clustering*. Dari data yang telah diolah sehingga memperoleh hasil bahwa penyakit ISPA dan COVID-19 memiliki jumlah data tertinggi diwilayah Tembesi dengan jumlah 419 pasien, penyakit kusta, DBD, dan campak memiliki jumlah data tertinggi di wilayah Sei Langkai dengan jumlah 43 pasien, sedangkan penyakit HIV dan TBC memiliki jumlah data tertinggi di desa Sei Pelunggut dengan jumlah 36 pasien. Kesimpulan dari penelitian ini adalah dengan menggunakan *Metode K-Means* dan pengujian aplikasi *RapidMiner* maka dapat mempermudah dalam pengolahan data dan memiliki nilai akhir yang akurat serta efektif digunakan dalam pengolahan data besar.

Kata Kunci : *Data mining*, Penyakit Menular, *K-Means*, *RapidMiner*

ABSTRACT

The use of data mining in technology is growing day by day. In this research, the author discusses the application of data mining in the medical field, that is data analysis of infectious diseases in humans and the use of infectious disease data in UPT Puskesmas Sei Langkai. Infectious diseases in humans are one type of disease that has a large amount of data and accumulates because basically infectious diseases have various causes and effects so that in this research the authors apply data mining which is useful to be able to manage stacked data into useful information. The purpose of this research was to identify and then analyze the highest to lowest levels of 7 types of data on infectious diseases in humans with a total of 1,212 patients in 2019 and 2020 using the K-Means clustering algorithm. From the data that has been processed get results that Acute Respiratory Infections and COVID-19 have the highest number of data in Tembesi village with 419 patients, leprosy, dengue fever, and measles have the highest number of data in Sei Langkai village with 43 patients, while HIV and TB has the highest number of data in Sei Pelunggut village with 36 patients. The conclusion of this research is using the K-Means method and testing the RapidMiner application, it can facilitate data processing and has an accurate final value and is effectively used in big data processing.

Keywords: Data mining, Infectious disease, K-Means , RapidMiner

KATA PENGANTAR

Puji dan Syukur penulis panjatkan kepada Tuhan Yang Maha Esa sehingga penulis dapat menyelesaikan Skripsi, yang menjadi salah satu persyaratan untuk menyelesaikan program studi sarjana (S1) dalam program Studi Teknik Informatika di Universitas Putera Batam.

Penulis menyadari bahwa Skripsi ini jauh dari sempurna. Karena itu kritik dan saran akan senantiasa penulis terima dengan senang hati. Dengan segala keterbatasan, dan dorongan dari berbagai pihak. Untuk itu, dengan segala kerendahan hati, penulis menyampaikan ucapan terima kasih kepada:

1. Rektor Universitas Putera Batam, Ibu Dr. Nur Elfi Husda, S.kom.,M.SI.
2. Dekan Fakultas Teknik dan Komputer, Bapak Welly Sugianto, S.T.,M.M.
3. Ketua Program Studi Teknik Informatika Bapak Andi Maslan,.S.T.,M.SI.
4. Bapak Koko Handoko,S.Kom., M.Kom. Sebagai Pembimbing Skripsi.
5. Dosen dan Staf Universitas Putera Batam.
6. Pihak Tata Usaha dan Tim Medis di Puskesmas Sei Langkai.
7. Orang Tua dan keluarga terkasih yang selalu berdoa dan memberi semangat kepada penulis sehingga dapat menyelesaikan skripsi ini.
8. Rekan kerja yang mau membagikan ilmunya dan berbagi pendapat dalam pembuatan skripsi ini
9. Rekan-Rekan mahasiswa Universitas Putera Batam yang juga memberikan doa dan dukungan mereka.
10. Teman-Teman Kampus yang selalu memberikan semangat dan dukungan kepada penulis dalam menyelesaikan skripsi ini.

11. Serta pihak yang tidak dapat disebutkan satu per satu oleh penulis, semoga senantiasanya diberkati Oleh Tuhan Yang Maha Esa, amin.

Batam, 24 Juli 2021

A handwritten signature in black ink, appearing to read 'Susi Susanti'.

Susi Susanti Tampubolon

DAFTAR ISI

HALAMAN SAMPUL	i
HALAMAN JUDUL	ii
HALAMAN PERNYATAAN	iii
HALAMAN PENGESAHAN	iv
ABSTRAK	v
ABSTRACT	vi
KATA PENGANTAR	vii
DAFTAR ISI	ix
DAFTAR GAMBAR	xii
DAFTAR TABEL	xiii
DAFTAR RUMUS	xiv
BAB I	1
PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Identifikasi Masalah	2
1.3 Batasan Masalah.....	3
1.4 Rumusan Masalah	3
1.5 Tujuan Penelitian.....	4
1.6 Mafaat Penelitian.....	4
BAB II	6

KAJIAN PUSTAKA	6
2.1 Teori Dasar	6
2.2 <i>Knowledge Discovery In Database (KDD)</i>	6
2.3 <i>Data Mining</i>	8
2.4 <i>K-Means Clustering</i>	11
2.5 Software Pendukung.....	13
2.5.1 <i>RapidMiner</i>	13
2.5.2 <i>Tanagra</i>	14
2.6 Penelitian Terdahulu.....	15
2.7 Kerangka Pemikiran	19
BAB III.....	21
METODE PENELITIAN	21
3.1 Desain Penelitian	21
3.2 Teknik Pengumpulan Data	23
3.2.1 Observasi	23
3.2.2 Wawancara.....	23
3.2.3 Studi Pustaka.....	23
3.3 Operasional Variabel.....	24
3.3.1 Kusta	24
3.3.2 <i>Demam Berdarah Dengue (DBD)</i>	24
3.3.3 <i>HIV (Human Immunodeficiency Virus)</i>	25
3.3.4 <i>TBC (Tuberculosis)</i>	25
3.3.5 Campak/Rubeola.....	26
3.3.6 <i>ISPA (Infeksi Saluran Pernafasan Akut)</i>	26
3.3.7 <i>Coronavirus (COVID-19)</i>	26
3.4 Metode Perancangan Sistem.....	27

