

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 mining*. *Data mining* 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.