

BAB II

LANDASAN TEORI

2.1 Data mining

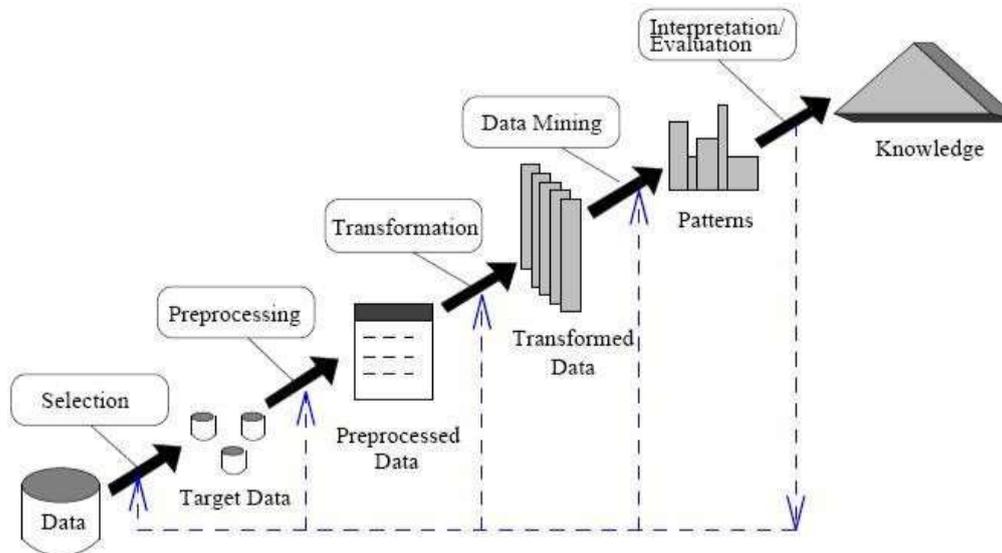
Data Mining adalah teknik analisis data berbasis pada aplikasi statistik yang bertujuan untuk mengekstrak informasi itu sebelumnya tidak dapat ditentukan, dari jumlah Big data [8].

Definisi lainnya yaitu, Data mining adalah suatu proses menganalisis pola data yang tersembunyi menurut berbagai perspektif untuk kategorisasi menjadi informasi yang berguna, yang dikumpulkan di area umum, data warehouse untuk analisis yang efisien, algoritma data mining, memfasilitasi pengambilan keputusan bisnis, dan informasi lainnya [9].

Dari beberapa pendapat diatas mengenai *data mining*, maka dapat disimpulkan bahwa *data mining* adalah sebuah prose untuk menemukan hasil dari data yang sebelumnya tidak diketahui dari berbagai sumber untuk menggali data yang besar dan mengesktrak data tersebut sehingga menjadi suatu informasi yang berguna dan dapat diamanfaatkan.

2.2 Tahapan Proses Data Mining

Data mining merupakan bagian dari proses *knowledge discovery from data* (KDD), dimana salah satu tahapan dalam keseluruhan proses *knowledge discovery from data* (KDD) merupakan data mining yang digambarkan pada gambar 2.1



Gambar 2. 1 Proses KDD Widya Cholil dkk [10]

1. Selection (seleksi data)

Pemilihan atau seleksi data yang dilakukan dari sekumpulan data. Data yang ada pada basis data seringkali tidak dipakai secara menyeluruh, oleh karena itu hanya data yang sesuai untuk proses analisis yang akan diambil dari basis data tersebut.

2. Preprocessing (Pemilihan data)

Pada tahap preprocessing, perlu dilakukan tahap perbersihan atau cleaning pada data. Pembersihan data mencakup pemilihan memeriksa data yang tidak konsisten, menghilangkan data noise dan data duplikasi. setelah data dibersihkan selanjutnya dilakukan pembagian data atau split data menjadi data training dan data testing yang kemudian akan di kelola menggunakan metode mining.

3. Transformation (Data Transformation)

Pada tahap ini data ditransformasi sesuai jenis atau pola informasi yang akan dicari sehingga sesuai untuk proses data mining. Berdasarkan data warehouse yang sudah dihasilkan oleh tahap sebelumnya, maka semua data tersebut diubah atau disesuaikan dengan kebutuhan peneliti untuk mendapatkan hasil analisis yang lebih tepat sesuai dengan tujuan penelitian.

4. Data mining

Data Mining adalah proses pencarian pola atau informasi yang menarik dalam data yang terpilih dengan menggunakan teknik atau metode tertentu. Teknik, metode dan algoritma dalam Data Mining sangat bervariasi. Pemilihan metode dan algoritma yang akan digunakan sangat bergantung pada tujuan proses KDD secara menyeluruh.