3.5 Lokasi dan Jadwal Penelitian	29
BAB IV	30
HASIL DAN PEMBAHASAN	30
4.1 Analisa Data	30
4.2 Penyeleksian Data	30
4.3 Analisa Proses Algoritma	31
4.4 Pengujian Data menggunakan Aplikasi <i>RapidMiner 5.3</i>	37
BAB V	41
KESIMPULAN DAN SARAN	41
5.1 Kesimpulan.....	41
5.2 Saran	42
DAFTAR PUSTAKA	43
Lampiran 1. Pendukung Penelitian.....	45
Lampiran 2. Daftar Riwayat Hidup.....	46
Lampiran 3. Surat Keterangan Penelitian.....	47

DAFTAR GAMBAR

Gambar 2. 1 Aplikasi <i>RapidMiner</i>	13
Gambar 2. 2 Aplikasi <i>Tanagra</i>	14
Gambar 4 3 Tampilan awal Aplikasi <i>RapidMiner</i>	37
Gambar 4 4 Tampilan Menu Operasional Data pada <i>RapidMiner</i>	37
Gambar 4 5 Tampilan menu processing import data.....	38
Gambar 4 6 Tampilan menu pilih file dari pc/laptop	38
Gambar 4 7 Tampilan Reload data	39
Gambar 4 8 Tampilan connecting data excel ke <i>K-Means</i> pada <i>RapidMiner</i>	39
Gambar 4 9 Tampilan Output pada <i>RapidMiner</i>	40
Gambar 4 10 Tampilan output <i>RapidMiner</i> dengan Centroid Plot View	40

DAFTAR TABEL

Tabel 4 1 Analisa data Penyakit menular di Puskesmas Sei Langkai	30
Tabel 4 2 Data Penyakit menular di Puskesmas Sei Langkai	31
Tabel 4 3 Nilai Rata-rata	33
Tabel 4 4 Hasil Literasi pertama	33
Tabel 4 5 Perhitungan data menggunakan Microsoft Excel.....	34
Tabel 4 6 Hasil Literasi Kedua.....	34
Tabel 4 7 Perbandingan hasil literasi pertama dan kedua	35
Tabel 4 8 Nilai Rata-rata dari Literasi Kedua	35
Tabel 4 9 Hasil Perhitungan data Literasi ketiga	36
Tabel 4 10 Perbandingan Hasil Literasi pertama, kedua,dan ketiga	36

DAFTAR RUMUS

Rumus 2 1 Euclidean Distance.....	12
Rumus 2 2 Nilai rata-rata	12
Rumus 4 1 Mencari jarak Euclidean	31
Rumus 4 2 Mencari nilai rata-rata.....	32

BAB I

PENDAHULUAN

1.1 Latar Belakang

Perkembangan teknologi pada masa kini mendukung berbagai sektor untuk semakin berkembang dalam bidangnya, ada dibidang perdagangan seperti membantu dalam pengelompokan penjualan produk, dan ada pula dibidang kesehatan seperti memprediksi persediaan alat-alat kesehatan yang diperlukan. Beberapa sektor tersebut menggunakan *data mining* sebagai upaya dalam mempermudah pekerjaan serta memanfaatkan berbagai data dari berbagai sumber yang telah menumpuk lama agar dapat dijadikan sebuah informasi yang berguna bagi banyak kalangan.

Penggunaan *data mining* sebagai bentuk pemberdayaan teknologi serta dihubungkan dengan bidang kesehatan, mengingat semakin maraknya penyakit yang dialami oleh manusia. Pertama penyakit menular yaitu penyakit yang ditularkan dari satu orang terhadap orang lain baik langsung ataupun tidak langsung dengan berbagai cara penularan. Kedua penyakit tidak menular yaitu penyakit yang disebabkan oleh adanya ketidakseimbangan metabolisme tubuh dan bukan karena bakteri/kuman. Ketiga penyakit kronis yaitu penyakit dengan kurun waktu yang Panjang serta berpeluang besar mengakibatkan kematian pada penderitanya (Swastati, 2017).

Dalam hal ini penulis akan membahas tentang penyakit menular yang mempunyai banyak jenis terlebih penyakit menular ini memiliki sebab akibat yang

banyak pula sehingga memiliki jumlah data yang telah banyak tertumpuk selama beberapa waktu. *Data mining* merupakan suatu proses yang saling berhubungan antara lain dalam mencari sebuah pengetahuan yang baru pada sebuah set database yang berupa pengetahuan secara tidak langsung. Di bidang kecerdasan buatan (*artificial intelligent*), *machine learning*, statistik dan basis data serta tidak jarang pula dikaitkan dengan beberapa teknik dalam penerapan *Data Mining* yaitu: pengelompokan, klasifikasi, aturan asosiasi, jaringan saraf tiruan, algoritma genetic (Handoko, 2016)

Pengolahan data yang paling tepat adalah dengan menggunakan Metode *Algoritma K-Means Clustering*, dimana *Algoritma K-Means Clustering* merupakan metode yang populer praktis dalam pengelompokan data dengan karakteristiknya. Hal ini relevan dengan data pasien dari Puskesmas Sei Langkai yang banyak sehingga nantinya akan dapat mengelompokkan data pasien Puskesmas menjadi beberapa cluster dengan cluster lain yang memiliki kemiripan. Oleh karena itu Penulis berupaya menerapkan *data mining* sebagai solusi dalam pengelolaan data pasien yang bertumpuk serta jenis penyakit yang beragam. Didukung oleh perangkat lunak yang kini dapat digunakan dengan mudah serta menjadikan efisiensi waktu yang baik.

1.2 Identifikasi Masalah

Identifikasi masalah yang ditemukan oleh penulis melalui latar belakang yang telah diuraikan adalah sebagai berikut:

1. Menumpuknya data pasien di Puskesmas Sei Langkai yang tidak diolah menyebabkan data dapat hilang.

2. Banyaknya jenis penyakit menular pada manusia serta penanganan yang dilakukan oleh beberapa tenaga medis yang berbeda sehingga sulit dalam melakukan wawancara.
3. Proses pengelompokan data dengan penerapan *K-Means clustering* terhadap data pasien di Puskesmas Sei Langkai.

1.3 Batasan Masalah

Batasan masalah yang ditetapkan oleh penulis guna untuk membatasi skripsi ini sehingga tidak menyimpang dan tidak terlalu luas pembahasannya, maka ditetapkan sebagai berikut:

1. Di dalam proses penerapan *data mining*, membahas 7 jenis penyakit menular pada manusia.
2. Data yang digunakan oleh penulis adalah data pasien dari Puskesmas Sei Langkai dalam kurun waktu 2 tahun yaitu tahun 2019 dan 2020.
3. Penelitian ini dilakukan dengan penerapan prosesnya menggunakan metode *K-Means* serta Aplikasi perangkat lunak *RapidMiner*.

