

BAB I

PENDAHULUAN

1.1 Latar Belakang

Data mining adalah proses menganalisa data dari perspektif yang berbeda dan menyimpulkannya menjadi informasi-informasi penting yang dapat dipakai untuk meningkatkan keuntungan, memperkecil biaya pengeluaran, atau bahkan keduanya. Secara teknis, data mining dapat disebut sebagai proses untuk menemukan korelasi atau pola dari ratusan atau ribuan *field* dari sebuah relasional *database* yang besar (Kiki Fatmawati & Agus Perdana, 2020).

Pada penelitian ini akan digunakan data mining dengan teknik *K-Means Cluster* adalah salah satu bentuk pengelompokan yang paling sederhana. Prosedurnya sederhana dan mudah untuk mengklasifikasikan data yang diberikan melalui sejumlah cluster. Penentuan centroid dilakukan dengan cara mengambil data pertama sebagai centroid pertama, data kedua sebagai centroid kedua, dan seterusnya hingga jumlah centroid yang diperlukan. Langkah berikutnya adalah dengan menghitung jarak dari titik yang akan di cluster ke setiap centroid yang ada dan dikelompokkan sesuai dengan jarak terdekat kepada centroid-nya. Bila semua titik sudah masuk kedalam pengelompokan maka langkah pertama selesai. Kemudian langkah berikutnya, kita perlu menghitung kembali k-centroid baru sebagai barycenters dari kelompok yang dihasilkan. Setelah memiliki k-centroid yang baru, pengelompokan di uji kembali terhadap k-centroid (Try Iryanto & Rini Arianty, 2019).

Mewabahnya COVID-19 di dunia juga mempengaruhi Indonesia terutama Provinsi Sumatera Barat yang merupakan nomor 11 diantara 34 Provinsi yang ada di Indonesia (covid-19.go.id, 2021). Identifikasi pasien infeksi Covid-19 diperoleh setelah dilakukan pemeriksaan laboratorium dimana didasari oleh penyakit bawaan (komorbid) seperti: Hypertension, Diabetes, COPD, CVD, Liver diseases, Obesity, Renal disesases dan Malignancy (Hasan Ejaza, Dkk tahun:2020).

Data yang digunakan meliputi diagnosa pasien yang telah dikodekan dengan standar internasional untuk klasifikasi penyakit dan gangguan kesehatan lainnya yaitu ICD-10, Umur, Data sekunder dan Primer, Selanjutnya data ini diolah dengan *Software rapidminer* agar mendapatkan data cluster. Selanjutnya digunakan juga teknik perhitungan secara matematika untuk melihat hasil akurasinya.

Karena data kematian covid-19 dirumah sakit hanya tersimpan di rekam medis dan belum ada pemetaan penyakit penyerta (komorbid) yang meninggal karena Covid-19 maka penulis membuat pemetaan terhadap penyakit penyerta (*Komorbid*)

Adapun yang menjadi lokasi penelitian adalah RSUP. Dr. M. Djamil Padang yang merupakan Rumah Sakit Umum Pusat Dr. M. Jamil Padang yang dapat memberikan Layanan kesehatan bagi masyarakat khususnya yang berada di Kota Padang. Berdasarkan hasil observasi penelitian yang ada di RSUP Dr. M. Jamil Padang, peneliti mengambil data rekam medis kematian akibat Covid-19 serta penyakit bawaan (*komorbid*) pada pasien COVID-19.

Algoritma clustering mencoba untuk membagi kumpulan data menjadi klaster yang anggotanya relatif sama, dimana kemiripan dokumen di klaster yang sama tinggi, dan kemiripan dokumen di klaster lain kecil. Algoritma clustering mencoba membuat klaster dokumen yang between-clusters variation (BCV) lebih tinggi dibandingkan dengan within-cluster cluster variation (WCV) (Robani & Widodo,2016).

Dari latar belakang yang telah diuraikan dapat disimpulkan perlunya cluster pendeteksi penyakit bawaan (*komorbid*) menggunakan program *rapidminer*. Berdasarkan hal tersebut maka peneliti mengangkat judul **“Identifikasi Tingkat Kematian Penderita COVID-19 berdasarkan Penyakit Penyerta (Komorbid) Menggunakan Metode K-Means Cluster”**. Metode ini bertujuan menghasilkan cluster data kematian akibat *COVID-19* berdasarkan penyakit bawaan (*Komorbid*) menggunakan teknik data mining metode *K-Means Clustering*.