5. Interpretation / evaluation

Hasil dari data mining perlu diinterpretasikan sehingga mudah dimengerti oleh pihak lain. Tahap ini merupakan proses menerjemahkan pola informasi atau data yang telah didapat ke dalam bentuk yang lebih mudah dimengerti oleh end-user, dan pihak yang berkepentingan serta pihak luar.

6. Knowledge

Merupakan proses tahap terakhir, Dalam hal ini di gunakan teknik visualisasi yang bertujuan membantu user dalam mengerti dan menginterpretasikan hasil dari pengolahan data (Data Mining).

2.3 Pengelompokan Data Mining

Menurut ismai [9] beberapa tugas yang dapat dilakukan oleh Data Mining dalam proses pemecahan masalah dan pencarian pengetahuan baru di antaranya adalah sebagai berikut :

1. Klastering (Clustering)

Digunakan untuk mengelompokan atau mengidentifikasi data yang memiliki karakteristik tertentu. Contoh algoritma: K-Means, K-Medoids, dan lain-lain.

2. Klasifikasi (Classification)

Digunakan untuk menemukan model atau fungsi yang menjelaskan atau membedakan konsep atau kelas data, dengan tujuan untuk dapat memperkirakan kelas dari suatu objek yang labelnya tidak diketahui.

Contoh algoritma : C4.5, K-Nearest Neighbor (KNN), dan lain-lain.

3. Asosiasi (Association)

Digunakan untuk mengatasi masalah bisnis yang khas, yakni dengan menganalisa tabel transaksi penjualan dang mengidentifikasi produk- produk yang seringkali dibeli bersamaan oleh customer, misalnya apabila orang membeli sambal, biasanya juga dia membeli kecap.

Contoh Algoritma: Apriori, Frequent Pattern Growth (FP- Growth).

4. Estimasi (Estimation)

Digunakan untuk memperkirakan atau menilai sesuatu hal yang belum pernah ada sebelumnya yang disajikan dalam bentuk hasil kuantitatif (angka).

Contoh algoritma: Regresi Linier, Confidence Interval Estimation.

5. Prediksi (Predictions)

Digunakan untuk memperkirakan atau meramalkan suatu kejadian yang belum pernah terjadi.

Contoh algoritma : Decision Tree, K-Nearest Neighbor (KNN).

2.4 KLASIFIKASI

Klasifikasi berguna sebagai pengumpulan data atau metode mengorganisasi data yang telah dikumpulkan, agregasi dan disagregasi serta untuk memberi makna pada suatu set data untuk maksud analisis termasuk untuk penyusunan indeks. Klasifikasi adalah salah satu pembelajaran yang paling umum di data mining.

Klasifikasi adalah proses pengelompokan artinya mengumpulkan benda/entitas yang sama serta memisahkan benda/entitas yang tidak sama. Secara umum dapat dikatakan bahwa batasan klasifikasi adalah usaha menata alam pengetahuan ke dalam tata urutan sistematis.[11].

Klasifikasi adalah proses untuk membangun suatu model atau fungsi yang menggambarkan konsep suatu data. Pada proses klasifikasi sejumlah data dikumpulkan kemudian dibagi menjadi kategori-kategori tertentu yang disebut class [12].

Beberapa algoritma yang dapat digunakan dalam klasifikasi *data mining* adalah sebagai berikut :

1. *Convolutional Neural Network*

convolutional neural network adalah jenis jaringan saraf tiruan yang digunakan dalam pengenalan dan pemrosesan gambar. *convolutional neural network* meniru cara sel-sel saraf kita berkomunikasi dengan neuron yang saling berhubungan dan

convolutional neural network memiliki arsitektur yang sama [13].

2. *Decision tree*

Decision Tree atau Pohon keputusan adalah sebuah struktur data yang terdiri dari simpul (*node*) dan rusuk (*edge*) simpul pada sebuah pohon dibedakan menjadi tiga, yaitu simpul akar (*root node*), simpul percabangan (*branch node*) dan simpul daun atau *leaf node* [14].

3. *Naive Bayes*

Bayesian classification merupakan pengklasifikasian statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu class [15]

4. *K-Nearest Neighbor*

Algoritma KNN adalah salah satu algoritma klasifikasi yang paling terkenal digunakan untuk memprediksi kelas dari catatan atau (sampel) dengan kelas yang tidak ditentukan berdasarkan kelas dari catatan tetangganya, (data terdekat) [16].