1.4 Rumusan Masalah

Rumusan masalah yang ditemukan dari latar belakang adalah sebagai berikut:

1. Bagaimana analisa *Data Mining* terhadap data penyakit menular di Puskesmas Sei Langkai?
2. Bagaimana penerapan metode *K-means clustering* dalam pengolahan data pasien penyakit menular pada manusia?

3. Bagaimana implementasi *software* aplikasi *RapidMiner* dalam menentukan tingkat tertinggi sampai terendah dari data penyakit menular pada manusia?

1.5 Tujuan Penelitian

1. Untuk menerapkan *data mining* khususnya Metode *K-Means Clustering* sebagai pengolah data pasien penyakit menular pada Puskesmas sei Langkai.
2. Untuk mengetahui implementasi *data mining* pada jenis penyakit menular yang diderita manusia dari tertinggi sampai terendah di Pukesmas Sei Langkai.
3. Untuk memanfaatkan data pasien penyakit menular menjadi sebuah informasi untuk Puskesmas Sei Langkai, pembaca dan juga masyarakat dalam tingkatan penyakit menular dimasa yang akan datang.

1.6 Mafaat Penelitian

Penulis membagi dua manfaat dalam penelitian ini, yaitu manfaat secara teoritis dan manfaat secara praktis.

a. Manfaat Teoritis

1. Referensi bagi pembaca atau peneliti yang akan melakukan penelitian lebih lanjut mengenai *data mining* penerapan *K-means clustering* ataupun metode lainnya dalam *data mining*.
2. Penerapan ilmu yang didapat selama belajar di Universitas Putera Batam Jurusan Teknik Informatika.

b. Manfaat Praktis

1. Menambah ilmu terlebih untuk saya dalam proses penerapan data *mining* serta penggunaan perangkat lunak Aplikasi *RapidMiner*.
2. Data pasien penyakit menular yang bertumpuk bertahun-tahun dapat dijadikan informasi oleh Puskesmas sebagai bahan kajian sosialisasi kesehatan terhadap masyarakat.

BAB II

KAJIAN PUSTAKA

2.1 Teori Dasar

Penelitian ini memiliki teori dasar yaitu mencakup KDD (*Knowledge Discovery In Database*), *Data mining*, *Algoritma K-Means Clustering*, *Software pendukung*, penelitian terdahulu dan kerangka pemikiran.

2.2 Knowledge Discovery In Database (KDD)

Data Mining sering disebut juga sebagai *Knowledge Discovery In Database* (KDD) menemukan pengetahuan baru yang didapatkan dari hasil pengolahan data. Menurut (Handoko, 2016) Konsep *Data Mining* pada saat ini adalah seperti mengumpulkan, menggunakan sebuah data besar untuk menemukan ketergantungan atau pola data yang berhubungan satu sama lain. Menurut (Sari & Harman, 2020) KDD (*knowledge discovery in database*) merupakan sebuah tahapan yang tidak mudah dalam pengidentifikasian pola pada sebuah data dimana pola tersebut juga bersifat baru, juga dapat berguna. *Knowledge Discovery in Database* (KDD) adalah proses penemuan sebuah informasi baru yang berguna dalam sebuah set *database* yang terdiri dari pemahaman di bidang aplikasi, kemudian membuat data target dalam *database*, *cleaning* data dan *preprocessing* data (Fiandra et al., 2017).

Menurut (Sinaga & Handoko, 2021) tahap-tahap dalam *Knowledge Discovery in Database* (KDD) terdiri dari:

1. *Cleaning Data*

Tahap yang dilakukan adalah membersihkan data berupa pengurangan data yang tidak tetap.

2. *Data Integration*

Tahap ini dilakukan untuk menggabungkan data secara keseluruhan dari sumber-sumber data.

3. *Data Selection*

Pada tahap ini, yang dilakukan dalam penyeleksian data adalah pemilihan data dari *database* yang sesuai dengan tujuan peneliti.

4. *Data Transformation*

Tahapan ini melakukan perubahan data sesuai dengan metode atau teknik data mining yang diterapkan.

5. *Data Mining*

Tahap ini, dilakukan implementasi terhadap metode *data mining* yang sesuai agar mendapatkan suatu pola dari data.

6. *Patten Evaluation*

Tahap ini dilakukan untuk mengidentifikasi data dan pola data.

7. *Knowledge Presentation*

Tahap ini, *data mining* menghasilkan informasi yang dapat dipresentasikan lalu dapat dijadikan informasi bagi tempat objek penelitian.

2.3 Data Mining

Menurut (Santoso, 2017) *Data Mining* merupakan sebuah metode pengolahan data guna menemukan pola baru yang terdapat pada data tersebut. Pemanfaatan *Data Mining* memang berguna sebagai bahan untuk menambahkan informasi dalam berbagai kalangan mulai dari bisnis hingga medis, ini dibuktikan juga setelah mengkaji kembali pengertian *data mining* menurut para ahli. Menurut (Kurnia et al., 2020) *Data Mining* adalah gabungan sejumlah disiplin ilmu komputer.

Selain itu, beberapa bidang dalam kehidupan sehari-hari yang mengaplikasikan *data mining* diantaranya adalah (Kurnia et al., 2020):

1. Marketing dan Bisnis perusahaan memiliki data yang berguna dalam strategi marketing dan bisnisnya. Seperti strategi dalam pemasaran produk agar menghasilkan diagram penjualan yang tinggi, pemilihan vendor yang tepat. Berikut contoh aplikasi *data mining* dalam marketing dan bisnis:
 - a) *Market Basket Analysis* atau analisis keranjang belanja dimana konsumen akan ditampilkan pada jenis belanjaan yang biasa dikonsumsinya. MBA juga dikenal dengan *association rule* (aturan asosiasi) yaitu salah satu konsep dalam *data mining* yang berusaha menemukan asosiasi atau keterkaitan data.
 - b) *Recommender System*, adalah sistem yang merekomendasikan beberapa variabel dengan tingkatan tertinggi sehingga dapat memilih dengan lebih tepat seperti dalam pemilihan rekomendasi

supplier mana yang menunjukkan performansi baik. Teknik yang digunakan aplikasi ini adalah teknik kalsterisasi ataupun klasifikasi.

