

BAB II

TINJAUAN PUSTAKA

A. Penelitian Terdahulu

Penelitian terdahulu menjadi bahan acuan dalam penelitian sehingga dapat membantu dalam mengkaji penelitian yang dilakukan. Berikut penelitian terdahulu yang masih berkaitan dengan penelitian yang dilakukan penulis:

1. Dalam Penelitian Sulastri dan Yusuf Sulisty Nugroho pada jurnal yang berjudul “PENERAPAN DATA MINING UNTUK PREDIKSI RATING PENJUALAN BUKU MENGGUNAKAN METODE *NAÏVE BAYES*” dijelaskan bahwa melihat banyaknya buku yang beredar di pasaran dengan segala kelebihan dan kelemahannya maka diperlukan sebuah sistem prediksi untuk mengetahui faktor yang mempengaruhi rating penjualan buku untuk menghasilkan sebuah buku yang memenuhi kriteria pembaca. Teknik yang digunakan dalam prediksi rating penjualan buku ini memakai algoritma *Naïve Bayes*. *Naïve Bayes* digunakan untuk mencari nilai probabilitas paling besar pada setiap variabel yang sudah ada.
2. Dalam penelitian Kevin Moniarga Suharta dalam judul “PENERAPAN ALGORITMA *NAIVE BAYES CLASSIFIER* UNTUK MENGETAHUI MINAT BELI PELANGGAN TERHADAP SOFA (STUDI KASUS DI MEBEL KELUMER BAYAU)” menyebutkan permasalahan utama yang dihadapi oleh Mebel Kelumer Bayau adalah bagaimana memprediksi minat beli pelanggan dari penjualan barang Mebel Kelumer Bayau pada masa mendatang dari data yang telah diperoleh sebelumnya. Penelitian ini menggunakan metode Algoritma *Naive Bayes Classifier* yang mempunyai akurasi dan kecepatan untuk dilakukan penggalian pengetahuan yang sudah ada pada database.

3. Dalam Penelitian Dicky Norfiansyah yang berjudul “PENERAPAN DATA MINING DENGAN ALGORITMA *NAIVE BAYES CLASIFIER* UNTUK MENGETAHUI MINAT BELI PELANGGAN TERHADAP KARTU INTERNET XL (STUDI KASUS DI CV. SUMBER UTAMA TELEKOMUNIKASI)” menyebutkan metode data mining bisa meringankan penyelesaian problem yang dilakukan di CV. Sumber Utama Telekomunikasi. Oleh sebab itu algoritma naive bayes dipakai untuk mencari peluang di masa akan datang dari kejadian di masa lalu. Hasil dari penelitian ini bisa memprediksi laku atau tidak terjual kartu perdana, sehingga perusahaan bisa mendapatkan keputusan sehingga bisa menaikkan strategi penjualan.
4. Dalam penelitian Rizha Franstika Wicaksana yang berjudul “IMPLEMENTASI DATA MINING PADA PERSEWAAN ALAT PESTA MENGGUNAKAN METODE APRIORI” menyebutkan berbagai macam dan model pada penyewaan alat-alat pesta dapat mengetahui adanya persoalan saat memasang berbagai macam alat pesta dengan harga yang diinginkan dan ketersediaan barang. Dari permasalahan tersebut muncul penyelesaian untuk membuat sistem informasi persewaan alat pesta dengan memakai data mining dengan metode apriori di UD. Reza Jaya yang yang kedepannya bisa membuat model alat-alat pesta dan memperoleh saran kepada pemilik persewaan untuk mencocokkan barang sesuai dengan harganya..

Dari tiga penelitian terdahulu yang sudah dibahas dapat disimpulkan jika penelitian ini terdapat perbedaan dengan penelitian terdahulu. Pada penelitian terdahulu membahas tentang minat beli pelanggan terhadap suatu produk, sedangkan pada penelitian ni membahas tentang prediksi minat pelanggan pada penyewaan.

B. Landasan Teori

a. Prediksi

Prediksi adalah memperkirakan sesuatu yang peluang yang mungkin terjadi dimasa mendatang menggunakan informasi di masa sebelumnya dan saat ini yang sudah terjadi, sehingga kesalahan strategi dapat lebih sedikit dilakukan. Untuk prediksi hanya mendapatkan jawaban yang bisa terjadi di masa selanjutya yang terkadang tidak sesuai dengan realita pada masa depan (Budi Santosa, 2007; 5).

Menurut kamus besar Bahasa Indonesia, prediksi adalah hasil dari kegiatan memprediksi nilai pada masa depan dengan memakai data masa lalu.

b. *Naïve Bayes*

Teori keputusan Bayes adalah pendekatan statistik yang fundamental dalam pengenalan pola (*pattern recognition*). Pendekatan yang dimaksud untuk keputusan suatu klasifikasi memakai probabilitas yang ditimbulkana dari suatu keputusan.

Naïve Bayes ialah salah satu metode data mining yang cukup banyak dipakai dikarenakan lebih sederhana proses perhitungan datanya. Algoritma *naïve bayes* mempunyai nilai akurasi yang lumayan tinggi apabila digunakan ke dalam database yang lebih banyak. *Naïve bayes* merupakan pendekatan statistik pada pengenalan pola (*pattern recognition*) yang ada dalam tingkat sederhana dari nilai atribut. (Sulastri, 2017;59-60)

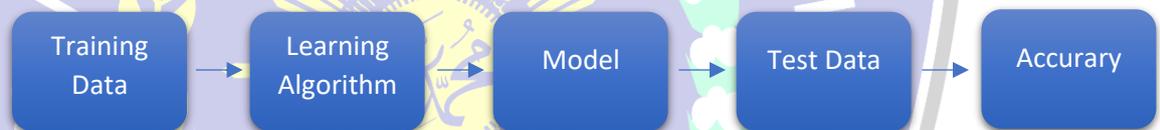
Salah satu algoritma yang sering digunakan untuk data mining adalah *Naïve Bayes*, dikarenakan *Naïve Bayes* memiliki performa yang kompetitif untuk proses klasifikasi walaupun tidak ada keterkaitan antara atribut.

Metode *Naïve Bayes* merupakan pengelompokan data statistik yang didapatkan dari teorema Bayes yang bisa biasanya digunakan saat prediksi probabilitas dari kelompok suatu *class*. (Jananto, 2013)

Naïve Bayes pertama kali dikemukakan oleh ilmuwan Inggris Thomas Bayes, yang disebut dengan teorema bayes yang artinya memprediksi di masa yang mendatang dari usaha yang dilakukan pada masa sekarang. Kemudian teorema bayes tersebut digabungkan dengan “*Naïve*” yang diartikan sebagai kondisi antar atribut saling bebas.

Dalam kasus studi sebelum algoritma klasifikasi sudah ditemukan lebih dahulu sistem Bayesian yang sering disebut *Naïve Bayes Classifier*. Apabila memakai *Naïve Bayes Classifier* dapat menemukan tingkat ketepatan dengan kecepatan lumayan tinggi dengan model perhitungan yang sederhana jika diterapkan untuk analisis pada database yang besar.

Naïve Bayes Classifier untuk penyelesaiannya hampir sama dengan *Nearest Neighbor*. Ada beberapa tahapan metode klasifikasi yang diawali dengan proses *training* dan diakhiri dengan proses *testing* yang kemudian menghasilkan keputusan yang lebih akurat. Berikut tahapan penyelesaian klasifikasi:



Gambar 2.1 Tahapan Klasifikasi

Teorema Bayes termasuk dasar dari metode *Naïve Bayes Classifier*. Berikut perhitungan algoritma *Naïve Bayes*:

$$P(H|X) = \frac{p(X|H), p(H)}{p(X)} \dots\dots\dots(1)$$

Keterangan:

X : Data yang sudah dimasukkan ke dalam kelas.

H : Data atau nilai X dimasukan kelas C

P(H/X): Probabilitas H pada kondisi x (*posteriori probability*)

P(H) : Probabilitas hipotesis H (*prior probability*)

P(X/H): Probabilitas X pada kondisi H

P(X) : Probabilitas dari X.