Dari beberapa pendapat diatas dapat diartikan klasifikasi berupa metode algoritma yang dapat digunakan untuk mengelompokan data kedalam kelas tertentu seperti *Neural Network*, *Decision Tree*, *Naive Bayes*, dan *K-Nearest Neighbor*.

2.5 Penilaian Kinerja Karyawan

Suatu organisasi maupun perusahaan didirikan karena memiliki tujuan tertentu yang ingin dicapai. Salah satu upaya yang dilakukan perusahaan dalam rangka mencapai tujuan adalah penilaian prestasi kerja. Penilaian kinerja karyawan sering diartikan sebagai pencapaian tugas, dimana karyawan yang bekerja harus sesuai dengan program kerja

perusahaan untuk menunjukkan tingkat kinerja dalam mencapai visi, misi dan tujuan perusahaan tersebut.

Karyawan adalah tenaga kerja yang melakukan pekerjaan dan memberikan hasil kerjanya kepada pengusaha yang mengerjakannya dimana hasil kerjanya itu sesuai dengan profesi atau pekerjaan atas dasar keahlian sebagai mata pencahariannya. [17]

Menurut Hanggraeni. [18] menyatakan bahwa “Penilaian kinerja adalah sebuah proses dimana perusahaan melakukan evaluasi dan penilaian kinerja individu setiap perkerjaannya”.

Menurut Sofyandi. [19] menyatakan bahwa “Penilaian kinerja karyawan adalah penilaian tentang prestasi kerja karyawan. Dalam persaingan global, perusahaan-perusahaan menuntut karyawannya untuk memiliki kinerja yang tinggi. Seiring dengan hal tersebut, karyawan membutuhkan *feedback* (umpan balik) atas kinerja mereka sebagai pedoman perilaku dan keputusan dimasa yang akan mendatang. Penilaian kinerja mencakup aspek kualitatif maupun kuantitatif dari pelaksanaan kerja”.

Dari beberapa definisi menurut para ahli diatas, maka dapat disimpulkan bahwa penilaian kinerja merupakan sebuah sistem yang digunakan untuk mengevaluasi kinerja karyawan agar sesuai dengan standar kerja yang telah ditetapkan oleh perusahaan. Lain dari pada itu penilaian kinerja berguna untuk menentukan kebutuhan pelatihan kerja secara tepat, sehingga dapat melaksanakan pekerjaan yang lebih baik di masa mendatang dan sebagai dasar untuk menentukan kebijakan dalam peningkatan suatu jabatan.

2.6 NAIVE BAYES

Algoritma Naive Bayes merupakan suatu algoritma yang dikemukakan oleh Thomas Bayes seorang ilmuwan Inggris yang dapat memprediksi peluang masa depan berdasarkan pengalaman sebelumnya dan merupakan algoritma yang terdapat pada teknik klasifikasi probabilitas dan statistik.

Metode Naive Bayes hanya membutuhkan jumlah yang kecil dari data yang diperlukan untuk proses pengklasifikasian merupakan suatu keuntungan penggunaan algoritma *Naive Bayes* [9].

Naive Bayes Classifier atau NBC merupakan proses pengklasifikasian probabilitas sederhana yang mengacu pada *Teory Bayes*. Teori tersebut menyatakan bahwa kemungkinan terjadinya suatu peristiwa sama dengan probabilitas intrinsik (dihitung dari data yang tersedia sekarang) dikalikan probabilitas bahwa hal serupa akan terjadi lagi di masa depan (berdasarkan pengetahuan yang terjadinya di masa lalu) [20]

Naive Bayes mempunyai keunggulan yaitu cepat, sederhana, dan memiliki akurasi tinggi [21]. Dibandingkan dengan *classifier* lain *Naive Bayes* bekerja lebih baik dan memiliki tingkat akurasi yang baik dari *Teorema Bayes*. Nilai *Bayes* yang dipilih merupakan persentase tertinggi.

Dari beberapa pendapat diatas *naive bayes* merupakan salah satu metode algoritma yang memiliki keunggulan yaitu dengan pengklasifikasian probalistik sederhana yang cepat, serta memiliki akurasi yang tinggi. Serta digunakan untuk memprediksi kemungkinan (peluang) di masa depan.

Teorema Bayes memiliki bentuk umum sebagai berikut :

$$P(H | X) = \frac{P(X|H)P(H)}{P(X)}$$

Keterangan :

X = Data dengan class yang belum diketahui.

H = Hipotesis Data X merupakan suatu class spesifik.

$P(H|X)$ = probabilitas hipotesis H berdasarkan kondisi x (posteriori prob.)

$P(H)$ = Probabilitas hipotesis H (prior prob.)