- c) *Churn Prediction* merupakan analisis dari loyal atau tidaknya suatu pelanggan berdasarkan variabel-variabelnya. Sebagai Contoh perusahaan telekomunikasi yang memiliki pelanggan hampir ratusan juta ingin melihat pelanggan apakah tetap loyal atau tidak dengan menggunakan teknik *data mining* sehingga hal tersebut menjadi mudah dan cepat dilakukan. Teknik yang digunakan adalah teknik klasifikasi dan kalsterisasi.
 - d) *Fraus Decection* digunakan dalam menemukan pelanggan yang mungkin melakukan kecurangan. Sejumlah data yang besar apabila dilakukan secara manual akan membutuhkan biaya dan waktu yang lama sehingga penggunaan teknik *data mining* dapat mempercepat dalam penemuan kecurangan di dalam suatu basis data pelanggan. Sistem ini dibengun menggunakan teknik *anomaly detection*.
2. Sains dan Teknik. Beberapa teknik *data mining* dapat digunakan dalam dunia sains dan teknik untuk menyelesaikan permasalahan yang kompleks, seperti genetika, medis, teknik elektro, dan sebagainya.
 3. Seni dan Hiburan. *Data Mining* juga dapat diaplikasikan ke dalam seni dan hiburan, seperti menentukan lagu kesukaan yang sering kali diputar ataupun merekomendasikan jenis lagu ataupun video yang memiliki kemiripan yang sama dengan lagu atau video favorit.

4. Dekripsi

Deskripsi merupakan penggambaran suatu objek yang bertujuan untuk mengidentifikasi kemudian menganalisis dan mengubah bentuk suatu data yang muncul berulang menjadi bentuk yang bisa dibaca atau dipahami oleh domain aplikasinya.

5. Prediksi

Prediksi mempunyai kesamaan dengan klasifikasi, namun data yang dikelompokkan sesuai dengan perilaku atau nilai yang diperkirakan untuk waktu yang akan datang. Contoh dari pada prediksi yaitu seperti memprediksi adanya pengurangan jumlah pelanggan dalam waktu dekat atau memprediksi harga saham dalam tiga bulan yang akan datang.

6. Estimasi

Memiliki pengertian yang hampir serupa dengan prediksi, estimasi memiliki target lebih kearah numerik disbanding kearah kategori. Model dibangun menggunakan record lengkap yang menyediakan nilai dari variabel target dibuat berdasarkan nilai variabel prediksi.

7. Klasifikasi

Klasifikasi adalah proses dimana ditemukannya sebuah model atau fungsi yang menggambarkan serta membedakan data kepada beberapa kelas. Klasifikasi melibatkan proses pemeriksaan karakteristik dari objek dan memasukkan objek ke dalam salah satu kelas yang sudah didefenisikan sebelumnya.

8. *Clustering*

Clustering adalah kegiatan mengelompokkan data berdasarkan beberapa kelas dan digabungkan dengan objek yang sama. Tujuannya adalah untuk mengelompokkan objek-objek yang sama atau hampir sama kedalam beberapa kelompok. Semakin besar kemiripan objek dalam suatu cluster dan semakin besar perbedaan tiap cluster maka kualitas analisis cluster semakin baik.

9. Asosiasi

Asosiasi dalam *data mining* merupakan penemuan atribut yang muncul dalam suatu waktu. Dalam dunia bisnis lebih umum disebut analisis keranjang belanja (*market basket analysis*). Tugas asosiasi adalah untuk menemukan aturan untuk mengukur hubungan antara dua atau lebih atribut yang berhubungan.

2.4 *K-Means Clustering*

Metode *K-Means Clustering* atau dikenal juga *Algoritma K-Means Clustering* adalah metode yang sudah tidak asing lagi dalam pengolahan data serta terkenal praktis. Tujuannya juga untuk pengelompokan data atau objek menjadi beberapa cluster (grup) maka pada setiap cluster akan diisi dengan beberapa data dengan cluster terdekatnya. Menurut (Rochcham, 2020) *Algoritma K-Means* adalah metode *data mining* yang sering digunakan untuk mengidentifikasi dan menganalisis kemiripan dalam pengelompokan data. Menurut (Thabit et al., 2020) *K-Means* yaitu metode algoritma yang menganalisa data dengan menentukan nilai pada data yang

akan dikelompokkan secara acak dan menemukan objek pada satu kelompok yang sama atau memiliki hubungan atau yang tidak berhubungan dengan objek kelompok lainnya.

Langkah-langkah dalam membentuk cluster secara iteratif yaitu:

Step 1 : Tentukan dulu jumlah cluster K yang akan dibentuk.

Step 2 : Menentukan titik *clustering* secara acak berdasarkan cluster nya.

Step 3 : Menghitung jarak antar data dengan titik clustering menggunakan rumus *Euclidean Distance*:

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

Rumus 2 1 *Euclidean Distance*

Step 4 : Setelah data dikelompokkan berdasarkan jarak yang terdekat dengan setiap titik *clustering*, untuk menentukan ataupun menghitung titik clustering yang baru ditemukan yaitu dengan menghitung nilai rata-rata dari titik data yang ada di cluster masing-masing dengan menggunakan rumus:

$$\mu_k = \frac{1}{N_k} \sum_{q=1}^{N_k} x_q$$

Rumus 2 2 Nilai rata-rata

Step 5 : Lakukan proses literasi sampai selesai. Literasi terakhir berakhir apabila nilainya sama dengan iterasi sebelumnya.

K-Means adalah metode yang membagi data menjadi beberapa cluster berbasis jarak dengan atribut numeric. *Algoritma K-Means* terkenal karena kemudahannya dalam mengolah data besar dan cepat. *Algoritma K-Means clustering* adalah

algoritma yang membutuhkan parameter input sebanyak k dan membagi sekumpulan n objek kedalam k cluster membentuk grup dengan tingkat kemiripan yang tinggi sedangkan pada anggota cluster lain memiliki tingkat kemiripan yang rendah.

2.5 Software Pendukung

Berikut beberapa *software* pendukung yang dapat digunakan dalam pengolahan *data mining* dengan menggunakan metode *K-Means Clustering*.

