**BAB I**

**PENDAHULUAN**

**1.1 Latar Belakang**

Provinsi sumatera selatan merupakan salah satu provinsi di Indonesia yang menerapkan kebijakan memberikan pengobatan gratis kepada rakyat miskin. kebiaksanaan yang dikenal dengan program berobat gratis, dimana setiap warga masyarakat sumsel yang miskin dapat berobat secara gratis. Program Pengobatan gratis pemerintah provinsi sumatera selatan berupa pemberian asuransi untuk rakyat miskin (askin), agar pasien yang kurang mampu dapat berobat sebagaimana mestinya dan mendapatkan perawatan yang layak. Memang dengan adanya program tersebut, maka pembelanjaan daerah (APBD) pemerintah provinsi sumatera selatan harus memberikan dana anggaran yang lebih dalam bidang kesehatan agar program berobat gratis di sumatera selatan tetap berjalan karena memang program tersebut sangat membantu warga miskin yang ada di sumatera selatan.

Besarnya anggaran yang diperlukan untuk mendukung program berobat gratis tersebut, mau tak mau membuat pemerintah provinsi sumatera selatan harus melakukan perencanaan yang tepat dan baik agar anggaran yang besar tersebut memang tepat sasaran dan sesuai dengan kebutuhan. Untuk melakukan perencanaan yang tepat dan baik, maka pemerintah sumatera selatan harus mempunyai data yang yang akurat, berupa data jumlah diagnosa penyakit pasien pengguna askin yang termasuk dalam program berobat gratis. Dengan menerapkan ukuran prioritas, diharapkan anggaran yang akan dicairkan nantinya untuk berobat gratis tersebut, memang sesuai dengan apa yang dibutuhkan dan tepat sasaran berdasarkan penyakit yang sring diderita oleh pasien pengguna askin. Maka pemerintah sumatera selatan harus bekerja sama dengan seluruh rumah sakit di Palembang yang terdapat program askin, untuk memiliki data yang akurat mengenai jumlah diagnosa penyakit pasien pengguna askin. Jika data jumlah diagnosa penyakit pasien pengguna askin yang didapat telah akurat, maka pemerintah sumatera selatan dapat memberikan dana yang efesien baik secara finansial dan tepat sasaran maupun sesuai dengan kebutuhan, berupa penambahan kamar, fasilitas kesehatan berupa alat-alat medis obat-obatan maupun yang lainnya yang dibutuhkan sesuai dengan apa yang dibutuhkan oleh pasien pengguna askin.

Sebagai salah satu rumah sakit rujukan di sumatera selatan, RSUP Dr. Mohammad Hoesin Palembang memliki data tentang pasien pengguna askin yang jauh lebih banyak dari rumah sakit lainnya melalui cara mengolah data untuk menghasilkan informasi. Dengan meningkatkannya pasien pengguna askin, maka data jumlah diagnosa penyakit pasien pengguna askin meningkat pula. Untuk menghasilkan informasi yang bermanfaat untuk data yang bermanfaat untuk data perencanaan anggaran, maka harus dilakukan pengelompokkan diagnosa jumlah penyakit untuk menemukan diagnosa penyakit mana yang paling sering diderita oleh pasien pengguna askin

Untuk melakukan pengelompokkan jumlah diagnosa penyakit pasien pengguna asking tersebut, maka teknologi yang dibutuhkan adalah data mining. Dimana data mining adalah solusi dalam dunia teknologi untuk mengatasi masalah yang dihadapi rumah sakit dalam memberikan informasi yang tepat dan akurat serta yang efesien kepada yang membutuhkan informasi yang tersebut,

Data mining dalam prosesnya menggunakan teknik statistik, matematika, kecerdasan buatan, dan machine learning untuk mengektraksi dan mengidentifikasi informasi yang bermanfaat serta pengetahuan yang terkait dari berbagai database yang besar (Turban,dkk. 2005).

Dengan adanya masalah tersebut serta ada solusi untuk mengatasi keadaaan seperti itu, maka penulis tertarik untuk melakukan penerapan data mining terhadap jumlah pasien yang menggunakan askin pada RSUP Dr. Mohammad Hoesin. Untuk mengimplementasikan ketertarikan penulis terhadap masalah tersebut, maka penulis akan menuangkan penelitiannya dalam bentuk penulisan skripsi yang berjudul **“ Penerapan Data Mining Pada Rsup Dr.Moh Hoesin Sumatera Selatan Untuk Pengelompokkan Hasil Diagnosa Pengguna Asuransi Kesehatan Miskin (ASKIN)”.**

**1.2 Perumusan Masalah**

Berdasarkan penjelasan yang telah diuraikan diatas tadi, maka pokok permasalahan yang akan dirumuskan dalam penelitian ini adalah bagaimana menerapkan teknologi data mining yang dapat menghasilkan informasi mengenai pengelompokkan hasil diagnosa pasien yang menggunakan asuransi kesehatan miskin (askin).

**1.3 Batasan Masalah**

Agar penlitian ini terfokus pada pokok permasalahannya, maka penulis berinisiatif untuk membatasi permasalahan hanya pada :

a. Implementasi data mining pada RSUP Dr. Mohammad Hoesin

b. Informasi yang ditampilkan berupa pengelompokkan hasil diagnosa jenis penyakit yang sering diderita oleh pasien pengguna askin, berdasarkan hubungan data antara jenis diagnosa penyakit dengan jumlah pasien pengguna askin pada RSUP Moh.Hoesin Palembang.

**1.4 Tujuan Penelitian Dan manfaat penelitian**

**1.4.1 Tujuan Penelitian**

Penelitian ini mempunyai tujuan untuk menjelaskan implementasi data mining serta menampilkan informasi pengelompokkan Hasil diagnosa penyakit pengguna askin menggunakan teknik clustering dan algoritma K-means menggunakan aplikasi weka.

**1.4.2 Manfaat Penelitian**

Dengan adanya penerapan data mining ini diharapkan nantinya akan memberikan manfaat pada RSUP Dr.Moh Hoesin Palembang.

1. Pihak Internal

Pihak Internal yang dimaksud adalah RSUP Dr.Moh Hoesin Palembang yang dapat dapat memiliki laporan penyakit-penyakit yang banyak diderita oleh pasien pengguna askin melalui pemanfaatan teknologi data mining tersebut, sehingga nantinya akan membantu pihak RSUP Dr.Moh Hoesin Palembang dalam hal perencanaan berupa penambahan kamar maupun fasilitas kesehatan seperti obat-obat’an maupun alat-alat medis lainnya dimana penambahan tersebut sesuai dengan kebutuhan pasien pengguna askin.

2. Pihak Eksternal

Pihak Eksternal yang dimaksud dapat termasuk yaitu pemerintah sumatera selatan, dinas kesehatan, maupun masyarkarakat sumsel yang miskin. dimana dengan adanya data pengelompokan diagnosa jumlah penyakit pasien pengguna askin, pihak tersebut dapat dihara pkan pemerintah sumatera selatan dapat melakukan efesiensi anggaran APBD dengan penambahan fasilitas kesehatan mauun obat-obatan perihal penyakit apa saja yang sering diderita oleh pengguna askin. Melalui efesiensi pemberian dana APBD sesuai dengan apa yang dibutuhkan oleh pasien pengguna askin serta kemudian melakukan usaha pencegahan sejak dini dengan cara penyuluhan kepada masyrakat sumsel yang miskin sehingga pemerintah sumatera selatana dan masyarkat sumsel yang miskin mampu bekerja sama untuk meminimalisir penyakit yang umumnya diderita oleh masyarakat sumsel yang miskin sehingga akan selaras dengan visi dan misi pemerintah sumatera selatan untuk menciptakan sumsel yang gemilang.

**1.5 Metodologi Penelitian**

**1.5.1 Waktu dan Tempat**

Waktu penelitian ini dimulai pada april 2013 sampai dengan Juni 2013. Sedangkan tempat penelitian adalah RSUP Dr. Mohammad Hoesin Palembang yang beralamat di Jalan jenderal Sudirman Kilometer 3.5 Palembang 30126.

**1.5.2 Alat dan Bahan**

1. *Hardware*

Kebutuhan perangkat keras (*hardware*) yang diperlukan untuk melakukan penelitian ini adalah *Laptop, Printer, Scanner* dan *Camera*

1. *Software*

Kebutuhan perangkat lunak yang diperlukan dalam penelitian ini adalah sebagai berikut :

1. *Microsoft Windows Xp-3 Profesional,* sebagai sistem operasi.
2. *Microsoft Google Chorome,* sebagai *browser.*
3. *MySQL,* sebagai aplikasi basis data.
4. *Microsoft Office,* sebagai aplikasi pengolahan data untuk penulisan penelitian.