1.2 Perumusan Masalah

Berdasarkan latar belakang permasalahan tersebut dapat dirumuskan permasalahan sebagai berikut:

1. Bagaimana menerapkan metode *k-means Cluster* untuk Cluster penyakit penyerta (*Komorbid*) penderita COVID-19 di RSUP Dr. M. Jamil?

2. Bagaimana mengelompokkan data kematian COVID-19 berdasarkan Penyakit penyerta (*Komorbid*)?

1.3 Batasan Masalah

Agar pembahasan dalam penelitian ini tidak menyimpang dari rumusan masalah penulis memberikan pembatasan masalah sebagai berikut:

1. Peneliti hanya meneliti penyakit bawaan berupa data Sekunder dengan kode *ICD10*.
2. Pengelompokkan data angka kematian penderita *COVID-19* berdasarkan penyakit bawaan (*komorbid*).
3. Ujicoba Menggunakan Aplikasi *Rapid Miner*.

1.4 Tujuan Penelitian

Adapun tujuan dari penulisan proposal ini adalah:

1. Penelitian ini bertujuan menghasilkan cluster kematian penderita *COVID-19* berdasarkan penyakit penyerta (*Komorbid*).
2. Menggunakan Metode *K-Means Clustering*.

1.5 Manfaat Penelitian

Ada beberapa manfaat penelitian dalam penulisan proposal:

1. Bagi Mahasiswa
 - 1) Memperdalam dan memahami ilmu tentang data mining Terutama *K-Means Clustering*.
 - 2) Menerapkan disiplin ilmu dan memanfaatkan serta menambah bekal pengetahuan yang dapat digunakan untuk persiapan dalam rangka menghadapi dunia kerja dimasa yang akan datang.
2. Bagi Universitas Putra Indonesia “YPTK” Padang
 - 1) Mengetahui kemampuan mahasiswa dalam menguasai dan menerapkan materi ataupun teori yang telah di peroleh selama perkuliahan.

- 2) Memberikan gambaran tentang kesiapan mahasiswa dalam menghadapi dunia kerja dari hasil yang di peroleh selama belajar di perkuliahan.
 - 3) Memperbanyak referensi bagi peneliti berikutnya di bidang sistem informasi.
3. Bagi RSUP Dr.M. Jamil Kota Padang
- 1) Hasil penelitian ini dapat bermanfaat untuk membantu RSUP Dr.M. Jamil dalam mengelompokkan penyakit penyerta (*Komorbid*) penderita COVID-19.
 - 2) Sebagai bahan masukan bagi RSUP Dr.M. Jamil agar dapat membuat tindakan bagi pasien COVID-19.
 - 3) Dalam penelitian diharapkan dapat di jadikan suatu gambaran untuk mengetahui pengelompokkan tingkat kematian akibat penyakit penyerta (*Komorbid*)

1.6 Sistematika Penulisan

Laporan tesis ini di bagi menjadi 6 bab yang saling berhubungan satu sama lain. Adapun sistematika penulisan laporan ini sebagai berikut:

BAB I PENDAHULUAN

Menjelaskan tentang latar belakang, perumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian dan sistematika penulisan laporan penelitian.

BAB II LANDASAN TEORI

Menjelaskan tentang landasan teori yang berhubungan dengan penelitian penerapan metode data mining K-Means Clustering untuk mengelompokkan data cluster pada pasien COVID-19 dengan penyakit bawaan (komorbid) di RSUP Dr. M. Jamil Padang.

BAB III METODOLOGI PENELITIAN

Menjelaskan tentang kerangka kerja penelitian secara sistematis metode K-Means cluster untuk pengelompokkan tingkat kematian pasien COVID-19 dengan penyakit bawaan (Komorbid) di RSUP Dr. M. Jamil Padang.

BAB IV ANALISA DAN PERANCANGAN

Menjelaskan tentang analisa dan perancangan dari sistem yang dibuat, seperti analisa data yang dibutuhkan dalam sebuah penelitian.

BAB V IMPLEMENTASI DAN HASIL

Menjelaskan tentang hasil penelitian, seperti spesifikasi implementasi perangkat keras (*hardware*), spesifikasi perangkat lunak (*software*), dan implementasi sistem.

BAB VI KESIMPULAN DAN SARAN

Menjelaskan tentang kesimpulan dari beberapa bab-bab sebelumnya. Dalam bab ini juga berisi saran-saran yang diajukan pengembangan pada penelitian selanjutnya.