2.5.1 RapidMiner

Rapidminer merupakan sebuah perangkat lunak yang difungsikan dalam membantu analisis *data mining*, *text mining* serta analisis prediksi. Sebagai *software* yang bersifat *Open source* atau terbuka *RapidMiner* ini telah menempati peringkat pertama sebagai *Software data mining* pada pemilihan oleh KDnuggets, sebuah portal *data mining* pada tahun 2010-2011. Dikenal sebagai alat bantu dalam membuat keputusan yang baik, *RapidMiner* menggunakan teknik *deskriptif* dan prediksi terhadap wawasan penggunanya.



Gambar 2. 1 Aplikasi *RapidMiner*

Sumber: *RapidMiner Studio* (2019)

2.5.2 *Tanagra*

Tanagra adalah salah satu *software* dalam *data mining* yang memiliki *User interface* sederhana sehingga dikenal mudah dalam pengoperasiannya. Aplikasi *Tanagra* juga memberikan akses terhadap algoritma dalam teknik *data mining*, penggunaan aplikasi *Tanagra* juga di set mengidentifikasi data dengan *database* dalam bentuk *txt* pada notepad dan *xls* pada Microsoft excel. Akan tetapi perangkat lunak ini tidak memasukkan set sumber data yang luas, akses langsung ke data warehouses, dan *database*, data *cleansing* dan *interactive utilization* seperti yang ada pada *software* komersil saat ini.



Gambar 2. 2 Aplikasi *Tanagra*

Sumber: *Tanagra Data Mining*

2.6 Penelitian Terdahulu

Ada beberapa penelitian terdahulu yang dijadikan sebagai acuan dalam penelitian ini, berikut adalah beberapa penelitian tersebut:

1. Menurut (Swastati, 2017) Penelitian yang berjudul **“Pengenalan Penyakit Pada Manusia Berbasis Android Menggunakan Metode *Sequential Search*”** membahas tentang pengenalan penyakit pada manusia yang sering diderita yaitu berupa penyakit tidak menular, penyakit menular, dan kronis dengan menggunakan metode *sequential search* dimana penggunaan metode ini adalah unruk mencari sebuah data dari kumpulan data dari awal sampai akhir fungsinya adalah agar dapat membantu masyarakat dalam mencari data. Kemudian aplikasi yang digunakan adalah android 4.4 (KitKat), android (5.0) Lollipop, android (5.1) Lollipop dengan hasil cukup baik artinya masih perlu penambahan dan hanya dapat dijalankan pada android.
2. Menurut Penelitian (MURTI, 2017) yang berjudul **“Penerapan Metode *K-Means Clustering* untuk mengelompokkan potensi produksi buah-buahan di Provinsi Daerah Istimewa Yogyakarta”** membahas tentang pengelompokan hasil produksi buah-buahan dalam bebrapa daerah dengan menggunakan metode *K-Means* seperti berdasarkan luas panen, produksi dan tahun panen berdasarkan data di beberapa daerah tujuannya adalah untuk memudahkan pengelompokan suatu daerah dengan hasil produksi buah yang paling banyak, sedang dan rendah. Hasilnya adalah akan ditemukan pengelompokan daerah dengan potensial produksi buah yang paling tinggi.

3. Menurut Penelitian (Bastian et al., n.d.) pada tahun 2018 Penelitian yang berjudul **“Penerapan Algoritma *K-Means clustering* Analisis Pada Penyakit Menular Studi kasus Kabupaten Majalengka)**” membahas tentang penerapan Algoritma *K-Means Clustering* dalam pengelompokan data penyakit menular pada manusia berdasarkan data yang diperoleh dari puskesmas di Kabupaten Majalengka yang terdapat ada 32 kantor Puskesmas dengan mengangkat 6 jenis data penyakit menular yang dikumpulkan dari sejumlah Puskemas di Kabupaten Majalengka. Hasil dari penelitian tersebut akan diketahui Puskesmas yang mendominasi dengan tingkat tertinggi penderita penyakit menular serta jenis penyakitnya sehingga tiap-tiap puskesmas dari kabupaten Majalengka dapat mengendalikan persediaan obat serta penanganan yang lebih intensif sesuai dengan hasil data yang diperoleh.
4. Menurut Penelitian (Handoko & Lesmana, 2018) yang berjudul **“*Data Mining* pada jumlah penumpang menggunakan metode *Clustering*”** membahas tentang pengelompokan jumlah penumpang di bandar udara hang Nadim dengan menggunakan metode *clustering* dengan beberapa variabel yaitu variabel pertama penumpang yang datang, kedua penumpang yang berangkat dan ketiga penumpang yang transit dengan data banyak, sedang dan sedikit yaitu di tahun 2015 hingga tahun 2017 juga dilakukan pengujian dengan menggunakan Aplikasi *RapidMiner*. Hasil yang diperoleh adalah pengetahuan tentang jadwal padat pelanggan bandar udara hang nadim perbulannya sehingga dapat membantu dalam mengantisipasi petugas bandar udara hang

Nadim terhadap penumpang yang berangkat, penumpang yang datang serta penumpang yang transit di bulan-bulan tertentu.

5. Menurut Penelitian (Dhuhita, 2015) yang berjudul **“Clustering Menggunakan Metode K-Means untuk menentukan Status Gizi Balita”** membahas tentang pengelompokan antara tinggi badan balita (TB) dan berat badan balita (BB), dikelompokkan menjadi status gizi balita kedalam 5 cluster status gizi yaitu gizi buruk, gizi kurang, gizi baik, gizi lebih dan obesitas menggunakan tabel *Growth Chart* dengan menggunakan data balita dengan jumlah 50 balita di Desa Karang Songo dengan usia < 3 tahun. Kemudian dilakukan perhitungan cluster dengan menggunakan SPSS, Analisa hasil data output lalu melakukan pengujian dengan membandingkan hasil pengelompokan *algoritma K-Means* dan tabel *Growth Chart*. Hasilnya adalah didapatkan 17 data yang memiliki kelompok yang sama sehingga disimpulkan *algoritma K-Means* memiliki nilai akurasi 34% benar dan dapat berubah sesuai dengan data yang ditambahkan.
6. Menurut Penelitian (Prediction & Syndromes, 2019) yang berjudul **“Disease Prediction through syndromes using K-Means algorithm”**. *This study describe a research work aiming to find out how much efficient K-Means can be build an expert system to detect human disease by evaluating symptoms data to improve the quality of health evaluation. This research collected 61 symptoms and there certain 20 diseases, that is Anemia, Angina, Asthma, Bacillary Dysentery, Bronchiolitis, Chickenpox, Dengue Fever, Diabetes Mellitus, Diarrhea, Jaundice, Leukemia, Malaria, Myocardial Infarction (MI), Peptic Ulcer, Pneumonia, Rheumatic Fever, Scurvy, Stroke, Tuberculosis,*