**1.5.3 Metode Penelitian**

 Dalam penelitian ini, penulis menjelaskan penerapan teknologi data mining pada RSUP Dr.Moh Hoesin Palembang untuk mengetahui penyakit-penyakit mana saja yang banyak diderita oleh pasien pengguna askin pada RSUP Dr.Moh Hoesin Palembang.

**1.5.4 Metode Pengumpulan Data**

 Data yang diperlukan dalam penelitian ini adalah sebagai berikut :

**1** Data laporan jumlah diagnosa penyakit pasien pengguna askin pada RSUP Dr.Moh Hoesin Palembang periode mulai dari tahun 2010 sampai dengan tahun 2012.

2. Studi pustaka, Untuk mendukung agar penelitian ini dapat berjalan sesuai dengan sebagaimana mestinya maka dibutuhkan data pendukung untuk penelitian ini melalui buku-buku serta referensi lainnya yang berkaitan dengan judul penlitian ini

**1.5.5 Metode Analisis Data**

 Adapun untuk menganalisis data dalam penerapan *data mining* ini menggunakan tahapan *Knowledge Discovery in Databases (KDD)* yang terdiri dari beberapa tahapan, yaitu *data selection, preprocessing, transformation, data mining,* dan *evaluation.*

**1.6 SISTEMATIKA PENULISAN**

Sistematika ini secara garis besar dapat memberikan gambaran isi, yang berupa susunan bab dari penelitian

**BAB I PENDAHULUAN**

Dalam bab ini penulis menguraikan latar belakang, perumusan masalah, batasan masalah, tujuan dan manfaat, metodologi penulisan laporan, serta sistematika penulis.

**BAB II LANDASAN TEORI**

Bab ini membahas tentang pengertian, istilah, dan teori–teori pendukung yang digunakan untuk menguraikan dan menjelaskan mengenai penerapan *data mining* yang dilakukan penulis.

**BAB III GAMBARAN UMUM**

Bab ini menjelaskan secara singkat mengenai Sejarah, Visi dan Misi, Profil dan Struktur RSUP Dr.Moh Hoesin Palembang.

**BAB IV HASIL DAN PEMBAHASAN**

Bab ini membahas dan menjelaskan hasil dari proses *data mining* yang dilakukan dengan menguraikan teknik clustering dan algoritma k-means menggunakan aplikasi weka.

**BAB V KESIMPULAN DAN SARAN**

Pada bab terakhir ini penulis akan membuat dan mengambil kesimpulan dari pembahasan sebelumnya dan mencoba untuk mengutarakan saran yang mungkin dapat dijadikan bahan pertimbangan bagi RSUP Dr.Moh Hoesin Palembang. dalam pengambilan keputusan.

**BAB II**

**TINJAUAN PUSTAKA**

**2.1 Data mining**

*Data mining* adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan di dalam database. Data mining adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan machine learning yang mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terakit dari berbagai *database* besar (turban, dkk. 2005).

 Selain itu juga *Data mining* disebut sebagai suatu proses untuk menemukan hubungan yang berarti, pola, dan kecenderungan dengan memeriksa dalam sekumpulan besar data yang tersimpan dalam penyimpanan dengan menggunakan teknik pengenalan pola seperti teknik statistik dan matematika (Larose, 2005).

 Selain definisi di atas, ada juga beberapa definisi maupun pengertian mengenai data mining seperti yang tertera yang ada dibawah ini :

a. “ *Data mining* adalah serangkaian proses untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui secara manual.” (Pramudiono,2006)

b. “ *Data mining* adalah analisis otomatis dari data yang berjumlah besar atau kompleks dengan tujuan untuk menemukan pola atau kecenderungan yang penting yang biasanya tidak disadari keberadaannya.” (Pramudiono, 2006)

c. “ *Data mining* merupakan analisis dari peninjauan kumpulan data untuk menemukan hubungan yang tidak diduga dan meringkas data dengan cara yang berbeda dengan sebelumnya, yang dapat dipahami dan bermanfaat bagi pemilik data.” (Larose, 2005)

d. “ *Data mining* merupakan bidang dai beberapa bidang keilmuan yang menyatukan teknik dari pembelajaran mesin, pengenalan pola, statistik, database, dan visualisasi untuk penanganan permasalahan pengambilan informasi dari database yang besar.” (Larose, 2005).

**2.2 Tugas Data Mining Secara Umum**

 Sebelum melakukan proses data mining, baiknya mengetahui terlebih dahulu apa yang bisa dilakukan oleh *data mining*, agar apa yang dilakukan nantinya memang sesuai dengan apa yang dibutuhkan serta menghasilkan sesuatu yang sebelumnya tidak diketahui dan bersifat baru serta bermanfaat bagi penggunanya sendiri (Prasetyo Eko, 2012).

 Pada dasarnya data mining mempunyai kegunaan serta tugas untuk mengspesifikasikan pola yang harus ditemukan dalam proses data mining. Secara umum tugas data mining dapat dibagi menjadi dua kategori yaitu :

a. Prediktif

 Tujuan dari tugas prediktif adalah untuk memprediksi nilai dari atribut tertentu berdasarkan pada nilai dari atribut-atribut lainnya. Atribut yang diprediksi umumnya dikenal sebagai target atau variable tak bebas, sedangkan atribut-atribut yang digunakan untuk membuat prediksi dikenal sebagai variabel bebas.

b. Deskriptif

 Tujuan dari tugas deskriptif adalah menurunkan pola-pola (korelasi. Trend, cluster, trayektori, dan anomali) yang meringkas hubungan yang pokok dalam data. Tugas data mining deskriptif sering disebut sebagai penyelidikan dan seringkali memerlukan teknik *postprocessing* untuk validasi dan penjelasan hasil.

**2.2.1 Tugas-Tugas Data mining**

Berdasarkan tugas data mining secara umum, maka data mining mempunyai tugas-tugas yang berkaitan dengan data mining itu sendiri yaitu sebagai berikut (Prasetyo Eko, 2012) :

a. Model prediksi

 Model Prediksi berkaitan dengan pembuatan model yang dapat melakukan pemetaan dari setiap himpunan variabel ke setiap targetnya, kemudian menggunakan model tersebut untuk memberikan nilai target pada himpunan baru yang didapat. Ada dua jenis model prediksi, yaitu klasifikasi dan regresi. Klasfikasi digunakan untuk variabel target diskret, sedangkan regresi untuk variabel target kontinu.

b. Analisis kelompok

 Analisis kelompok melakukan pengelompokkan data-data ke dalam sejumlah kelompok (cluster) berdasarkan kesamaan karakteristik masing-masing data pada kelompok-kelompok yang ada. Data-data yang masuk dalam batas kesamaan dengan kelompoknya akan bergabung dalam kelompok tersebut, dan akan terpisah dalam kelompok yang berbeda jika keluar dari batas kesamaan dengan kelompok tersebut

c. Analisis asosiasi

 Analisis asosiasi (*association analysis*) digunakan untuk menemukan pola yang menggambarkan kekuatan hubungan fitur dalam data. Pola yang ditemukan biasanya mempresentasikan bentuk aturan implikasi atau subset fitur. Tujuannya adalah untuk menemukan pola yang menarik dengan cara yang efesien.

d. Deteksi Anomali

 Deteksi anomali (*anomaly detection*) berkaitan dengan pengamatan sebuah data dari sejumlah data yang secara signifikan mempunyai karakteristik yang berbeda dari sisa data yang lain. Data-data yang karakteristiknya menyimpang (berbeda) dari data yang lain disebut *outlier.* Algoritma deteksi anomali yang baik harus mempunyai laju deteksi yang tinggi dan laju error yang rendah. Deteksi anomali dapat diterapkan pada sisterm jaringan untuk mengetahui pola data yang memasuki pola data yang memasuki jaringan sehingga penyusupan bisa ditemukan jika pola kerja data yang datang berbeda.

**2.3 *Clustering***

 *Clustering* merupakan pengelompokkan record, pengamatan, atau memperhatikan dan membentuk kelas objek-objek yang memiliki kemiripan. *Cluster* adalah kumpulan *record* yang memiliki kemiripan satu sama lainnya dan memiliki ketidakmiripan dengan *record-record* dalam *cluster* lain.

 Pengclusteran berbeda klasifikasi yaitu tidak adanya variabel target dalam pengclusteran. Pengclusteran tidak mencoba untuk melakukan klasifikasi, mengestimasi, atau memprediksi nilai dari suatu variabel target. Akan tetapi, algoritma pengclusteran mencoba untuk melakukan pembagian terhadap keseluruhan data menjadi kelompok-kelompok yang memiliki kemiripan (Homogen), yang mana kemiripan record dalam suatu kelompok akan bernilai maksimal, sedangkan kemiripan dengan record dalam kelompok lain akan bernilai minimal.