$P(X|H)$ = probabilitas X berdasarkan kondisi tersebut

$P(X)$ = probabilitas dari X

2.7 WEKA



Gambar 2. 2 Logo Aplikasi WEKA

WEKA merupakan *software* terintegrasi yang berisi implementasi dari metode-metode *data mining*. *WEKA* dikembangkan oleh Universitas Waikato, Selandia Baru menggunakan bahasa pemrograman Java. Oleh karena itu, *WEKA* merupakan singkatan dari *Waikato Environment for Knowledge Analysis*. Dengan mengadopsi konsep *open source software*, menjadikan *WEKA* dapat digunakan dan dimodifikasi siapapun secara

gratis.

Perangkat sistem operasi *WEKA* dipergunakan dalam pembuatan aplikasi penerapan *data mining*, karena sarana akses data yang lebih cepat dan terdapat beberapa kelengkapan mengenai *tools* teknik *data mining* [22]. *WEKA* mengandung *tools* untuk *pre-processing* data, klasifikasi, regresi, *clustering*, aturan asosiasi dan visualisasi ini juga cocok untuk mengembangkan skema pembelajaran mesin baru [23].

Dengan menggunakan *software WEKA*, *user* dapat melakukan *preprocess* pada data, memasukkannya dalam sebuah skema pembelajaran dan menganalisis *classifier* yang dihasilkan dan performanya, semua itu tanpa menulis kode program sama sekali. Contoh penggunaan *WEKA* adalah dengan menerapkan sebuah metode pembelajaran ke dataset dan menganalisis hasilnya untuk memperoleh informasi tentang data, atau menerapkan beberapa metode dan membandingkan performanya untuk dipilih.

2.8 PENELITIAN SEJENIS

Penulis memulai penelitian ini terlebih dahulu lalu melakukan studi kepustakaan dari penelitian-penelitian sebelumnya dan sumber-sumber lain. Di antaranya penelitian banyak dilakukan dengan teknik data *mining* untuk menggali informasi atau menciptakan informasi baru dari sebuah data. Penelitian tersebut membahas tentang topik yang terkait dengan penelitian penulis, berikut adalah beberapa penelitian mengenai algoritma yang akan digunakan penulis:

Tabel 2. 1 Penelitian Sejenis

No	Penulis dan tahun	Judul	Masalah	Metode	Hasil
1.	Rino Gupitha (2018) [24]	Penerapan Klasifikasi Status Pegawai Menggunakan Metode Naïve Bayes Di Rsu H.Syaiful Anwar	Pada laporan ini Dengan berdirinya rumah sakit H.Syaiful Anwar , pemerintah Kabupaten Subang berharap agar permasalahan pasien yang tidak tertampung di beberapa rumah sakit di Subang bisa teratasi dengan Klasifikasi status pegawai yang tepat akan menentukan posisi karyawan di suatu institusi atau perusahaan. Dengan kesesuaian tersebut maka akan terbentuk suatu kinerja yang profesional.	Metode yang digunakan adalah <i>naïve bayes</i>	Sistem klasifikasi status pegawai ini dapat digunakan untuk mengetahui Status Pegawai baru.
2.	Ajat Sudrajat (2022) [25]	Penerapan Metode Naïve Bayes Untuk Menentukan Penilaian Kinerja Karyawan PT.Sinergi Guna Solusindo	Di dalam dunia pekerjaan adanya karyawan terbaik menjadi tolak ukur kemajuan dari perusahaan itu sendiri. Dalam penentuan biasanya dengan melihat kinerja karyawan tersebut misal dari kerajinan,	Metode yang digunakan adalah <i>naïve bayes</i>	Berdasarkan hasil Analisa pada penelitian yang dilakukan , didapatkan hasil prediksi tingkat akurasi sebesar

No	Penulis dan tahun	Judul	Masalah	Metode	Hasil
			kedisiplinan dan juga prestasi lainnya. Dengan cara seperti ini agak kurang efektif dan akurat.		94.00%. yang dapat membantu HRD dalam mengambil kesimpulan.
3.	Viny Novika Sari (2020) [26]	Analisis dan Penerapan Algoritma Naive Bayes untuk Evaluasi Kinerja Karyawan pada PT. Pelita Wira Sejahtera	Pada PT. PELITA WIRA SEJAHTERA penilaian terkadang dilakukan secara subjektif dan keterbatasan dalam mengontrol setiap karyawan yang bekerja. Oleh karena itu penulis melakukan analisis <i>data mining</i> pada data-data penilaian karyawan tersebut agar dapat mengetahui mana karyawan yang memiliki kinerja yang sangat baik, baik, cukup, dan kurang. Penulis menggunakan data-data karyawan sebanyak 149 data yang	Metode yang digunakan adalah <i>naive bayes</i>	Presentasi akurasi terbesar diperoleh dengan menggunakan <i>Use Training Set Correctly</i> sebesar 95,302%, menggunakan <i>5-Fold Cross Validation Correctly</i> sebesar 93,9597% dan <i>10-Fold Cross Validation Correctly</i> sebesar 93,9597%.