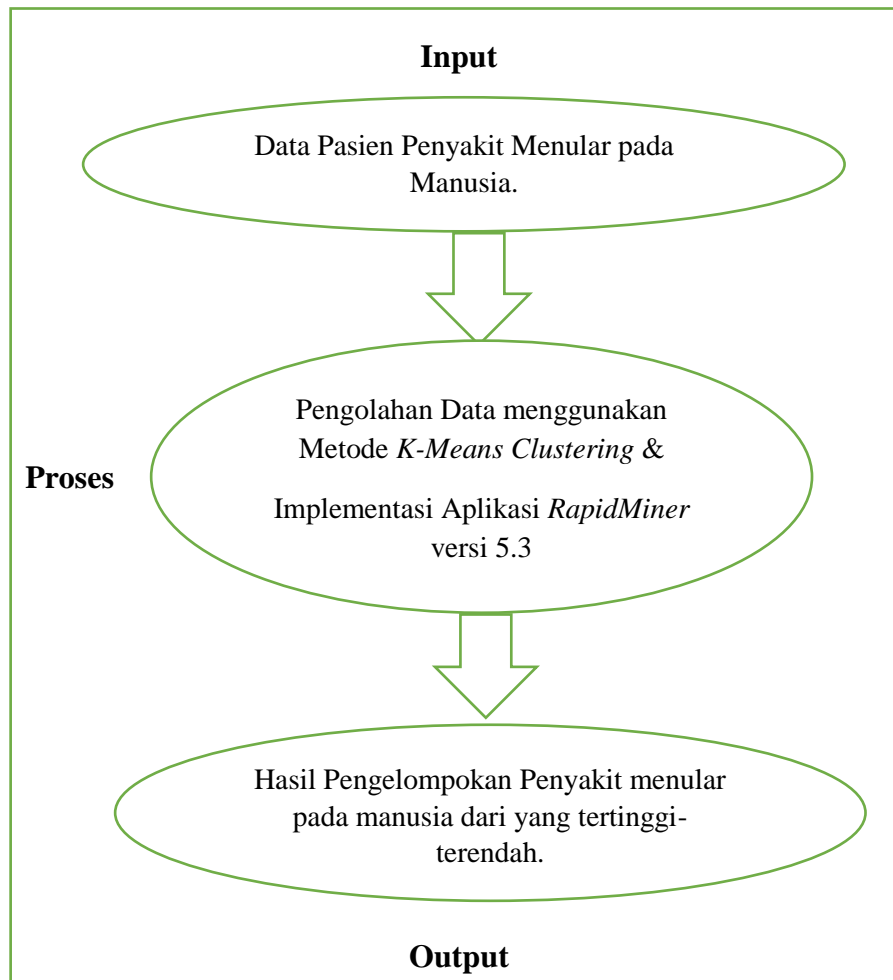
Typhoid Fever. Then applied K-Means to the dataset to predict the outcome of each point based on selected independent data and evaluated the performance of the model 'f1-score' does it is combines precision can give a better intuition about the performance. Conclusion of this research is analize disease in the healthcare domain to discover a new range of patterns and information using K-Means clustering and to make K-Means more effective it can be used in combination with other algorithms to produce accurate, relevant and useful results.

7. Menurut Penelitian (Li, 2019) yang berjudul **“Study on the Grouping of Patients with Chronic Infectious Diseases Based on Data Mining”**. *This study describes data mining technology research that is used to classify chronic infectious disease patients in order to predict patients according to the level of infectious disease suffered. K-Means clustering algorithm was used to classify chronic infectious disease patients, then C5.0 decision tree algorithm was used to predict the state of chronic infectious disease patients with 170,246 outpatient data, 43,448 data formed after cleaning data. The C5.0 decision tree algorithm was used to predict the treatment situation of patients with chronic infectious diseases, 99.94% accuracy rate verified by the confusion model. The conclusion of this research study is that medical institutions should be better at socializing chronic infectious diseases to patients and their communities, providing solutions to help them improve medication adherence. To accelerate the development of hospital information as well as in handling patients to build a database of chronic infectious diseases to determine the level of ups and*

downs of chronic infectious diseases from time to time, strengthen the blocking of transmission from mother to child, to effectively curb chronic infectious diseases, reduce the burden of disease and death.

2.7 Kerangka Pemikiran

Kerangka Pemikiran merupakan sebuah diagram atau tabel yang menggambarkan secara garis besar berjalannya suatu penelitian. Adapun kerangka pemikiran dari penelitian ini adalah sebagai berikut:



Gambar 2.3 Kerangka Pemikiran

Sumber: Data Peneliti (2021)

Sebagai input dalam penelitian ini adalah data pasien penyakit menular pada manusia di UPT Puskesmas Sei Langkai yang meliputi 3 kelurahan/desa yaitu Sei Langkai, Tembesi, dan Sei Pelunggut yang akan diproses dengan menggunakan Metode *K-Means Clustering* kemudian di Implementasikan juga menggunakan *Software Aplikasi data mining RapidMiner* sebagai pengujian data serta akan menghasilkan pengelompokan Penyakit menular pada manusia dari yang tertinggi sampai terendah di UPT Puskesmas Sei Langkai sebagai Outputnya.

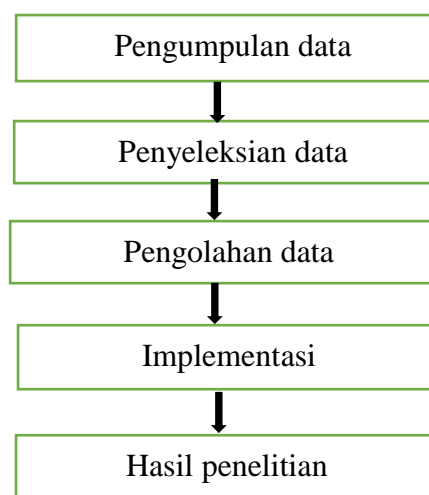
BAB III

METODE PENELITIAN

3.1 Desain Penelitian

Dalam bab ini Peneliti membuat desain penelitian sebagai gambaran dari rancangan penelitian *data mining* dimulai dengan pengumpulan data sebagai awal penelitian hingga hasil akhir penelitian menggunakan metode *K-Means clustering* serta Penerapan *software* aplikasi *RapidMiner*.