**2.3.1 Contoh *Clustering***

 Beberapa contoh pengclusteran dalam dunia bisnis dan penelitian adalah sebagai berikut :

a. *Clustering* untuk mendapatkan kelompok-kelompok konsumen untuk target pemasaran dari suatu produk bagi perusahaan yang tidak memiliki dana pemasaran yang besar.

b. *Clustering* untuk tujuan audit akutansi, yaitu melakukan pemisahan terhadap perilaku finansial yang baik maupun yang termasuk mencurigakan.

c. *Clustering* terhadap ekspresi dari gen, untuk mendapatkan kemiripan perilaku dari gen dalam jumlah besar

**2.4 Algoritma *K-means***

 Dalam statistik dan mesin pembelajaran *K-means* merupakan metode analisis kelompok yang mengarah pada permatisian N objek pengamatan dalam K kelompok (cluster) dimana setiap objek pengamatan dimiliki oleh sebuah kelompok dengan mean (rata-rata) terdekat.

 K-means merupakan salah satu metode pengelompokkan data nonhierarki (sekatan) yang berusaha mempartisi data yang ada ke dalam bentuk dua atau lebih kelompok. Metode ini mempartisi data kedalam sebuah kelompok sehingga data berkarakteristik sama dimasukkan kedalam satu kelompok yang sama dan data yang berkarakteristik berbeda dikelompokkan ke dalam kelompok yang lain. Adapun tujuan pengelompokkan data ini adalah untuk meminimalkan fungsi objektif yang diset dalam proses pengelompokkan, yang pada umumnya berusaha meminimalkan variasi di dalam suatu kelompok dan memaksimalkan variasi antar kelompok.

 Pengelompokkan dengan K-means mempunyai tahapan-tahapan sebagai berikut :

a. Menentukan jumlah kelompok.

b. Mengalokasikan data kedalam kelompok secara acak.

c. Menghitung pusat kelompok(sentroid/rata-rata) dari data yang ada di masing-masing kelompok.

d. Mengalokasikan masing-masing data ke sentroid/rata-rata terdekat.

e. Kembali kelangkah 3, apabila masih ada data yang berpindah kelompok, atau apabila ada perubahan nilai sentroid diatas nilai ambang yang ditentukan, atau apabila perubahan nilai pada fungsi objektif yang digunakan masih diatas nilai ambang yang ditentukan.

**2.4 Metode *Knowledge Discovery in databases (KDD)***

*Knowledge Discovery in Databases (KDD)* merupakan keseluruhan proses konversi data mentah menjadi pengetahuan yang bermanfaat yang terdiri dari serangkaian tahap transformasi meliputi data *preprocessing* dan *postprocessing.* Dimana data *preprocessing* merupakan langkah awal untuk mengubah data mentah menjadi format yagn sesuai utnuk tahap analisis selanjutnya. Data *preprocessing* kemungkinan akan membutuhkan waktu yang sangat lama, hal ini dikarenakan data yang mentah kemungkinan disimpan dengan format dan *database* yang berbeda. Sedangkan *Postprocessing* meliputi semua operasi yang harus dilakukan agar hasil data mining dapat diakses dan lebih mudah diinpretasikan oleh para pihak analisis (Han & Kamber, 2001).

 Memang sering kali terdengar bahwa *Knowledge Discovery in Databases* mempunyai ikatan yang ikatan sangat kuat dengan data mining, bahkan data mining mempunyai nama lain yaitu *Knowledge Discovery in Databases* (KDD). Memang anggapan itu benar adanya, mengingat tujuan data mining itu sendiri. Namun dalam arti sebenarnya *data mining* dan *knowledge in databases* tidaklah sama karena data mining merupakan bagian proses dalam *Knowledge Discovery in Databases* itu sendiri (KDD), walaupun begitu memang *data mining* merupakan proses yang terpenting dalam proses *Knowledge Discovery in Databases* (KDD). Dan hal itu bisa dilihat pada gambar 2.3 (Han & Kamber, 2001) untuk lebih menjelaskan proses *Knowledge Discovery in Databases*

 Gambar 2.1 Proses tahapan-tahapan *Knowledge Discovery in Databases* (KDD)

Untuk lebih lebih mengetahui mengenai tahapan- tahapan proses pada knowledge in databases, berikut penjelasan mengenai tahapan-tahapan tersebut :

1. *Database*

Koleksi data yang saling berhubungan untuk dipergunakan secara bersama kemudian dirancang untuk memenuhi kebutuhan informasi organisasi

2. *Data Cleaning*

Pada umumnya, data yang diperoleh, baik dari *database* suatu perusahaan maupun eksperimen, memiliki isian-isian yang tidak sempurna seperti data yang hilang, data yang tidak valid atau hanya sekedar salah ketik. Data-data yang tidak relevan itu juga lebih baik dibuang karena keberadaanya bsia mengurangi mutu atau akurasi dari hasil data mining nantinya. Pembersihan data juga akan mempengaruhi performasi dari system data mining karena data yang akan ditangani akan berkurang jumlah dan kompleksitasnya.

3. *Data integration*

Integrasi data dilakukan pada attribut-attribut yang mengidentifikasikan *entitas-entitas* yang unik seperti attribut nama,nomor pegawai, tempat lahir,agama dan lain sebagainya. Integrasi data perlu dilakukan secara cermat karena kesalahan pada integrasi data bisa menghasilkan hasil data yang menyimpang dan bahkan menyesatkan pengambilan aksi nantinya.

4. *Task relevan data*

Setelah semua sumber data digabung atau diintegrasikan menjadi satu keseluruhan *database*, maka tahap selanjutnya adalah melakukan task *relevant data*. Pada tahap ini melakukan relevansi attribut dari data yang *relevant* atau yang sesuai dengan target atau *output* yang akan dihasilkan.

 5. *Data transformation*

*Coding* adalah proses transformasi pada data yang telah dipilih, sehingga data tersebut sesuain untuk proses data mining. Proses *coding* dalam KDD merupakan proses kreatif dan sangat tergantung pada jenis atau pola informasi yang akan dicari dalam *database*.

6. *Data mining*

*Data mining* adalah proses mengeksplorasi dan menganalisa data dalam jumlah yang besar yang bertujuan untuk menemukan suatu pola atau informasi yang menarik dari data yang tersimpan dalam jumlah yang besar dengan menggunakan teknik atau metode tertentu. Teknik, metode, atau algoritma yang tepat sangat bergantung pada tujuan dan proses KDD secara keseluruhan. Tahap ini merupakan inti dari tahapan KDD yang dilakukan untuk menganalisis data yang telah dibersihkan.

7. *Pattern evaluation*

Dalam tahap ini, merupakan hasil dari teknik data mining berupa pola-pola yang khas maupun model prediksi dievaluasi untuk menilai apakah hipotesa yang ada memang tercapai. Bila ternyata hasil yang diperoleh tidak sesuai dengan hipotesa, ada beberapa alternatif yang dapat diambil seperti menjadikannnya umpan balik untuk memperbaiki *data mining* lain yang lebih sesuai, atau menerima hasilnya sebagi suatu hasil yang diluar dugaan yang mungkin bermanfaat.

8. *Knowledge*

Tahap terakhir dari proses *data mining* adalah bagaimana memformulasikan keputusan atau aksi dari hasil analisa yang didapat. Ada kalanya hal ini harus melibatkan orang-orang yang tidak paham mengenai *data mining*. Karenanya presentasi hasil *data mining* dalam bentuk pengetahuan yang bisa dipahami semua orang adalah atu tahapan yang diperlukan dalam proses *data mining*. Dalam presentasi ini, *visualisasi* juga bisa membantu mengkomunikasikan hasil data mining.

**2.5 Weka**

Weka adalah aplikasi data mining open source yang berbasis java. Aplikasi ini dikembangkan pada tahun 1994 dan pertama kali oleh sebuah universitas di selandia baru yang bernama universitas Waikato. aplikasi weka ini mulai menjadi aplikasi *data mining open source* yang sangat terkenal pada awal perkembangannya. Hal itu dikarenakan aplikasi weka memiliki kelebihan yang tidak dimiliki aplikasi *data mining* lainnya yaitu pada aplikasi weka terdapat banyak algoritma yang terdapat didalam aplikasi dan disertai juga *machine learning*, lalu juga dalam penggunaanya tidak terlalu rumit sehingga tidak menyulitkan penggunanya, dan ditambah dengan kelebihan lainnya bahwa algoritma-algoritma yang terdapat pada aplikasi weka selalu baru dan ter *update*, sehingga dengan beberapa kelebihan aplikasi weka tersebut digunakan banyak perusahaan dalam dunia bisnis untuk membantu dalam usaha bisnisnya, akademik pun juga tak mau ketinggalan untuk menggunakan aplikasi weka ini, serta instansi dalam bidang kesehatan yaitu rumah sakit juga saat ini juga menggunakan aplikasi weka.

 Aplikasi weka merupakan *software* yang terdiri dari koleksi algoritma *machine learning* yang dapat digunakan untuk melakukan generalisasi atau formulasi dari sekumpulan data *sampling*. Inti dari kekuatan pada aplikasi weka terletak pada algoritma yang makin lengkap dan canggih, namun walaupun begitu canggihnya aplkasi weka tersebut, letak keberhasilan *data mining* tetap ditentukan oleh manusia itu sendiri sebagai penggunanya/*user*. Keberhasilan *data mining* itu berdasarkan pada pengumpulan data yang berkualitas tinggi, penggunaan model dan algoritma yang tepat. Sehingga secanggih serta sehebat apapun aplikasi *data mining*, tanpa kemampuan sang penggunanya untuk menerapkannya maka tidak akan menghasilkan data mining yang tepat dan bermanfaat.

**2.6 Analisis Data Mining**

 Analisis data mining merupakan proses yang melakukan pengumpulan bentuk dan relasi dari sekumpulan data yang besar tanpa harus melakukan percobaan hipotesa yang tidak menggunakan rumus. Juga diesbut sebagai proses pembentukan data yang mentah menjadi data yang dibutuhkan agar dapat melakukan proses mining pada data tersebut.

**2.6.1 Data selection**

 Pada data selection merupakan langkah awal dalam penelitian untuk menggunakan data yang dibutuhkan dalam melakukan proses mining , dimana data yang dibutuhkan merupakan data yang berasal dari tahun RSUP Dr.moh Hoesin Palembang yang merupakan tempat penelitian selama satu tahun yang terdiri dari dua tabel yaitu tabel hasil Diagnosa penyakit serta tabel Jumlah Diagnosa penyakit per transaksi yang memiliki penjelasan sebagai berikut :

a. Data Hasil Diagnosa Pasien, terdapat tabel hasil diagnosa penyakit yang berisi informasi hasil diagnosa penyakit yang diderita oleh pasien pengguna askin, dan itu dapat dilihat melalui tabel 4.2:

Tabel 2.1 Atribut Hasil Diagnosa

|  |  |
| --- | --- |
| Atribut | Keterangan |
| Kd\_Diagnosa | Berupa primary key dari tabel hasil diagnosa yang digunakan dalam relasi terhadap tabel jumlah diagnosa pernyakit per transaksi |
| Nama\_Diagnosa | Berupa informasi mengenai jenis asuransi yang dipilih. |

b. Data Jumlah Diagnosa penyakit per transaksi, terdapat tabel jumlah diagnosa penyakit per transaksi yang berisi informasi jumlah diagnosa penyakit yang diderita pasien pengguna askin transaksi per bulannya, dan hal itu dapat dilihat melalui tabel 4.3 :

Tabel 2.2 Atribut Jumlah Diagnosa penyakit per transaksi

|  |  |
| --- | --- |
| Atribut | Keterangan |
| Kd\_Diagnosa | Berupa Foreign key digunakan dalam relasi terhadap Jumlah Diagnosa penyakit per transaksi |
| TglBuktiaskin | Berupa informasi mengenai tanggal transaksi klaim asuransi kesehatan orang miskin (Askin) |
| Kd\_Pasien | Sebagai Foreign key,dimana nantinya untuk mengetahui pasien askin yang melakukan transaksi |
| Bulan | Berupa informasi mengenai jumlah hasil diagnosa penyakit per bulannya |
| Jumlah | Berupa informasi mengenai total keseluruhan hasil diagnosa penyakit yang di akumulatifkan berdasarkan jumlah Diagnosa penyakit per transaksi pada tiap bulannya |

 Lalu setelah mengetahui beberapa tabel tersebut beserta attribute yang terdapat didalamnya, maka untuk melakukan proses mining sesuai dengan tahapan knowledge discovery in databases maka dilakukan pemilihan beberapa attribute yang dibutuhkan adalah sebagai berikut :

a. Kd\_Diagnosa merupakan atribut yang terdapat pada ketiga tabel tersebut yaitu Tabel pasien, Tabel hasil diagnosa, serta tabel jumlah diagnosa penyakit per transaksi, dimana atribut ini berperan sebagai foreign key serta menjadi primary key pada tabel hasil diagnosa.

b. Nama\_Diagnosa merupakan atribut yang terdapat pada tabel hasil\_diagnosa. Dimana atribut ini digunakan untuk mengetahui nama hasil diagnosa penyakit yang diderita oleh pasien pengguna askin nantinya.

c. Bulan merupakan atribut yang terdapat pada tabel Jumlah Diagnosa penyakit per transaksi, dimana pada atribut bulan ini dapat terlihat jumlah hasil diagnosa per transaksi hasil diagnosa selama satu bulan.

d. Jumlah merupakan jumlah atribut yang terdapat pada tabel jumlah hasil diagnosa per transaksi yang dimana jumlah tersebut merupakan hasil selama satu tahun yang didapat dari hasil akumulatif jumlah hasil diagnosa per bulannya.

**2.6.2 Preprocessing**

Pada tahapan ini akan dilakukan proses integrasi data untuk menghubungkan tabel hasil diagnosa dan tabel jumlah diagnosa penyakit per transaksi dan proses ini juga disebut juga preprocessing. Yang dimana nantinya hal itu bertujuan untuk mendapatkan dataset yang bersih sehingga nantinya data tersebut dapat digunakan dalam tahap yang terpenting dalam proses ini yaitu melakukan proses mining pada data tersebut dengan mempunyai tujuan untuk memperoleh pengelompokkan hasil diagnosa pasien pengguna askin. Berikut penjelasan dari kedua proses tersebut :

a. Integrasi data, merupakan tahap integrasi data yang mempunyai definisi sebagai tahapa penggabungan data dari beberapa sumber. Dan proses integrasi data tersebut dapat dilihat pada gambar 2.2 berikut :

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Data Source**(Data Hasil Diagnosa)

|  |
| --- |
| Atribut |
| Kd\_Diagnosa |
| Nama\_Diagnosa |

**Data Source**(Data jumlah hasil diagnosa per transaksi)

|  |
| --- |
| Atribut |
| Kd\_Diagnosa |
| TglBuktiAskin |
| Kd\_Pasien |
| Bulan  |
| Jumlah |

 | **Data Warehouse**(Integrasi Data)

|  |
| --- |
| Atribut |
| Kd\_Diagnosa |
| Nama\_Diagnosa |
| Kd\_Diagnosa |
| Tglbuktiaskin |
| Kd\_Pasien |
| Bulan |
| Jumlah |
| UP |

 | **Data Target**

|  |
| --- |
| Atribut |
| Kd-Diagnosa |
| Nama\_Diagnosa |
| Bulan |
| Jumlah |

 |

Gambar 2.2. Integrasi Data

Pada Gambar 4.1 menggambarkan proses penggabungan dua data source yaitu tabel hasil diagnosa dan jumlah hasil diagnosa per transaksi hingga didapatkan data yang sesuai dengan kebutuhan dalam melakukan proses data mining

Gambar 2.3 Hasil data set

b. Data Cleaning, tahap data cleaning merupakan tahap awal dari proses KDD. Dimana pada tahap ini data yang tidak relevan, missing value, dan redundant harus dibersihkan. Hal ini bertujuan agar data yang tidak relevan dan tidak redundant merupakan syarat awal dalam melakukan data mining. Dengan pengertian bahwa data dikatakan missing value apabila jika terdapat atribut dalam dataset yang tidak berisi nilai atau kosong, sedangkan data yang yang bersifat redundant jika dalam satu data set terdapat lebih dari satu record yang berisi nilai yang sama.

2.6.3 Tranformasi data

 Tahapan berikutnya adalah melakukan proses tahapan transformasi data yang merubah data ke dalam bentuk yang sesuai untuk di mining. Perubahan awal yang dilakukan adalah menambah 14 atribut yaitu janutari,februari,maret,april,mei,juni,juli,agustus,September,oktober,november,desember sebagai penggant atribut sebelumnya yaitu atribut bulan.dan proses tranformasi data tersebut dapat dilihat pada gambar berikut :

|  |  |
| --- | --- |
|  |  |
| **Data Target**

|  |
| --- |
| Atribut |
| Kd\_Diagnosa |
| Nama\_Diagnosa |
| Bulan |
| Jumlaj |

 |

|  |
| --- |
| Januari |
| Februari |
| Maret |
| April |
| Mei |
| Juni |
| Juli |
| Agustus |
| September |
| Oktober |
| November |
| Desember |

|  |
| --- |
| Bulan |

|  |
| --- |
| Jumlah |
| Nama\_Diagnos |
| Januari |
| Februari |
| Maret |
| April |
| Mei |
| Juni |
| Juli |
| Agustus |
| September |
| Oktober |
| November |
| Desember |

 |
|  |  |
|  |  |

Gambar 2.4 Proses Transformasi

Dari proses yang dilakukan diatas maka diperoleh hasilnya pada gambar dibawah ini :

Gambar 2.5 Dataset Hasil transformasi data

**BAB III**

**GAMBARAN UMUM**

**3.1 Sejarah RSUP dr. Mohammad Hoesin**

RSUP dr. Mohammad Hoesin Palembang didirikan pada tahun 1953 atas prakarsa Menteri Kesehatan RI dr. Mohammad Ali (Dr. Lee kiat Teng) dengan biaya pemerintah pusat pada tanggal 03 Januari 1957 rumah sakit ini dimulai operasional yang dapat melayani masyarakat se-Sumbagsel di mana saat itu meliputi provinsi se-Sumatera Selatan, Lampung, Jambi, Bengkulu dan Bangka Belitung.

Pada awalnya RSUP dr. Moh. Hoesin Palembang baru memiliki fasilitas yang sederhana yaitu berupa pelayanan rawat jalan dan rawat inap (fasilitas78 tempat tidur). Lalu kemudian RSUP Palembang resmi menggunakan nama RSUP Dr.Mohammad hoesin Palembang pada tanggal 4 oktober 1997, berdasarkan SK menteri kesehatan RI No : 1297/ Menkes/ SK/ XI/ 1997. Beberapa waktu kemudian RSUP Dr.Mohammad hoesin Palembang memiliki pelayanan Laboratorium, Apotek, Radiologi, Emergency dan peralatan penunjang lainnya, seiring dengan perkembangan waktu, rumah sakit ini semakin berkembang, baik fasilitas, sarana dan prasarana Sumber daya manusianya tersedia para spesialis lengkap dan beberapa sub spesialis sehinga mengubah tipenya dari kelas C menjadi Rumah Sakit Umum Pusat tipe B dan menjadi rumah sakit terbesar dan sebagai pusat rujukan layanan kesehatan se-Sumatera Selatan,Jambi, Bengkulu, Lampung, dan Bangka Belitung.

Tahun 1993–1994 RSUP Palembang mengubah status dari RS Vertikal (RS Penerima Negara Bukan Pajak) menjadi RS Swadana sesuai SK. Menkes RI No.1279/Menkes/SK/XI/1997 RSUP Palembang resmi bernama RSUP dr. Mohammad Hoesin Palembang. Dengan UU No. 20/1997 menjadi Rumah Sakit Instansi Pengguna PNBP, dimana rumah sakit dapat memanfaatkan dana dari hasil pendapatan sesuai dengan anggaran yang diproyeksikan rumah sakit dan diselaraskan dengan pendapatan melalui prosedur KPKN disamping itu subsidi Pemerintah tetap seperti sediakala.

 Tahun 2000 dengan PP No 122/2000, RSUP dr. Mohammad Hoesin Palembang ditetapkan menjadi salah satu dari 13 Rumah Sakit Pemerintah menjadi Rumah Sakit Perusahaan Jawatan di Indonesia dan operasionalnya dimulai tanggal 01 Januari 2002, sebagai Rumah Sakit Perjan ( perusahaan jawatan ) secara operasional RSMH Palembang masih tetap melaksanakan fungsi pelayananan bagi masyarakat ekonomi kurang mampu melalui Program ASKESKIN.

**3.2 Visi Dan Misi**

**3.1.1 Visi**

 Sebagai salah satu rumah sakit yang terbesar di provinsi sumatera selatan yang pastinya ingin mmberikan pelayanan kesehatan yang baik dan bermutu bagi masyarakat, maka RSUP Dr.Mohammad hoesin Palembang memliki visi yaitu “Menjadi rumah sakit pusat layanan kesehatan, pendidikan dan penelitian terbaik dan bermutu se-sumatera”

**3.1.2 Misi**

 Sebagai salah satu rumah sakit yang terbesar di Sumatera selatan, RSUP Dr.Mohammad hoesin Palembang juga memliki misi sebagai berikut:

1. Menyelenggarakan pelayanan kesehatan yang komprehensif dan berkualitas tinggi.

2. Menyelenggarakan jasa pendidikan dan penelitian dalam bidang kedokteran dan kesehatan.

3. Menjadi pusat promosi

**3.3 Tujuan berdiri RSUP Dr.Mohammad Hoesin Palembang**

Selain itu juga RSUP Dr.Mohammad hoesin Palembang juga mempunyai tujuan sebagai berikut

1. Meningkatkan derajat kesehatan dan senantiasa berorientasi kepada kepentingan masyarakat.

2. Meningkatkan citra pelayanan pemerintah kepada masyrakat di bidang kesehatan.

3. menghasilkan tenaga dokter umum, spesialis dan sub spesialis serta tenaga keperawatan yang berkualitas dan bermoral tinggi.

**3.4 Struktur Organisasi RSUP Dr.Mohammad hoesin Palembang**

 Direktur Utama

Direktur Pelayanan Publik Direktur Keuangan Dir. Umum,SDM

Sumber : http ://www.rsmh.co.id, 2013

**Gambar 3.1 Strukutur Organisasi Rumah sakit Moh.Hoesin Palembang**

**3.5 Motto Dan Budaya**

RSUP Dr.Mohammad hoesin Palembang sebagai salah satu rumah sakit yang ada di provinsi sumatera selatan berkomitmen penuh untuk melayani masyarakat di bidang kesehatan sehingga RSUP Dr.Mohammad hoesin Palembang mempunyai motto yaitu “KESEHATAN DAN KEPUASAN ANDA MERUPAKAN KEBAHAGIAAN KAMI”. Dan jelas melalui motto tersebut RSUP Dr.Mohammad hoesin Palembang, sangat memperhatikan mutu dan kualitas pelayanan mereka terhadap masyrakat.

Waktu yang telah dijalani RSUP Dr.Mohammad hoesin Palembang mulai dari awal berdiri sampai dengan saat ini, bukanlah waktu yang singkat untuk sebuah rumah sakit. Boleh dibilang tidak semua rumah sakit bisa bertahan sampai dengan saat ini bila tidak mempunyai budaya yang positif , dan budaya tersebut tetap dipertahankan oleh setiap pegawai yang berada di lingkungan RSUP Dr.Mohammad hoesin Palembang mulai dari pimpinan sampai dengan jajaran dibawahnya. sehingga sampai saat ini RSUP Dr.Mohammad hoesin Palembang tetap ada dan semakin berkembang dari waktu ke waktu, berikut budaya yang diterapkan di lingkungan RSUP Dr.Mohammad hoesin Palembang adalah sebagai berikut :

1. Semua petugas rumah sakit dalam memberikan pelayanan kepada masyrakat harus bersikap ramah tamah dengan menunjukkan air muka yang jernih dan ikhlas.

2. Dalam melaksanakan semua aktifitas di RS, petugas harus selalu melakukan efesiensi untuk mencapau tujuan.

3. Terdapat keserasian atau kerhamonisan dalam kerja sama antar petugas medis, paramedis, dan non medis dengan pasien dan keluarga pasien serta pengguna jasa rumah sakit lainnya.

4. Semua kegiatan pelayanan dan transaksi keuangan harus dapat dipertanggug jawabkan secara akademis dan ilmiah sesuai dengan peraturan perundang-undangan yang berlaku.

5. Ada keterbukaan di segenap aspek kegiatan.

**3.5 Jenis Pembayaran**

 Dalam urusan pembayaran yang dilakukan di RSUP Dr.Mohammad hoesin Palembang, juga terdapat berbagai jenis yang disesuaikan dengan kemampuan serta kebutuhan dari masyrakarakat yang berobat di RSUP Dr.Mohammad hoesin Palembang, berikut jenis pembayaran yang dilayani di RSUP Dr.Mohammad hoesin Palembang :

1. Askes

Askes merupakan badan usaha milik Negara yang ditugaskan khusus oleh pemerintah untuk menyelenggarakan jaminan pemeliharaan kesehatan bagi pegawai negeri sipil, penerima pensiun PNS dan TNI/POLRI, veteran, perintis kemerdekaan beserta keluarganya dan badan usaha lainnya.

2. Askin

Askin merupakan jaminana kesehatan untuk masyarakat yang kurang mampu atau masyarakat miskin yang ditujukan untuk membantu mereka yang berpenghasilan sangat rendah ketika terjadi gangguan kesehatan, misalnya karena sakit atau meninggal.

3. Umum

Umum yaitu pembayaran yang dilakukan oleh pribadi atau tidak ditanggung oleh jaminan apapun, dan biasanya pembayaran jenis ini tidak mendapatkan potongan biaya sedikit pun.

4. Perusahaan

Perusahaan merupakan pembayaran biaya pasien yang dibiayai atau ditanggung dan dijamin oleh suatu perusahaan yang bekerja sama dengan rumah sakit.

5. Asuransi Swasta

Pembayaran ini biasanya untuk pasien yang sebelumnya telah terdaftar pada asuransi yang telah bekerja sama dengan RSUP Dr.Mohammad hoesin Palembang dan besar biaya pembayaran tergantung dengan polis yang di miliki pada asuransi pasien tersebut dan juga berdasarkan jenis penyakit yang diklaim oleh pihak asuransi apabila berobat di RSUP Dr.Mohammad hoesin Palembang.

**BAB IV**

**HASIL DAN PEMBAHASAN**

**4.1 Data Mining**

Pada tahapan ini akan dicari pola atau informasi dalam sekumpulan data dengan menggunakan teknik dan algoritma tertentu. Pada penelitian ini penerapan data mining menggunakan teknik clustering dan algoritma k-means untuk menemukan jumlah diagnosa penyakit pasien pengguna askin.

**4.2 Penerapan Clustering dengan algoritma k-means**

 Setelah melakukan proses transformasi data menjadi data yang sesuai untuk penerapan data mining dengan teknik clustering maka tahapan selanjutnya adalah dengan melakukan pengelompokkan diagnosa penyakit.

 Dalam tahapan pengelompokkan ini, langkah awal yang dilakukan adalah menentukkan jumlah kelompok yang dimana kelompok tersebut dikelompokkan berdasarkan jenis penyakit. Dalam tahapan awal menghasilkan informasi berupa diagnosa penyakit pasien pengguna askin dengan cara mengelompokkan per diagnosa penyakit. Berikut pengelompokkan diagnosa berdasarkan jenis penyakit yang dikelompokkan berdasarkan jenis yang mengikuti sesuai dalam dunia kesehatan dan telah ditentukan berdasarkan sumber <http://upikblogs.blogspot.com/2012/05/klasifikasi-penyakit-kelompok-3.html> :

|  |  |  |
| --- | --- | --- |
| Traumatic (kelompok 0)Curturio CerebriTraumatic Subcutenous  | Infeksi (kelompok 1)Diarrohea gastroenteritis Dengue Haermogavis dengue | Degeneratf (kelompok2)Thalasselmia, UnspecifiedBronchopneunomia |

Tabel 4.1 Contoh Kelompok Jenis Penyakit

|  |  |  |  |
| --- | --- | --- | --- |
| Traumatic (kelompok 0)100 | Infeksi (kelompok 1)4839 | Degeneratf (kelompok2)10806 | Jumlah Pengguna askin15295 |

Tabel 4.2 Jumlah Pengguna askin selama 3 tahun

Kemudian setelah itu maka langkah selanjutnya adalah mengalokasikan dari data yang ada di masing-masing kelompok. Dimana setiap data jumlah diagnosa penyakit ke masing-masing jenis penyakit kemudian akan membentuk data jumlah diagnosa penyakit pasien pengguna askin per tahunnya.

 Lalu kemudian melakukan penghitungan pada pusat masing-masing kelompoknya (sentroid/rata-rata) dari data yang ada di masing-masing kelompok. Pada langkah ini, menentukan lokasi sentroid (titik pusat) setiap kelompok yang diambil dari rata-rata (mean) semua nilai pada setiap fiturnya harus dihitung kembali. Jika M menyatakan jumlah data dalam sebuah kelompok, *i* menyatakan fitur ke-*i* dalam sebuah kelompok, dan *p* menyatakan dimensi data, dan *X* menyatakan matriks set M yang merupakan jumlah data, lalu *j* menyatakan jarak, dan *1* menyatakan nilai di dalam data, maka untuk menghitung sentroid dinyatakan dalam rumus ini :

Tabel 4.3 Rumus menentukan sentroid

 C1=$\frac{1}{M}$ $\sum\_{j=1}^{m}Xj$

 Lalu langkah yang diambil selanjutnya adalah mengalokasikan masing masing data ke sentroid/rata-rata terdekat. Lalu kemudian langkah yang terakhir adalah kembali ke langkah 3, apabila dirasa masih ada data yang berpindah kelompok, atau apabila ada perubahan nilai sentroid diatas nilai ambang yang ditentukan, atau apabila perubahan nilai pada fungsi objektif yang digunakan masih diatas nilai ambang yang ditentukan.

 Proses mining untuk menemukan diagnosa penyakit yang sering diderita oleh pasien pengguna askin berdasarkan hubungan diagnosa penyakit pada pasien pengguna askin, maka proses akan dipilih sesuai dengan apa yang telah ditentukan. Pada contoh ini akan dilakukan pengelompokkan 10 data diagnosa penyakit pada set data pengelompokkan dua dimensi. Dimensi datanya dua fitur dengan tujuan memudahkan dalam visualisasi kordinat catersius nantinya. Fitur yang digunakan dalam pengelompokkan adalah x dan y. lalu akan dilakukan pengukuran jarak, jumlah kelompok (K) adalah 2. Ambang batas atau threshold (T) yang digunakan untuk perubahan fungsi objektif adalah 0.1. Kondisi awal data dapat dilihat pada tabel. Langkah-langkah yang akan dilakukan adalah inisialisasi dan iterasi 1,2,3,4.

1. Inisialisasi

a. Mengalokasikan semua data pada satu kelompok yang telah ditentukan sebelumnya. Setelah dilakukan alokasi data secara acak, masing-masing data akan bergabung ke dalam sebuah kelompok. Hasil alokasi secara acak ini akan ditampilkan pada tabel berikut :

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Data Ke-i | Fitur x | Fitur y | Kelompok 1 | Kelompok 2 | Kelompok 3 |
| 1 | 1 | 1 | \* |  |  |
| 2 | 4 | 1 |  |  | \* |
| 3 | 6 | 1 |  | \* |  |
| 4 | 1 | 2 |  | \* |  |
| 5 | 2 | 3 |  |  | \* |
| 6 | 5 | 3 |  | \* |  |
| 7 | 2 | 5 |  | \* |  |
| 8 | 3 | 5 |  |  | \* |
| 9 | 2 | 6 |  |  | \* |
| 10 | 3 | 8 |  | \* |  |

Tabel 4.4 Inisialisasi

b. Menghitung sentroid (pusat kelompok)

Lokasi sentroid (pusat kelompok) dihitung dengan rata-rata semua data yang bergabung dalam setiap kelompok.

|  |  |  |
| --- | --- | --- |
| Data Anggota | Fitur x | Fitur y |
|  |  |  |
| 2 |  | 1 |
|  |  |  |
|  |  |  |
| 5 |  | 3 |
|  |  |  |
|  |  |  |
| 8 | 3 | 5 |
| 9 | 2 | 6 |
|  |  |  |

 M Jumlah x Jumlah y 4 11 15

 Rata-rata 2,7500 3,7500

Tabel 4.5 Kelompok 0

|  |  |  |
| --- | --- | --- |
| Data Anggota | Fitur x | Fitur y |
|  |  |  |
|  |  |  |
| 3 | 6 | 1 |
| 4 | 1 | 2 |
|  |  |  |
| 6 | 5 | 4 |
| 7 | 2 | 5 |
|  |  |  |
|  |  |  |
| 10 | 3 | 8 |

M Jumlah x Jumlah y

 5 17 19

 Rata-rata 3,4000 3,8000

Tabel 4.6 Kelompok 1

|  |  |  |
| --- | --- | --- |
| Data Anggota | Fitur x | Fitur y |
|  |  |  |
| 2 |  | 1 |
|  |  |  |
|  |  |  |
| 5 |  | 3 |
|  |  |  |
|  |  |  |
| 8 | 3 | 5 |
| 9 | 2 | 6 |
|  |  |  |

M Jumlah x Jumlah y 4 11 15

 Rata-rata 2,7500 3,7500

Tabel 4.7 Kelompok 2

|  |  |  |
| --- | --- | --- |
| Kelompok | Fitur x | Fitur y |
| 1 | 1,0000 | 1,0000 |
| 2 | 3,4000 | 3,8000 |
| 3 | 2,7500 | 3,7500 |

 Tabel 4.8 Sentroid yang didapat

Fungsi Objektif

J lama = 0

J baru = 22,4007

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Data Ke-i | Fitur x | Fitur y | Kelompok 0 | Kelompok 1 | Kelompok 2 |
| 1 | 1 | 1 | 0,0000 |  |  |
| 2 | 4 | 1 |  |  | 3,0208 |
| 3 | 6 | 1 |  | 3,8210 |  |
| 4 | 1 | 2 |  | 3,0000 |  |
| 5 | 2 | 3 |  |  | 1,0607 |
| 6 | 5 | 3 |  | 1,7889 |  |
| 7 | 2 | 5 |  | 1,8439 |  |
| 8 | 3 | 5 |  |  | 1,2748 |
| 9 | 2 | 6 |  |  | 2,3717 |
| 10 | 3 | 8 |  | 4,2190 |  |

 0,0000 14,6728 7,7279

 Fungsi Objektif 22,4006

Tabel 4.9 jarak Data ke sentroid

Perubahan fungsi Objektif = |22.4007-0| = 22,4007 (Masih diatas T), kemudian langkah berikutnya dilakukan sampai dengan kondisi seperti algoritma terpenuhi.

2. Iterasi 2

a. Menghitung kembali jarak terpendek setiap data pada kelompok

terdekat. Jarak data ke setiap sentroid ditampilkan pada kolom 2 sampai 4. Jarak minimal yang dipilih ditampilkan pada kolom Min, sedangkan kelompok berikutnya yang diikuti oleh data ada di kolom kelompok.

 Tabel 4.10 Jarak terpendek setiap data

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Data Ke-i | Kelompok 0 | Kelompok 1 | Kelompok 2 | Min | Kelompok  |
| 1 | 0,0000 | 3,6878 | 3,2596 | 0,0000 | 1 |
| 2 | 3,0000 | 2,8636 | 3,0208 | 2,8636 | 2 |
| 3 | 5,0000 | 3,8210 | 4,2573 | 3,8210 | 2 |
| 4 | 1,0000 | 3,0000 | 2,4749 | 1,0000 | 1 |
| 5 | 2,2361 | 1,6125 | 1,0607 | 1,0607 | 3 |
| 6 | 4,4721 | 1,7889 | 2,3717 | 1,7889 | 2 |
| 7 | 4,1231 | 1,8439 | 1,4577 | 1,4577 | 3 |
| 8 | 4,4721 | 1,2649 | 1,2748 | 1,2649 | 2 |
| 9 | 5,0990 | 2,6077 | 2,3717 | 2,3717 | 3 |
| 10 | 7,2801 | 4,2190 | 4,2573 | 4,2190 | 2 |

**4.3 Proses Data mining menggunakan weka**

 Setelah dijelaskan proses penerapan data mining menggunakan metode clustering secara teoritis pada penjelasan diatas, maka kali ini akan dijelaskan proses data mining secara aplikatif dimana proses data mining yang akan dilakukan menggunakan software data mining weka weka V3.7.9 serta beberapa perangkat lainnya yang berbasis java untuk mendukung jalannya weka di system operasi yang digunakan.

 Seperti yang dijelaskan dalam bab sebelumnya weka merupakan aplikasi data mining open source yang berbasis java. Tampilan awal weka dapat dilihat dalam, gambar berikut

 Gambar 4.1 Tampilan Awal Aplikasi weka

Pada gambar 5.1 merupakan tampilan awal pada aplikasi weka yang dimana terdapar beberapa menu yang mempunyai fungsi masing masing funsgi masing-masing :

a. Explorer digunakan untuk menggali lebih jauh data dengan menggunakan aplikasi weka.

b. Experimenter digunakan untuk melakukan percobaan dengan pengujian statistic skema belajar.

c. knowledge flow digunakan untuk pengetahuan pendukung

d. Simple CLI antar muka dengan menggunakan tampilan command-line yang memungkinkan langusng mengeksekusi perintah weka untuk sisterm operasi yang tidak menyediakan langsung

Setelah mengenal beberapa menu tersebut, maka kemudian untuk melakukan proses data mining menu yang dipilih adalah explorer, setelah itu maka ada beberapa tampilan sub menu untuk dipilih sesuai dengan teknik yang akan diperlukan dan sebelumnya perlu diketahui bahwa kesuksesan implementasi data mining terletak pada factor pengetahuan manusia sebagai implementornya, kemudian pengumpulan data yang berkualitas tinggi dan menguasai permodelan dan penggunaan algoritma yang tepat sangat menentukan keakuratan hasil akhir sesuai dengan apa yang diharapkan. Dan factor terakhir sebelum akan melakuakan implementasi data mining menggunakan aplikasi weka, perlu memperhatikan factor tersebut agar hasil akhir nantinya sesuai dengan apa yang diharapkan.

Kemudian setelah itu, langkah yang selanjutnya adalah untuk menggunakan aplikasi weka melakukan tranformasi data dari data awal berupa data jumlah hasil diagnosa yang didapat sebelumnya dalam format Microsoft Excell 2007 (CSV). Lalu langkah berikutnya adalah memilih menu explorer kemudian pilih menu open file untuk selanjutnya kemudian memilih data yang telah ditranformasi sebelumnya, dan langkah tersebut dapat dilihat melalui gambar berikut :

Gambar 4.2 Menu Explorer

Dari pengaturan awal yang dilakukan, dimana data hasil diagnosa yang telah ditentukan untuk mendapatkan data yang berisi nilai relevan dan tidak bersifat missing value dan redundant. Pengaturan awal tersebut memiliki tujuan agar data yang yang akan diproses pada aplikasi weka telah memenuhi standar weka yang dimana data yang akan diproses ekstensi file telah diubah terlebih dahulu menjadi .csv (Command separated values). Apabila langkah awal tersebut telah dilakukan sesuai dengan petunjuk, maka akan didapat kan hasil sebagai berikut :

Gambar 4.3 Tampilan awal proses

Melalui tampilan tersebut, maka akan dilihat beberapa attribute yang tampil dalam gambar tersebut, berikut penjelasanya :

a. Jenis Penyakit : Merupakan jenis penyakit yang telah dibagi menjadi tiga kelompok melakukan perawatan dan perobatan di RSMH moh.hoesin Palembang.

b. Januari, Februari, dst : Merupakan nama bulan-bulan, dimana dalam setiap bulan tersebut terdapat angka-angka yang menunjukan jumlah per diagnosa selama satu bulan.

c. Jumlah : Merupakan jumlah hasil diagnose pernyakit yang telah diakumulatifkan dalam periode satu tahun.

Lalu langkah selanjutnya adalah menentukan metode serta algoritma, namun langkah ini harus disesuaikan dengan yang menjadi kebutuhan serta sesuai dengan pembahasan penelitian. Dalam hal ini teknik yang dipilih adalah teknik clustering serta algoritma yang dipilih adalah algoritma K-means. Dan proses langkah tersebut dapat dilihat pada gambar berikut :

Gambar 4.4 Tampilan pemilihan algoritma

Algoritma yang digunakan adalah algoritma K-means yang merupakan suatu penganalisaan data atau metode data mining yang melakukan proses permodelan tanpa supersive (Unspesived) dan merupakan salah satu metode yang melakukan pengelompokkan data dengan sistem partisi.

 Setelah melakukan langkah penentuan teknik serta algoritma yang dibutuhkan, maka setelah itu akan tampil beberapa pilihan sub menu yang terdiri dari use training set, supplied test set, percentage split, classes to clusters evaluation, dan hal itu dapat terlihat melalui gambar berikut :

Gambar 4.5 Cluster mode

 Untuk lebih mengenai sub menu pada cluster menu tersebut serta fungsinya, berikut fungsi serta penjelasannya :

a. Use training set adalah melakukan pengujian dengan menggunakan data yang dibutuhkan sesuai dengan kebutuhan dan berasal dari sendiri.

b. Supplied test set adalah melakukan pengujian yang dilakukan dengan menggunakan data yang berasal dari sumber yang berbeda. Dengan memilih submenu inilah, maka dapat dilakukan prediksi terhadap data yang akan dites.

c. Percentage Split adalah hasil clustering yang akan dites dengan menggunakan k% dari data yang akan dites tersebut. Dimana k merupakan masukan data oleh user.

d. Classe to clusters evalution adalah melakukan pengujian berdasarkan kelas untuk evaluasi kelompok data.