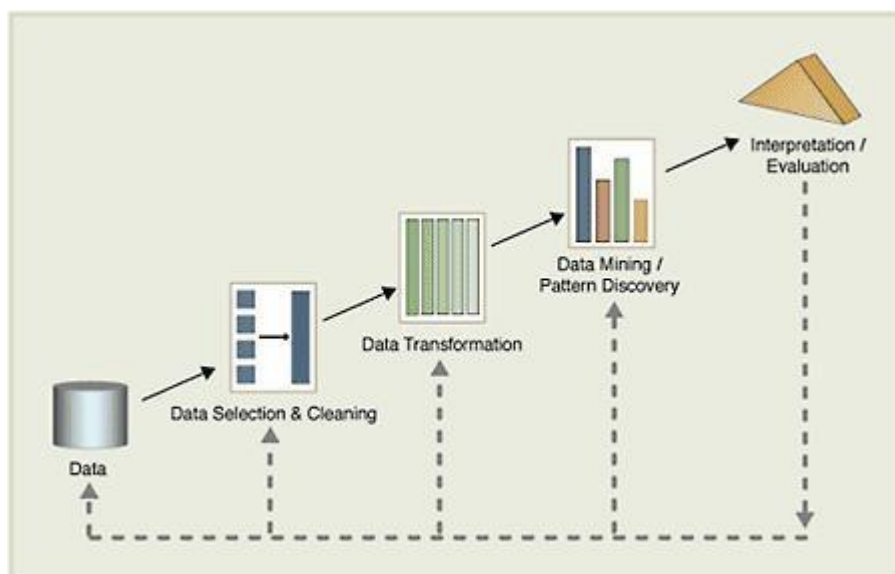
BAB II

LANDASAN TEORI

2.1 Knowledge Discovery In Database (KDD)

Knowledge Discovery In Databases (KDD) adalah keseluruhan proses nontrivial untuk mencari dan mengidentifikasi pola (pattern) dalam data, dimana pola yang ditemukan bersifat sah, baru, dapat bermanfaat dan dapat dimengerti. KDD berhubungan dengan teknik integrasi dan penemuan ilmiah, interpretasi dan visualisasi dari pola-pola sejumlah kumpulan data (Muhammad Iqbal, 2019).

Gambar 2.1 Tahapan dari KDD (Muhammad Iqbal, 2019)



1. Data Selection

- Menciptakan himpunan data target, pemilihan himpunan data, atau memfokuskan pada subset variabel atau sampel data, dimana penemuan (discovery) akan dilakukan.

- Pemilihan (seleksi) data dari sekumpulan data operasional perlu dilakukan sebelum tahap penggalian informasi dalam KDD dimulai. Data hasil seleksi yang akan digunakan untuk proses data mining, disimpan dalam suatu berkas, terpisah dari basis data operasional.

2. *Pre-Processing/Cleaning*

- Pemrosesan pendahuluan dan pembersihan data merupakan operasi dasar seperti penghapusan noise dilakukan.
- Sebelum proses data mining dapat dilaksanakan, perlu dilakukan proses cleaning pada data yang menjadi fokus KDD.
- Proses cleaning mencakup antara lain membuang duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data.
- Dilakukan proses enrichment, yaitu proses “memperkaya” data yang sudah ada dengan data atau informasi lain (eksternal).

3. *Transformation*

- Pencarian fitur-fitur yang berguna untuk mempresentasikan data bergantung kepada goal yang ingin dicapai.
- Merupakan proses transformasi pada data yang telah dipilih, sehingga data tersebut sesuai untuk proses data mining. Proses ini merupakan proses kreatif dan sangat tergantung pada jenis atau pola informasi yang akan dicari dalam basis data.

4. *Data Mining*

- Pemilihan tugas data mining, pemilihan goal dari proses KDD misalnya klasifikasi, regresi, clustering, dll.
- Pemilihan algoritma data mining untuk pencarian (searching).
- Proses Data mining yaitu proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Teknik, metode, atau algoritma dalam data mining sangat bervariasi. Pemilihan metode atau algoritma yang tepat sangat bergantung pada tujuan dan proses KDD secara keseluruhan.

5. *Interpretation/Evaluation*

- Penerjemahan pola-pola yang dihasilkan dari data mining.
- Pola informasi yang dihasilkan dari proses data mining perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan.