

BAB II

LANDASAN TEORI

2.1. Penelitian Terdahulu

Penelitian yang dilakukan Dilla Yulia dengan judul ‘Pembuatan Sistem Berbasis Android Untuk Memprediksi Penyakit Berdasarkan Gejala yang Ditimbulkan Menggunakan Metode Naïve Bayes’ bertujuan merancang sebuah sistem prediksi sesuai gejala yang dialami pasien untuk mencegah penyakit tersebut semakin memburuk, informasi tentang penyakit yang dikaitkan dengan gejala yang dialami diberitahukan kepada pasien untuk mengurangi kemungkinan makin buruknya penyakit yang dialami oleh pasien, yang diprediksi menggunakan metode naïve bayes. Yang dirancang dengan android yang dimaksudkan untuk memudahkan pengaplikasiannya di masyarakat. [2]

Penelitian yang dilakukan Agung Kurniadi dengan judul ‘Aplikasi Data Mining Menggunakan Naïve Bayes Classifier Untuk Persetujuan Pengajuan Kredit’ bertujuan untuk membantu pihak yang akan menggunakan sistem untuk menentukan kriteria nasabah yang akan mengajukan kredit, sehingga dapat meramalkan kriteria nasabah dengan melakukan pengenalan pola data historis serta menentukan apakah pengajuan nasabah tersebut diterima/layak atau ditolak/tidak layak. [3]

Persamaan peneliti ini dengan kedua penelitian yang dilakukan oleh Dilla Yulia dan Agung Kurniadi adalah menggunakan metode yang sama untuk

meramalkan kejadian yang akan dimasa yang datang, dengan menggunakan data historis yang telah ada pada tempat penelitian masing masing.

Perbedaannya permasalahan yang terjadi dalam kasus tersebut dalam bidang yang berbeda sehingga permasalahan yang dibahas hanya mengenai perbedaan perhitungan algoritma data mining, dan dalam dua kasus sebelumnya dengan kasus yang sedang diteliti tidak memiliki perbedaan perhitungan metode yang signifikan.

2.2. Sistem

Sistem berasal dari bahasa Latin (*systēma*) dan bahasa Yunani (*sustēma*) adalah suatu kesatuan yang terdiri komponen atau elemen yang dihubungkan bersama untuk memudahkan aliran informasi, materi atau energi untuk mencapai suatu tujuan. Istilah ini sering dipergunakan untuk menggambarkan suatu set entitas yang berinteraksi, di mana suatu model matematika seringkali bisa dibuat.

Sistem juga merupakan kesatuan bagian-bagian yang saling berhubungan yang berada dalam suatu wilayah serta memiliki item-item penggerak, contoh umum misalnya seperti negara. Negara merupakan suatu kumpulan dari beberapa elemen kesatuan lain seperti provinsi yang saling berhubungan sehingga membentuk suatu negara di mana yang berperan sebagai penggeraknya yaitu rakyat yang berada dinegara tersebut.

Kata "sistem" banyak sekali digunakan dalam percakapan sehari-hari, dalam forum diskusi maupun dokumen ilmiah. Kata ini digunakan untuk banyak hal, dan pada banyak bidang pula, sehingga maknanya menjadi beragam. Dalam pengertian

yang paling umum, sebuah sistem adalah sekumpulan benda yang memiliki hubungan di antara mereka

2.2.1. Pengertian Sistem

Menurut **Erwan Arbie** dalam bukunya yang berjudul “Pengantar Sistem Informasi Manajemen” , pengertian sistem adalah sebagai berikut :

“Sistem informasi adalah sistem di dalam suatu organisasi yang mempertemukan kebutuhan pengolahan transaksi harian, membantu dan mendukung kegiatan operasi, bersifat manajerial dari suatu organisasi dan membantu mempermudah penyediaan laporan yang diperlukan.”.[4]

Sedangkan menurut **Tafri D. Muhyuzir**, dalam bukunya yang berjudul “Analisa Sistem Informasi”. pengertian sistem adalah sebagai berikut :

“Sistem informasi adalah data yang dikumpulkan, dikelompokkan dan diolah sedemikian rupa sehingga menjadi sebuah satu kesatuan informasi yang saling terkait dan saling mendukung sehingga menjadi suatu informasi yang berharga bagi yang menerimanya”.[5]

Dari dua pengertian diatas dapat disimpulkan bahwa Sistem adalah hal yang ada dalam sebuah lingkungan kerja, yang mempermudah pengguna atau penerima sistem tersebut dalam mencapai suatu tujuan yang telah disepakati sebelumnya agar bermanfaat bagi penggunanya.

2.3. Informasi

Informasi adalah pesan (ucapan atau ekspresi) atau kumpulan pesan yang terdiri dari order sekuens dari simbol, atau makna yang dapat ditafsirkan dari pesan atau kumpulan pesan. Informasi dapat direkam atau ditransmisikan. Hal ini dapat dicatat sebagai tanda-tanda, atau sebagai sinyal berdasarkan gelombang. Informasi adalah jenis acara yang mempengaruhi suatu negara dari sistem dinamis. Para konsep memiliki banyak arti lain dalam konteks yang berbeda. Informasi bisa dikatakan sebagai pengetahuan yang didapatkan dari pembelajaran, pengalaman, atau instruksi. Namun, istilah ini memiliki banyak arti bergantung pada konteksnya, dan secara umum berhubungan erat dengan konsep seperti arti, pengetahuan, nentropy, Persepsi, Stimulus, komunikasi, kebenaran, representasi, dan rangsangan mental.

Dalam beberapa hal pengetahuan tentang peristiwa-peristiwa tertentu atau situasi yang telah dikumpulkan atau diterima melalui proses komunikasi, pengumpulan intelejen, ataupun didapatkan dari berita juga dinamakan informasi. Informasi yang berupa koleksi data dan fakta seringkali dinamakan informasi statistik. Dalam bidang ilmu komputer, informasi adalah data yang disimpan, diproses, atau ditransmisikan. Penelitian ini memfokuskan pada definisi informasi sebagai pengetahuan yang didapatkan dari pembelajaran, pengalaman, atau instruksi dan alirannya.

2.3.1. Pengertian Informasi

Menurut **Abdul Kadir** dalam bukunya yang berjudul “Pengenalan Sistem Informasi” , pengertian informasi adalah sebagai berikut :

“mendefinisikan informasi sebagai data yang telah diproses sedemikian rupa sehingga meningkatkan pengetahuan seseorang yang menggunakan data tersebut.”.[6]

Sedangkan menurut **Anton M. Meliono**, dalam bukunya berjudul “ Kamus Besar Bahasa Indonesia”. pengertian sistem adalah sebagai berikut :

“Informasi adalah data yang telah diproses untuk suatu tujuan tertentu. Tujuan tersebut adalah untuk menghasilkan sebuah keputusan.”.[7]

Dari dua pengertian diatas dapat disimpulkan bahwa informasi adalah sebuah data yang telah diproses sebelumnya untuk melakukan suatu tujuan tertentu, tujuan yang akan menghasilkan sebuah keputusan tertentu.

2.4. Promosi

Pada dasarnya promosi ditujukan untuk memperkenalkan produk baru, juga untuk mempertahankan suatu produk yang sedang dipasarkan yang pada akhirnya sasaran promosi itu untuk memperbesar penjualan. Jadi promosi berarti mempertahankan kelangsungan hidup perusahaan.[8]

Untuk lebih memahami konsep promosi, dapat dikemukakan pendapat dari Nitisemito sebagai berikut: “Promosi adalah salah satu kegiatan dibidang marketing yang bertujuan untuk meningkatkan omzet penjualan, dengan jalan mempengaruhi konsumen baik langsung maupun tidak langsung” [9].

2.5. Data Mining

Secara sederhana, *data mining* atau penambangan data dapat didefinisikan sebagai proses seleksi, eksplorasi, dan pemodelan dari sejumlah besar data untuk menemukan pola atau kecenderungan yang biasanya tidak disadari keberadaannya. *Data mining* dapat dikatakan sebagai proses mengekstrak pengetahuan dari sejumlah besar data yang tersedia. Pengetahuan yang dihasilkan dari proses *data mining* harus baru, mudah dimengerti, dan bermanfaat. Dalam *data mining*, data disimpan secara elektronik dan diproses secara otomatis oleh komputer menggunakan teknik dan perhitungan tertentu.[10]

2.6. Naïve Bayes

Naive Bayes merupakan sebuah pengklasifikasian probabilistik sederhana yang menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari dataset yang diberikan. Algoritma menggunakan teorema Bayes dan mengasumsikan semua atribut independen atau tidak saling ketergantungan yang diberikan oleh nilai pada variabel kelas[11]. Definisi lain mengatakan Naive Bayes merupakan pengklasifikasian dengan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya [12].

Naive Bayes didasarkan pada asumsi penyederhanaan bahwa nilai atribut secara kondisional saling bebas jika diberikan nilai output. Dengan kata lain, diberikan nilai output, probabilitas mengamati secara bersama adalah produk dari probabilitas individu[13]. Keuntungan penggunaan Naive Bayes adalah bahwa metode ini hanya

membutuhkan jumlah data pelatihan (Training Data) yang kecil untuk menentukan estimasi parameter yang diperlukan dalam proses pengklasifikasian. Naive Bayes sering bekerja jauh lebih baik dalam kebanyakan situasi dunia nyata yang kompleks dari pada yang diharapkan[14].

2.6.1. Persamaan Teorema Naïve Bayes [12]

$$P(C|X) = \frac{P(X|C) P(c)}{P(x)} \quad \text{Persamaan Teorema Bayes (1)}$$

x : Data dengan class yang belum diketahui

c : Hipotesis data merupakan suatu class spesifik

$P(c|x)$: Probabilitas hipotesis berdasarkan kondisi

$P(c)$: Probabilitas hipotesa

$P(x|c)$: Probabilitas berdasarkan kondisi pada hipotesis

$P(x)$: Probabilitas c

Rumus diatas menjelaskan bahwa peluang masuknya sampel karakteristik tertentu dalam kelas C (Prosterior) adalah sebuah peluang munculnya kelas C (sebelum masuknya sampel tersebut, seringkali disebut prior), dikali dengan peluang kemunculan karakteristik sampel pada kelas C (disebut juga likelihood), dibagi dengan peluang kemunculan karakteristik sampel secara global (disebut juga evidence), Karena itu, rumus diatas dapat pula ditulis sebagai berikut:

$$prosterior = \frac{prior \times likelihood}{evidence} \quad \text{Persamaan Teorema Bayes (2)}$$

Nilai evidence selalu tetap untuk setiap kelas pada satu sampel. Nilai dari posterior tersebut nantinya akan dibandingkan dengan nilai nilai prosterior kelas lainnya untuk menentukan ke kelas apa suatu sampel akan diklasifikasikan. Penjabaran lebih lanjut rumus Bayes tersebut dengan menggunakan $c|x_1, \dots, x_n$ menggunakan aturan perkalian sebagai berikut :

$$\begin{aligned}
 P(C|X_1, \dots, X_n) &= P(C)P(X_1, \dots, X_n|C) \\
 &= P(C)P(X_1|c)(X_2, \dots, X_n|C, X_1) \\
 &= P(C)P(X_1|c)P(X_2|C, X_1)(X_3, \dots, X_n|C, X_1, X_2) \\
 &= P(C)P(X_1|c) P(X_2|C, X_1)P(X_2|C, X_1, X_2) \dots P(X_n|C, X_1, X_2, \dots, X_{n-1})
 \end{aligned}$$

Persamaan Naïve Bayes Aturan Perkalian (3)

Dapat dilihat bahwa hasil penjabaran tersebut menyebabkan semakin banyak dan semakin kompleksnya factor factor syatat yang mempengaruhi nilai probabilitas, yang hamper mustahil untuk dianalisa satu persatu. Akibatnya, perhitungan tersebut menjadi sulit untuk dilakukan. Disinilah digunakan asumsi indepedensi yang sangat tinggi, bahwa masing masing petunjuk saling bebas satu sama lain. Dengan asumsi berikut, maka berlaku suatu persamaan sebagai berikut:

$$P(c|X_1, \dots, X_n) = P(C) \prod_{i=1}^n P(X_i|C)$$

$$P(c|X) = P(x_1|c)P(X_2|c) \dots P(X_n|c)P(c) \quad \text{Persamaan Naïve Bayes (4)}$$

