

BAB II. TINJAUAN PUSTAKA

2.1 Data Mining

Data Mining adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan dan juga informasi di dalam database. Data mining akan membuat pemilik data lebih memahami dan membuat data lebih bermanfaat dari data yang dimiliki dengan cara yang berbeda dari sebelumnya, akan ditemukan hubungan yang tidak terpikirkan sebelumnya dan data akan diringkas untuk menghasilkan analisis dari data tersebut.

Data Mining adalah suatu proses dari pengumpulan data serta pengolahan data yang selanjutnya akan di ekstrak menjadi informasi penting dari dalam data tersebut. Proses ini dapat dilakukan dengan menggunakan perangkat lunak (software) yang tentunya dibantu dengan adanya perhitungan seperti sistematisa, matematika ataupun juga teknologi Artificial Intelligence (AI) yang merupakan bidan dari berbagai bidang keilmuan. (Alim, n.d.)

Berdasarkan pengertian dari data mining dengan penjelasan di atas, maka data mining dapat diartikan pengetahuan yang tersembunyi di dalam database yang diolah lalu diproses untuk menemukan pola dan teknik statistik matematika, machine learning dan kecerdasan buatan untuk mengekstraksi dan mengidentifikasi informasi pengetahuan dari database tersebut.

Pada data mining ada beberapa metode yang dapat digunakan dan juga di terapkan pada kehidupan untuk menganalisis secara meluas pengetahuan secara mudah pada banyak orang salah metodenya yaitu salah satunya adalah metode klasifikasi. (Andriska et al., 2021)

Klasifikasi data mining adalah salah satu dari metode yang sering digunakan karena menjadi metode yang paling umum, data mining ini merupakan sebuah proses untuk menemukan definisi kesamaan karakteristik dalam suatu kelompok atau bisa disebut sebagai kelas (class) yang memiliki tujuan guna memberikan hasil perkiraan kelas dari suatu objek yang sebelumnya masih belum diketahui. (Alim, n.d.)

Adapun beberapa tujuan yang dapat dipenuhi oleh data mining sendiri :

- a. Dapat menjadi sarana menjelaskan (Explanatory)

Yang dimaksud menjelaskan disini yaitu pada kondisi tertentu di dalam berbagai bentuk penelitian agar lebih di mengerti dan akan lebih mudah

menemukan solusi untuk menyelesaikan proses, seperti contohnya dapat dengan mudah memenuhi keinginan user dalam pengembangan perusahaan untuk mengetahui adakah kekurangan dari salesman yang tidak mencapai target di bulan tertentu.

b. Dapat menjadi sarana konfirmasi (Confirmatory)

Konfirmasi disini yaitu untuk memastikan keinginan ataupun pernyataan dan juga mempertegas keinginan user, seperti contohnya plan untuk bulan berjalan kedepannya apa saja yang akan dijadikan poin utama dan akan dipantau ke stabilannya dalam pengumpulan suatu data, sehingga dapat dilihat dalam laporan tahunan apa yang harus diperbaiki kedepannya.

c. Dapat menjadi sarana eksplorasi (Exploratory)

Sarana eksplorasi disini juga dapat digunakan untuk menemukan pola baru yang mungkin sebelumnya tidak dapat terdeteksi ataupun terlewat dari poin yang di rencanakan, seperti contohnya user dapat mengajukan fitur lain yang akan lebih mempermudah dalam pencarian ataupun pengumpulan data seperti fitur filter KPI pada salesman contohnya.

Data mining juga dapat dimanfaatkan dalam dunia bisnis, edukasi, asuransi, perbankan dan lainnya untuk melakukan analisa pasar, mencari tahu kebutuhan pelanggan, mengetahui penawaran yang sesuai dengan kebutuhan, dan dapat memprediksi kemungkinan – kemungkinan yang akan terjadi.