Berikut penjelasan dari penelitian tersebut:



Gambar 3.1 Desain Penelitian

Sumber: Data Peneliti (2021)

Berdasarkan gambar 3.1 penjelasannya adalah sebagai berikut:

1. Pengumpulan Data

Pengumpulan data dilakukan dengan cara observasi, wawancara di Puskesmas Sei Langkai serta mengumpulkan data dari berbagai sumber yaitu *Ebook*, Jurnal nasional dan Internasional, buku yang terkait tentang *data*

mining guna untuk memperoleh data yang diperlukan dan akan dipilih nantinya sesuai dengan topik penerapan *data mining* dan aplikasi *RapidMiner*.

2. Penyeleksian Data

Penyeleksian terhadap data dilakukan guna untuk menseleksi data atau memilih data penyakit menular pada manusia dari berbagai data penyakit yang diperoleh dari Puskesmas Sei Langkai sebagai sumber datanya. Data yang diambil adalah data penyakit menular pada tahun 2019 dan 2020 dengan 7 jenis penyakit menular yang ada di Puskesmas Sei Langkai.

3. Pengolahan Data

Pengolahan terhadap data penyakit menular manusia yang diperoleh dari sumber guna untuk mempermudah proses tahapan dalam penerapan dan implementasi *data mining* sesuai dengan metode dan aplikasi yang digunakan dalam *data mining*.

4. Implementasi Data

Implementasi data dilakukan guna untuk mengolah data sesuai dengan penerapan *data mining* menggunakan metode *Algoritma K-Means clustering* kemudian dilanjutkan dengan implementasi menggunakan *software* aplikasi *RapidMiner* versi 5.3.

5. Hasil Penelitian

Hasil dari penelitian yang dilakukan oleh peneliti dengan menggunakan metode *K-Means clustering* dan bantuan *software* aplikasi *RapidMiner* akan menghasilkan tingkatan penyakit menular yang dari tertinggi hingga terendah.

3.2 Teknik Pengumpulan Data

Teknik pengumpulan data yang dilakukan peneliti dalam penelitian ini adalah sebagai berikut:

3.2.1 Observasi

Pengumpulan data dengan cara observasi oleh peneliti yaitu dengan mengunjungi Puskesmas Sei Langkai secara langsung guna untuk mendapatkan informasi tentang topik penelitian yang terkait.

3.2.2 Wawancara

Peneliti melakukan wawancara kepada Ibu Serlinawati sebagai bagian Tata Usaha di Puskesmas Sei Langkai guna meminta izin mengadakan penelitian dan memberitahukan tujuan daripada penelitian tersebut serta memperoleh informasi tentang data penyakit menular pada manusia yang ada di Puskesmas Sei Langkai. Wawancara juga dilakukan peneliti kepada beberapa ibu yang merupakan tenaga medis yang menangani penyakit menular pada manusia di Puskesmas Sei Langkai guna memperoleh informasi terkait dengan penelitian yang dilakukan.

3.2.3 Studi Pustaka

Sebagai pendukung penelitian ini, peneliti mengumpulkan referensi dari berupa buku, *ebook*, Jurnal nasional dan internasional serta berbagai sumber lainnya yang berkaitan dengan materi *data mining*, metode *K-Means clustering* dan aplikasi *RapidMiner*.

3.3 Operasional Variabel

Operasional variabel pada penelitian ini meliputi beberapa jenis penyakit menular pada manusia serta data yang digunakan oleh peneliti adalah data pasien di Puskesmas Sei Langkai. Data pasien yang ada di Puskesmas Sei Langkai mencakup beberapa kelurahan, yakni kelurahan Sei Langkai, Tembesi, dan Sei Pelunggut serta beberapa data penyakit menular yang didapat yaitu berupa Penyakit Kusta, Penyakit DBD, Penyakit HIV, Penyakit TBC, Penyakit campak/rubeola, dan Penyakit ISPA di tahun 2019 kemudian munculnya Penyakit *Coronavirus* ditahun 2020.

3.3.1 Kusta

Kusta atau lepra adalah penyakit menular yang ditularkan melalui percikan cairan dari saluran pernapasan saat bersin atau batuk. Penyakit kusta disebabkan oleh infeksi bakteri *Mycobacterium leprae* yang dapat menularkan penyakit tersebut dengan mudah. Pengobatan untuk penyakit kusta sendiri dapat memakan waktu selama 1 tahun. Gejala yang dialami penderita penyakit kusta yaitu Mati rasa, bercak putih seperti panu, mimisan dan otot melemah dan juga dapat mengakibatkan komplikasi.

3.3.2 Demam Berdarah Dengue (DBD)

Demam berdarah dengue merupakan penyakit menular yang disebabkan oleh virus DBD yang dibawa oleh nyamuk *Aedes Aegypti* yang merupakan nyamuk betina yang sudah terinfeksi virus *dengue*. Menyebabkan infeksi pada sel-sel tubuh yang sehat apabila sudah masuk kedalam darah dan mengalir keseluruh tubuh. Gejala yang biasanya dialami oleh penderita DBD yaitu demam tinggi mendadak,

nyeri otot dan sendi, kelelahan/panas dingin, mual, muntah, dan sakit kepala/tenggorokan. Waktu untuk penderita DBD mengalami gejala adalah sekitar 4-10 hari setelah gigitan nyamuk.

3.3.3 HIV (*Human Immunodeficiency Virus*)

HIV atau *Human Immunodeficiency Virus* merupakan jenis penyakit menular yang paling banyak ditularkan melalui hubungan intim yang tidak aman dan berganti-ganti pasangan serta penggunaan jarum suntik yang tidak steril ketika menggunakan narkoba. Seringkali HIV dan AIDS diartikan sama oleh masyarakat, padahal sebenarnya AIDS (*Acquired Immune Deficiency Syndrome*) itu sendiri merupakan jenis HIV tahap infeksi akhir. Gejala yang biasanya dialami oleh penderita HIV yaitu mengalami nyeri mirip seperti flu, demam, nyeri tenggorokan, kelelahan, nyeri otot, bahkan ada pula yang tidak menimbulkan gejala apapun selama beberapa tahun.