Naive Bayes tahap klasifikasi memakai beberapa proses yang digunakan saat mengelompokkan kelas yang paling baik untuk sebuah sampel yang akan digunakan untuk proses analisis. Oleh sebab itu, metode *Naive Bayes* diatas disesuaikan. Oleh sebab itu, metode *Naive Bayes* diatas disesuaikan menjadi:

$$P(C|F1 \dots Fn) = \frac{P(C)P(F1 \dots Fn|C)}{P(F1 \dots Fn)} \dots\dots\dots(2)$$

Keterangan:

pada variabel C menunjukkan kelas, variabel F1 ... Fn menampilkan nilai suatu data yang dipakai untuk proses klasifikasi. Jadi, rumus diatas mencari peluang suatu contoh ciri tertentu pada kelas C (*Posterior*) ialah suatu peluang kemunculan kelas C (ampel data yang belum dioalah, yang disebut *prior*), kemudian dikali bersama peluang menculnya ciri khusus contoh pada kelas C (*likelihood*), kemudian ciri khusus contoh pada kelas C (*evidence*) dibagi dengan *likelihood*. Oleh sebab itu,perhitungan rumus bisa ditulis menggunakan rumus yang lebih sederhana seperti berikut:

$$Posterior = \frac{prior \times likelihood}{evidence} \dots\dots\dots(3)$$

Keterangan:

Evidence adalah nilai yang tidak berubah pada kelas yang diapakai pada sampel. Nilai dari *posterior* tadi akan dibandingkan menggunakan nilai-nilai *posterior* kelas yang lain yang digunakan untuk memasukkan kelas sampel yang digunakan untuk klasifikasi.

Dari rumus *Bayes* ($C|F1 \dots n$) dapat dijabarkan dengan perkalian berikut:

$$\begin{aligned} (C|F1, \dots, Fn) &= p(C)P(F1, \dots, Fn|C) = (C)p(F1|C)P(F2, \dots, Fn|C, F1) \\ &= P(C)P(F1|C)P(F2|C, F1)P(F3, \dots, Fn|C, F1, F2) \\ &= P(C)P(F1|C)P(F2|C, F1)P(F3|C, F1, F2)P(F4, \dots, Fn|C, F1, F2, F3) \\ &= P(C)P(F1|C)P(F2|C, F1)P(F3|C, F1, F2) \dots \\ &P(Fn|C, F1, F2, F3, \dots, Fn - 1) \dots\dots\dots(4) \end{aligned}$$

Dari hasil perhitungan diatas dapat dilihat rumus yang digunakan lebih kompleks faktor-faktor syarat yang dipengaruhi dari nilai probabilitas, yang tidak mungkin untuk diteliti satu persatu. Sehingga bisa mempersulit proses perhitungan. Pada proses ini dipakai pendapat independensi yang lebih tinggi (*naif*), bahwa setiap petunjuk ($F_1, F_2 \dots F_n$) saling bebas (*independent*) dengan petunjuk yang lainnya.. Dari pendapat tersebut, maka digunakan rumus berikut:

$$P(F_i|F_j) = \frac{P(F_i \cap F_j)}{P(F_j)} = \frac{P(F_i)P(F_j)}{P(F_j)} = P(F_i) \dots\dots\dots(5)$$

Untuk $i \neq j$, jadi

$$P(F_i|C, F_j) = P(F_i|C) \dots\dots\dots(6)$$

Dari rumus diatas bisa dipakai untuk tahap klasifikasi teorema *Naïve Bayes*. Saat proses klasifikasi dengan data kontinyu memakai rumus *Densitas Gauss*, dengan perhitungan sebagai berikut:

$$P(X_i = x_i|Y = y_j) = \frac{1}{\sqrt{2\pi\sigma_{ij}}} e^{-\frac{(x_i - \mu_{ij})^2}{2\sigma_{ij}^2}} \dots\dots\dots(7)$$

Keterangan:

- P = Peluang
- X_i = Atribut ke-i
- x_i = Nilai atribut ke-i
- Y = Kelas yang dicari
- Y_i = Sub kelas Y yang dicari
- μ = *Mean* (nilai rata-rata)
- σ = *Deviiasi standar* (varian semua atribut)

Penggunaan *Naïve Bayes* juga terdapat kelebihanannya, diantaranya:

- a. Mudah untuk dipahami
- b. Hanya memakai perhitungan yang lebih mudah daripada algoritma yang lainnya.
- c. Lebih cepat dalam perhitungan