Klasifikasi merupakan metode yang ada pada data mining yang juga sering digunakan untuk menentukan suatu data yang sudah diolah dan disesuaikan dengan kebutuhannya, cara pengerjaannya dengan menilai objek data lalu memasukkannya kedalam kelas tertentu dengan banyaknya yang tentunya tersedia. Dari data latihan yang tersedia akan dilakukan pembangunan model oleh klasifikasi. Pekerjaan dari klasifikasi ini melakukan pelatihan ataupun pembelajaran bagi fungsi target yang memetakan setiap set atribut (fitur) ke dalam satu jumlah label kelas yang sudah tersedia.

Pada sebuah sistem diharapkan akan melakukan klasifikasi dengan benar untuk semua data set yang ada, tetapi kinerja pada sistem tidak bisa mencapai angka 100% benar, dalam hal ini kinerja sebuah sistem klasifikasi juga harus diukur. Pada umumnya cara mengukur dari kinerja klasifikasi ini dengan melakukan matriks konfusi.

Tahap Klasifikasi pada penelitian ini dilakukan menggunakan beberapa algoritma diantaranya:

- a. Decision Tree, merupakan algoritma pengambilan keputusan yang melakukan partisi rekursif atas ruang instance, dari pohon keputusan tipikal terdiri dari simpul internal, tepi dan simpul daun. Tes pada atribut ataupun subset atribut adalah perwakilan dari setiap simpul yang biasa disebut dengan simpul keputusan. Nilai spesifik ataupun rentang nilai atribut input akan diberikan label dari masing – masing edge. Akurasi yang serupa diperoleh dari klasifikasi decision tree bahkan terkadang lebih baik jika dibandingkan dengan metode klasifikasi lainnya.
- b. Naïve bayes, merupakan sebuah bentuk pengklasifikasian probabilistik sederhana yang dilakukan dengan cara menghitung sekumpulan dari probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari dataset yang diberikan, dikemukakan juga oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya. (Jaya Purnama & Rahayu, 2022)

2.2 Decision Tree

Klasifikasi adalah salah satu metode dari data mining dan banyak juga yang menggunakan metode tersebut untuk mengolah data pada bidang pembelajaran dan juga di bidang lainnya. Juga banyak penelitian menggunakan metode klasifikasi ini yang kemudian dijadikan informasi berguna dan juga mudah di pahami, dengan mengolah dan menganalisis data yang sudah tersedia lalu diubah menjadi informasi untuk memudahkan dalam menentukan keputusan yang akan diambil. Ada juga contoh klasifikasi yaitu Decision Tree, metode ini dapat digunakan untuk Menyusun kelas – kelas yang ada pada data lalu diubah menjadi wujud pohon berakar. (Andriska et al., 2021)

Ada pula Klasifikasi decision tree menjadi salah satu teknik yang cukup terkenal didalam dunia data mining, populer dalam hal membuat keputusan dalam suatu kasus, tidak diperlukan proses pengelolaan pengetahuan yang sudah terdahulu untuk menyelesaikan kasus – kasus yang memiliki sebuah dimensi yang besar.

Klasifikasi decision tree dapat di pergunakan untuk mengklasifikasikan beberapa aspek yang ingin di kembangkan atau juga menjadi acuan dari yang sudah pernah dialami agar kedepannya dapat memberikan keamanan dan kenyamanan kepada konsumen sehingga kedua belah pihak tidak merasa dirugikan.

Kelebihan utama decision tree sangat mudah dipahami, diinterpretasikan dan juga cocok untuk membuat visualisasi data, juga dapat menghasilkan informasi yang berguna tanpa menggunakan data yg matang dikarenakan hanya membutuhkan persiapan yang

sangat minimal, decision tree ini juga dapat memilih pilihan yang paling terbaik dari banyaknya pilihan yang ada.

Algoritma adalah bentuk dari urutan langkah yang logis, biasa digunakan untuk memecahkan masalah yang secara singkat masalah tersebut harus terdiri dari beberapa langkah yang logis. Pembuatan algoritma itu sendiri berfungsi untuk digunakan dalam bentuk pemecahan masalah pemrograman yang kompleks dan juga menyelesaikan program sederhana dan juga kompleks. Algoritma ini juga memiliki keuntungan lainnya seperti algoritma ini dapat digunakan secara berulang. Algoritma ini dapat dengan mudah dan sederhana dalam penulisan pemrogramannya, dengan menggunakan strategi *top-down* dan *divide-and-conquer* dalam mengimplementasikannya.