Apabila langkah untuk melakukan mining tersebut telah dilakukan sesuai dengan apa yang telah ditentukan, maka langkah selanjutnya adalah memilih sub menu start dan nantinya akan menampilkan hasil dari teknik data mining menggunakan teknik clustering disertai algoritma K-means. Kemudian akan menghasilkan ouput yang menghasilkan pengclusteran terhadap data class to clusters evolution (Hasil Diagnosa) dengan jumlah cluster (3) seed (1) yang dihasilkan sebagai berikut :

=== Run information ===

Scheme:weka.clusterers.SimpleKMeans -N 3 -A "weka.core.EuclideanDistance -R first-last" -I 500 -S 10

Relation: Jenis Diagnosa Penyakit

Instances: 3

Attributes: 14

 jan

 feb

 mar

 apr

 mei

 jun

 jul

 ags

 sep

 okt

 nov

 des

 jumlah

Ignored:

 Jenis Penyakit

Test mode:Classes to clusters evaluation on training data

=== Model and evaluation on training set ===

kMeans

======

Number of iterations: 2

Within cluster sum of squared errors: 0.0

Missing values globally replaced with mean/mode

Cluster centroids:

 Cluster#

Attribute Full Data 0 1 2

 (3) (1) (1) (1)

===================================================

jan 143.6667 4 300 127

feb 119 3 255 99

mar 157 4 331 136

apr 122.6667 4 252 112

mei 162.3333 4 335 148

jun 141.3333 2 305 117

jul 138.6667 2 295 119

ags 138.6667 2 303 111

sep 137 2 304 105

okt 129 2 273 112

nov 129.6667 2 263 124

des 181 2 386 155

jumlah 1699 32 3602 1463

Time taken to build model (full training data) : 0 seconds

=== Model and evaluation on training set ===

Clustered Instances

0 1 ( 33%)

1 1 ( 33%)

2 1 ( 33%)

Class attribute: Jenis Penyakit

Classes to Clusters:

 0 1 2 <-- assigned to cluster

 1 0 0 | Traumatic

 0 0 1 | Infeksi

 0 1 0 | Degeneratif

Cluster 0 <-- Traumatic

Cluster 1 <-- Degeneratif

Cluster 2 <-- Infeksi

Incorrectly clustered instances : 0.0 0 %

Gambar 4.6 Hasil Proses

Gambar 4.7 Hasil perhitungan per cluster

Dari penjelasan tabel diatas bahwa tabel ini hasil akhir yang telah diproses dalam data mining. Di dalam tabel mempunyai beberapa atribut yaitu atribut Januari, Februari, Maret, April, Mai, Juni, Juli, Agustus, September, Oktober, November, Desember, dan Jumlah. Untuk cluster terdapat tiga cluster yaitu cluster 0, cluster 1, cluster 2.

**4.4 Interpration/Evalution**

Pola informasi yang dapat dihasilkan dari proses data mining yang telah dilakukan sebelumnya, maka akan menghasilkan berupa grafik.

Pada grafik ini merupakan grafik hasil akhir tabel fitted mode yang dihasilkan berupa grafik. Di grafik ini memberikan informasi jumlah hasil diagnose penyakit pada RSUP Dr.Moh Hoesin Palembang.

 Sehingga dapat dihasilkan bahwa dari jenis penyakit yang ada RSUP Dr.moh Hoesin Palembang, jenis penyakit degeneratif yang dimana dalam setiap tahunnya penyakit jenis ini mengalami peningkatan yang signifikan dari tahun ke tahunnya dibandingkan dengan jenis penyakit lainnya.hal itu dapat dilihat melalui proses yang terdapat grafik berikut :

Gambar 4.8 Hasil grafik per cluster

 Dari grafik diatas maka merupakan hasil akhir dari proses fitted mode yang menghasilkan grafik. Dan melalui grafik ini lah dapat diketahui jenis penyakit yang jumlahnya paling tinggi dibandingkan jenis penyakit lainnya dalam pengelompokan hasil diagnosa penyakit pada RSUP Dr.Moh Hoesin Palembang. Setelah grafik tersebut ada proses selanjutnya yang dimana nantinya memberikan informasi secara detail berdasarkan jenis cluster nya masing-masing yang dapat dilihat melalui gambar



Gambar 4.9 Cluster 0



Gambar 4.10 Cluster 1



Gambar 4.11 Cluster 2

Dari keterangan yang dapat diambil Untuk Jenis penyakit yang terdapat jumlah paling besar dalam periode satu tahun adalah jenis penyakit dengan cluster 1 yaitu jenis penyakit Degeneratif dengan jumlah diagnosa 3602 dengan angka tertinggi pada bulan desember sebesar 386 jumlah diagnosa , lalu kemudian cluster 2 yaitu jenis penyakit Infeksi dengan jumlah diagnosa 1463 hasil diagnosa dengan angka tertinggi pada bulan Desember juga dengan jumlah diagnosa sebesar 155 dan yang terakhir cluster 0 yaitu jenis penyakit traumatic dengan jumlah diagnosa 32.

**BAB V**

**KESIMPULAN DAN SARAN**

**5.1 Kesimpulan**

Berdasarkan dari penelitian yang telah dilaksanakan dan sudah diuraikan dalam penerapan data mining untuk pengelompokan diagnose penyakit pasien pengguna askin pada RSUP Dr. Mohammad Hoesin Palembang, maka penulis dapat menarik kesimpulan sebagai berikut :

1. Penerapan *data mining* dengan teknik *clustering* dan algoritma *k-means* yang dilakukan menghasilkan sebuah informasi mengenai Jenis penyakit yang sering diderita pasien pengguna askin berdasarkan hubungan antara diagnosa penyakit dengan jumlah pasien pengguna askin, dimana dari informasi tersebut didapat jumlah diagnosa yang besar yaitu pada jenis penyakit Degeneratif sebesar dengan jumlah diagnosa 3602 dengan angka tertinggi pada bulan desember sebesar 386 jumlah diagnosa.

53

1. Berdasarkan penerapan *data mining* menggunakan *weka* yang dilakukan, maka weka dapat membantu pengelompokkan jenis diagnosa penyakit pada RSUP Dr. Mohammad Hoesin Palembang.
2. Selain disebabkan oleh factor keturunan, factor buruknya lingkungan juga turut menjadi penyebab jenis penyakit degeneratif. Oleh karena itu jelas mengapa diagnosa penyakit ini jumlahnya besar diderita oleh pasien pengguna askin, karena kebanyakan pasien pengguna askin umumnya tergolong miskin dan biasanya tinggal di daerah lingkungan yang kurang bersih dan sanitasi yang buruk sehingga turut membuat kerusakan sel darah merah dan hal tersebut dapat memicu penyakit yang mengakibatkan menurunnya fungsi organ tubuh, maka dapat disimpulkan rata-rata pasien pengguna askin mengidap penyakit ini tanpa diketahui sebelumnya dan kemungkinan pada saat datang kerumah sakit sudah dalam keadaan yang parah sehingga angka kematian untuk jenis penyakit ini meningkat dari tahun ke tahun dikarenakan memang jenis penyakit ini belum ditemukan obatnya dan perlu penanganan yang khusus dan intensif dan hal tersebut sangat sulit dilakukan dikarenakan factor biaya.
3. Jumlah penderita jenis penyakit degeneratif meningkat dalam 1 tahun pada pasien pengguna askin, oleh karena itu pihak RSUP Dr. Mohammad Hoesin Palembang dan Pemprov sumatera selatan memanfaatkan teknologi data mining yang akan membantu mereka melakukan perencanaan anggaran yang efesien dan tepat.

**5.2 Saran**

Berdasarkan kesimpulan yang telah diuraikan diatas, maka ada beberapa saran yang ingin disampaikan yaitu:

1. Pada penelitian selanjutnya dapat mencoba menggunakan *dataset* yang berbeda dan dengan jumlah diagnosa yang lebih besar lagi serta periode waktu yang lebih lama bahkan kalau bisa sampai dengan tahun ini sehingga nilai pengelompokkan diagnosa penyakit menghasilkan tingkat akurasi yang lebih tinggi.
2. Pada penelitian yang akan dating, dapat menggunakan teknik data mining dan algoritma lainnya ataupun juga menggabungkan menjadi dua teknik data mining sehingga akan menghasilkan hasil yang variatif dan tentunya bernilai informasi yang tinggi.
3. Selain informasi ini berguna untuk pihak internal yang terlibat, ada baiknya juga turut membagi informasi dengan pihak eksternal, yang dalam hal ini adalah masyrakat sumatera selatan yang kurang mampu yang umumnya pengguna askin. Sehingga nantinya pihak internal dan eksternal dapat bekerja sama untuk melakukan pencegahan serta pengobatan yang efesien dan tepat sasaran, diharapkan nantinya dapat mewujudkan serta selaras dengan visi dan misi pemerintah sumatera selatan untuk menghasilkan masyrakat sumatera selatan yang gemilang.