3.3.4 TBC (*Tuberculosis*)

TBC atau *tuberculosis* merupakan penyakit menular yang ditularkan melalui percikan ludah, baik ketika batuk atau bersin. Biasanya rentan terhadap seseorang yang memiliki imun tubuh yang rendah dan memiliki penyakit HIV. Gejala yang biasanya dialami oleh penderita TBC yaitu demam, lemas, berat badan menurun, nyeri dada. Kurun waktu yang diderita oleh penderita *tuberculosis* akan menimbulkan gejala selama lebih dari 3 minggu.

3.3.5 Campak/*Rubeola*

Campak atau disebut juga *rubeola* merupakan penyakit menular yang ditularkan melalui infeksi percikan cairan oleh penderita campak seperti pada saat bersin atau batuk, dapat juga ditularkan melalui benda-benda yang telah dihindangi virus tersebut. Gejala yang biasanya dialami oleh penderita campak yaitu demam tinggi, lemas dan letih, sakit dan nyeri, diare atau muntah-muntah dan bercak atau ruam berwarna merah kecokelatan muncul dikulit setelah beberapa hari terdampak penyakit tersebut.

3.3.6 ISPA (Infeksi Saluran Pernafasan Akut)

ISPA atau infeksi saluran pernafasan akut merupakan penyakit menular yang disebabkan oleh virus dan ditularkan melalui udara berupa percikan liur penderita ISPA, dapat juga menular lewat sentuhan terhadap benda yang telah terkena liur penderita atau berjabat tangan dengan penderita ISPA. Selain itu kelompok yang biasanya sangat mudah tertular oleh penyakit ISPA adalah anak-anak dan lansia. Gejala yang biasanya dialami oleh penderita ISPA yaitu seperti batuk, pilek disertai dengan demam. Namun dikarenakan oleh virus para penderita penyakit ISPA bisa sembuh dengan sendirinya tanpa pengobatan khusus.

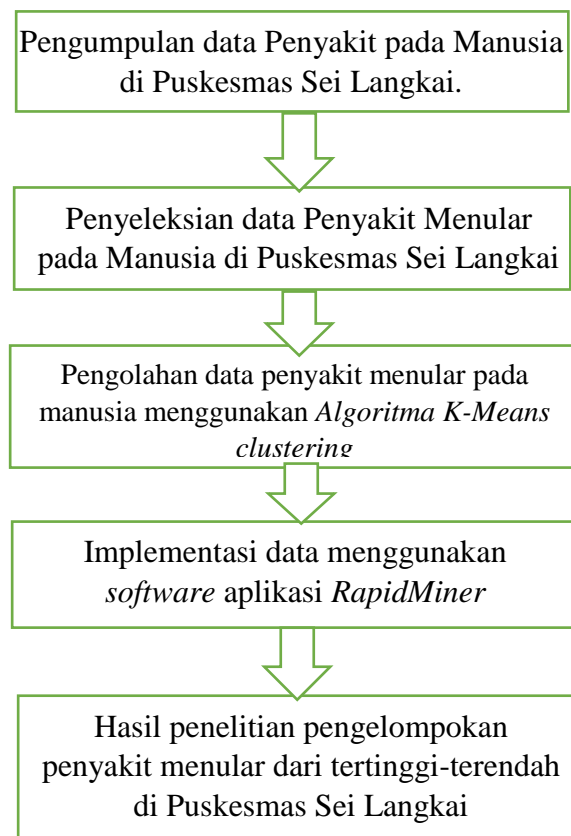
3.3.7 *Coronavirus* (COVID-19)

Coronavirus atau covid-19 merupakan penyakit menular yang muncul diakhir tahun 2019 pertama kali di Wuhan, China lalu kemudian menyebar dengan pesat ke berbagai negara lainnya salah satunya Indonesia. Pada umumnya virus corona didapati pada hewan namun banyaknya jumlah kasus covid-19 menunjukkan bahwa virus tersebut dapat ditularkan dari manusia ke manusia lain

melalui partikel air yang berukuran sangat kecil. Virus tersebut dikeluarkan pada saat batuk atau bersin kemudian apabila mengenai kornea mata maka dapat beresiko dapat tertular oleh penyakit ini. Gejala yang dialami oleh penderita *Coronavirus* bervariasi yaitu demam, batuk, pilek, sakit tenggorokan, diare, gagal napas hingga berakhir pada kematian.

3.4 Metode Perancangan Sistem

Metode perancangan sistem pada penelitian ini dapat dilihat sebagai berikut:



Gambar 3.1 Desain Penelitian

Sumber: Data Peneliti (2021)

Dari gambar 3.1 diatas dijelaskan bahwa proses awal yang dilakukan oleh peneliti adalah Pengumpulan data penyakit pada manusia di Puskesmas Sei Langkai yaitu memperoleh data dari beberapa tenaga medis sesuai dengan penyakit yang ditangani, kemudian melakukan seleksi terhadap data penyakit pada manusia tersebut dari Puskesmas Sei Langkai yang mana data yang diterima termasuk dalam penyakit menular sesuai dengan bahasan yang diambil oleh peneliti. Setelah tahap seleksi maka dilakukan pengolahan data penyakit menular pada manusia sesuai dengan teknik *data mining* dengan menggunakan *Algoritma K-Means Clustering*, selanjutnya data tersebut pun akan diuji pula dengan menggunakan aplikasi *RapidMiner* versi 5.3, untuk membuktian hasil yang diperoleh melalui perhitungan manual sesuai dengan hasil pengujian dari aplikasi *RapidMiner* versi 5.3.

3.5 Lokasi dan Jadwal Penelitian

Penelitian ini dilakukan di UPT Puskesmas Sei Langkai, Sagulung, Kota Batam, Kepulauan Riau. Jadwal penelitian dilakukan selama 5 bulan dan dijabarkan sebagai berikut:

Tabel 3.1 Jadwal Penelitian

Kegiatan	Waktu Kegiatan																					
	Februari 2021				Maret 2021				April 2021				Mei 2021				Juni 2021				Juli 2021	
	Minggu Ke				Minggu Ke				Minggu Ke				Minggu Ke				Minggu Ke				Minggu Ke	
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2
Pengajuan Judul																						
Penyusunan BAB I																						
Penyusunan BAB II																						
Penyusunan BAB III																						
Penyusunan BAB IV																						
Penyusunan BAB I-V																						
Pengumpulan Skripsi																						

Sumber: Data Peneliti (2021)