No	Penulis dan tahun	Judul	Masalah	Metode	Hasil
			kemudian disajikan kedalam format arff.		
4.	Sri W. Utami, Ahmad A. Supianto, Fitra A. Bcahtiar (2019) [27]	Analisis Sentimen Evaluasi Kinerja Dosen menggunakan Term Frequency-Inverse Document Frequency dan Naïve Bayes Classifier	Pengajaran yang baik dapat membantu mahasiswa dalam mencapai hasil yang maksimal. Untuk meningkatkan kualitas pembelajaran dan standarisasi akademik perlu dilakukan evaluasi sehingga dapat menghasilkan mahasiswa-mahasiswa yang berkualitas. Oleh karena itu, Jurusan Sistem informasi selalu melakukan evaluasi terhadap kinerja menggunakan kuisisioner yang diisikan oleh mahasiswa disetiap akhir semester. Hasil kolom saran dapat dilakukan analisis sentimen untuk mengetahui saran tersebut bernilai positif,	Menggunakan metode <i>data mining naïve bayes</i> .	Hasil pengujian terhadap 4 parameter menghasilkan akurasi sebesar 80,1%, <i>precision</i> 80,3%, <i>recall</i> 80,3%. Hasil dari <i>Usability testing</i> diperoleh nilai rata-rata kedalam kategori <i>Acceptance</i> dan pada rating "Good".

No	Penulis dan tahun	Judul	Masalah	Metode	Hasil
			negatif atau netral.		
5.	Ikhsan Romli (2020) [28]	Evaluasi Penilaian Kinerja Dalam Klasifikasi Data Mining Dengan Metode Naive Bayes	PT. Berkat Sinar Sentosa merupakan perusahaan yang bergerak di bidang penyediaan jasa security, office boy dan driver. Namun sangat disayangkan karena perusahaan masih manual dalam mengevaluasi penilaian kinerja security, sehingga pengambilan keputusan dalam menentukan security belum bisa dikatakan kompeten dan tidak kompeten. menjadi efektif dan efisien.	Menggunakan metode <i>data mining naïve bayes</i> .	Data yang digunakan dalam penelitian ini menggunakan 63 data latih tanpa menggunakan data uji, namun masih berupa data sampel uji sebagai acuan perhitungan <i>naïve bayes</i> . Tes ini dilakukan dengan menggunakan algoritma data mining <i>Naïve Bayes</i> dan <i>Rapid Miner Framework</i> . Dari perhitungan menggunakan algoritma <i>Naïve Bayes</i>

No	Penulis dan tahun	Judul	Masalah	Metode	Hasil
					diperoleh akurasi sebesar 85,71%, presisi sebesar 88,24% dan recall sebesar 85,71%.

Dari beberapa penelitian pada tabel 2.1, dapat disimpulkan bahwa metode *Naïve Bayes* memiliki akurasi yang baik dalam pengkalsifikasian. Maka dari itu dalam penelitian ini penulis tertarik dalam mengolah data kinerja karyawan rumah sakit dr.Bratanata kota jambi karena sangat relevan dengan metode yang akan digunakan oleh penulis yang dimana data – data nya sudah meiliki kelas atribut (label). Penelitian-penelitian yang telah dilakukan sebelumnya ini akan menjadi acuan penulis dalam melakukan penelitian.

Berdasarkan dari pemaparan penelitian sebelumnya, dapat dilihat persamaan adalah memiliki tujuan yang sama yaitu mengenai evaluasi penilaian kinerja karyawan serta metode yang digunakan untuk mengetahui tingkatan kinerja karyawan selama bekerja di instansi tersebut. Penulis mengolah data hasil evaluasi kinerja karyawan rumah sakit dr.Bratanata Kota Jambi dengan cara mengklasifikasikan data karyawan di 1 tahun terakhir yaitu tahun 2022 dan atribut dalam pengklasifikasian data. Sedangkan perbedaan dari penenelitian sebelumnya yaitu atribut class serta tools yang digunakan. Hasil diperoleh bertujuan untuk mengetahui akumulasi persentasi kinerja karyawan selama 1

tahun terakhir bekerja di rumah sakit dr.Bratanata Kota Jambi.