Berikut cara menghitung nilai gain menggunakan rumus :

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i)$$

Keterangan :

- S = himpunan kasus
- A = atribut
- n = jumlah partisi atribut A
- |S_i| = jumlah kasus pada partisi ke i
- |S| = jumlah kasus dalam S

Menentukan akar dari pohon. Akar akan diambil dari atribut yang terpilih, dengan cara menghitung nilai gain dari masing –masing atribut, nilai gain yang paling tinggi yang akan menjadi akar pertama. Sebelum menghitung nilai gain dari atribut, hitung dahulu nilai entropy. Untuk menghitung nilai entropy digunakan rumus. (Ainnur et al., 2021)

Berikut untuk rumus entropy :

$$Entropy(S) = \sum_{i=1}^n p_i * \log_2 p_i$$

Keterangan :

- S = himpunan kasus
- A = fitur
- n = jumlah partisi S

p_i = proporsi dari S_i terhadap S

2.3 Metode Naïve Bayes

Metode naïve bayes merupakan algoritma dari machine learning dalam mengatasi permasalahan klasifikasi dengan menerapkan teorema probabilitas bayes. Metode ini dikenal dengan kesederhanaannya dan juga keefektifannya juga cepat dalam membangun model dan membuat prediksi dengan algoritma naïve bayes. Algoritma pada Naïve bayes merupakan algoritma yang perhitungannya menggunakan perhitungan probabilitas, pada algoritma ini sangat umum digunakan pada saat menyelesaikan bentuk permasalahan prediksi yang berupa klasifikasi.

Algoritma dari naïve bayes sangat sering dikenali sebagai salah satu algoritma yang memiliki tingkat akurasi cukup tinggi, pada prosesnya klasifikasi berikut ini terdapat dari dua fase yaitu fase training dan juga fase testing. Fase training ini bisa juga disebut sebagai fase learning dapat diartikan Sebagian dari datanya telah diketahui oleh kelas datanya teruntuk model perkiraan. Fase kedua yaitu fase testing atau juga biasa disebut sebagai fase classify model yang sudah terbentuk diuji menggunakan sebagian data yang lainnya agar dapat diketahui akurasi atas model yang sudah terbentuk.

Algoritma naïve bayes juga menjadi salah satu algoritma yang cukup banyak dikenal atau biasa disebut populer untuk digunakan pada banyaknya studi kasus pada data mining ataupun text mining. Pada algoritma ini masuk kedalam salah satu metode klasifikasi yang pengklasifikasiannya memanfaatkan probabilitas dan juga statistik berguna untuk memprediksi peluang di masa yang akan datang dengan memanfaatkan pengalaman yang ada di masa lalu atau masa sebelumnya. (Yudhi Putra & Ismiyana Putri, n.d.)

Metode ini memiliki asumsi yang kuat bahwa semua predictor independent satu dengan yang lainnya. Dengan kata lain kehadiran fitur nya bergantung pada fitur yang ada pada kelas yang lainnya. Naïve bayes ini bertujuan untuk melakukan klasifikasi probabilitas berdasarkan dari pembelajaran mesin atas probabilitas lainnya. Dapat juga dimanfaatkan untuk melakukan evaluasi banyak hal seperti memperkirakan di bulan selanjutnya Performance KPI salesman apa yang harus menetap di setiap bulannya.

Dapat pula menampilkan hasil – hasil presentase pencapaian KPI salesman pada bulan yang sudah lalu sehingga dapat memberikan perusahaan untuk mengambil solusi yang lebih awal guna menghindari hal – hal yang tidak diinginkan terjadi.

Naïve bayes tentunya memiliki beberapa kelebihan seperti, cepat dan mudah nya

dalam mengimplementasikan, juga memiliki sifat yang sangat skalabel dengan jumlah predictor titik data. Tetapi Adapun kelemahan dari naïve bayes ini seperti dalam hal keakuratannya tidak dapat diukur dengan hanya satu probabilitas saja, melainkan membutuhkan bukti – bukti lainnya untuk memastikan keakuratannya.

Naive Bayes teruji mempunyai akurasi serta kecepatan yang besar dikala diaplikasikan ke dalam database dengan informasi yang besar. Keuntungan pemakaian Naive Bayes ialah cuma membutuhkan beberapa kecil informasi latihan buat memastikan parameter mean serta varians dari variabel yang dibutuhkan buat klasifikasi. (Alim, n.d.)

Metode ini banyak diterapkan pada hal – hal yang berkenaan dengan diagnosa secara statistik yang berhubungan dengan probabilitas serta kemungkinan dari penyakit dan gejala gejala yang berkaitan. (Putra et al., 2022)

Torema Bayes memiliki rumus sebagai berikut :

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

Keterangan :

X = Data dengan *class* yang belum diketahui

H = Hipotesis data merupakan suatu *class* spesifik

$P(H|X)$ = Probabilitas hipotesis H berdasar kondisi X (*Posteriori Probabilitas*)

$P(H)$ = Probabilitas hipotesis H (*Prior Probabilitas*)

$P(X|H)$ = Probabilitas X berdasarkan kondisi hipotesis H

$P(X)$ = Probabilitas X

2.4 Key Performance Indicator

Key Performance Indicator atau juga biasa disingkat menjadi KPI ini adalah suatu matriks atau juga nilai terukur yang berfungsi sebagai acuan dari seberapa efektif suatu perusahaan dalam menunjukkan pencapaian dari salah satu tujuan bisnis yang sedang dijalankan.

KPI berupa serangkaian indikator kunci yang memiliki sifat terukur dan memberikan informasi beban sasaran yang strategis dari pencapaian suatu organisasi, dimana organisasi tersebut secara interaktif terintegrasi dalam sebuah strategi organisasi

dengan secara menyeluruh. (Ghufroni An & Kurniawan, 2022).

Key Performance Indicator sebagai yang paling kritikal untuk kesuksesan organisasi pada kondisi sekarang dan di masa datang. Key Performance Indicator merupakan ukuran berskala dan kuantitatif yang digunakan untuk mengevaluasi kinerja organisasi dalam tujuan mencapai target organisasi. KPI juga digunakan untuk menentukan objektif yang terukur, melihat tren, dan mendukung pengambilan keputusan. (Dwi et al., n.d.). KPI sebagai alat ukur yang dapat menggambarkan tingkat efektivitas pada perusahaan dalam mencapai tujuan bisnisnya.

KPI menjadi salah satu acuan untuk salesman mendapatkan insentif, tidak hanya satu KPI tetapi terdapat beberapa KPI yang harus dipenuhi syarat nya oleh salesman, jika salesman tidak mencapai batas presentase salah satu KPI maka tidak akan mendapatkan insentif. Adapun beberapa karakteristik dari KPI, yaitu sebagai berikut :

- Ukuran non finansial
- Ukuran yang sering sekali digunakan
- Ukuran yang tentunya diketahui manajemen
- Semua orgn dalam sebuah organisasi paham dan telah mengerti mengenai KPI
- Merupakan tanggung jawab kepada individu pada tim
- KPI memiliki efek yang signifikan
- Juga memiliki efek positif

KPI juga digunakan untuk menentukan objektif yang terukur, untuk mendukung dalam pengambilan keputusan juga untuk melihat tren. Key Performance Indicator adalah ukuran yang bersifat kuantitatif dan bertahap bagi perusahaan serta memiliki berbagai perspektif dan berbasis data konkret, dan menjadi titik awal penentuan tujuan dan penyusunan strategi organisasi. (Nurwahidah et al., 2021)

KPI itu sendiri dapat diukur dalam beberapa periode mulai dari harian, mingguan, dan juga bulanan, KPI yang baik adalah suatu hal yang penting dan akan terus menerus mendapatkan perhatian dari manajemen.