

BAB II

LANDASAN TEORI

Pada bab ini berisi tentang teori-teori yang digunakan dalam penyelesaian permasalahan pada penelitian khususnya pada perancangan dan implementasi sistemnya.

2.1. Keluarga Mampu dan Tidak Mampu

Keluarga adalah unit terkecil dari masyarakat yang terdiri atas kepala keluarga dan beberapa orang yang terkumpul dan tinggal di suatu tempat di bawah suatu atap dalam keadaan saling ketergantungan (Wikipedia, 2014). Keluarga bisa di golongkan menjadi dua golongan yaitu keluarga mampu dan tidak mampu, keluarga mampu yaitu bisa dikatakan keluarga yang bisa memenuhi kebutuhan dan keinginan bahkan lebih dan keluarga tidak mampu yaitu bisa dikatakan keluarga yang belum bisa memenuhi kebutuhan.

Ketentuan sekolah dalam menentukan keluarga mampu dan tidak mampu ialah dari hasil survei yang telah dilakukan pihak sekolah dari data keluarga, dilihat dari fisik rumah dan keluarga tersebut bisa memenuhi kebutuhan keluarga atau tidak.

2.2. Definisi Sistem

Sistem secara fisik adalah kumpulan dari elemen-elemen yang beroperasi bersama-sama untuk menyelesaikan suatu sasaran (Gordon, 1991).

Menurut (Jagianto, 2005) Sistem adalah kumpulan dari elemen-elemen yang berinteraksi untuk mencapai suatu tujuan tertentu.

2.3. Klasifikasi

Klasifikasi adalah proses 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. Dalam mencapai tujuan tersebut, proses klasifikasi membentuk suatu model yang

mampu membedakan data kedalam kelas-kelas yang berbeda berdasarkan aturan atau fungsi tertentu. Model itu sendiri bisa berupa aturan "Jika-maka", berupa pohon keputusan, atau formula matematis (Bustami, 2011)

Tahapan dari klasifikasi dalam data mining menurut (Han dan Kamber, 2006) terdiri dari :

- **Pembangunan Model**

Pada tahapan ini dibuat sebuah model untuk menyelesaikan masalah klasifikasi class atau atribut dalam data. Tahap ini merupakan fase pelatihan, dimana data latih dianalisis menggunakan algoritma klasifikasi, sehingga model pembelajaran direpresentasikan dalam bentuk aturan klasifikasi.

- **Penerapan Model**

Pada tahapan ini model yang sudah dibangun sebelumnya digunakan untuk menentukan atribut/class dari sebuah data baru yang atribut/classnya belum diketahui sebelumnya. Tahap ini digunakan untuk memperkirakan keakuratan aturan klasifikasi terhadap data uji. Jika model dapat diterima, maka aturan dapat diterapkan terhadap klasifikasi data baru.

2.4. Data Mining

Data mining adalah suatu proses mengeksplorasi dan menganalisis data dalam jumlah besar baik secara otomatis maupun semi otomatis untuk mendapatkan suatu pola yang bermakna dari data (Nawawi, 2013).

Menurut (Goela., et al, 2012), Data Mining (DM) adalah proses menganalisis data dengan berbagai perspektif dan meringkasnya kedalam informasi yang berguna, dimana informasi tersebut bisa digunakan untuk meningkatkan pendapatan, memotong biaya atau keduanya.

Dan ada juga menurut (Turban, E., J.E. Aronson dan T.P. Liang. 2005), Data mining adalah proses yang menggunakan teknik statistik, matematika,

kecerdasan buatan, dan machine learning untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar.

2.5. Naïve Bayes Classifier

2.5.1 Teorema Bayes

Menurut (Presetyo, E. 2012). Bayes merupakan teknik prediksi probalistik sederhana yang berdasar pada penerapan teorema Bayes (atau aturan bayes) dengan asumsi independensi (ketidak tergantungan) yang kuat (naif). Dengan kata lain, dalam *Naïve Bayes*, model yang di gunaka adalah "model fitur independen"

Dalam Bayes (terutama Naïve Bayes), maksud independensi yang kuat pada fitur adalah bahwa sebuah fitur pada sebuah data tidak berkaitan dengan ada atau tidak adanya fitur lain dalam data yang sama. Contohnya, pada kasus klasifikasi hewan dengna fitur penutup kulit, melahirkan, berat dan menyusui. Disini ada ketergantungan pada fitur menyusui karena hewan yang menyusui biasanya melahirkan, atau hewan bertelur tidak menyusui. Dalam Bayes, hal tersebut tidak dipandang sehingga masing-masing fitur seolah tidak memiliki hubungan apa pun.

Prediksi Bayes didasarkan pada teorema Bayes dengna formula umum sebagai berikut :

$$P(H|E) = \frac{P(E|H)P(H)}{P(E)} \dots\dots\dots(2.1)$$

Keterangan :

- P(H | E) = Probabilitas akhir bersyarat (conditional probability) suatu hipotesis H terjadi jika diberikan bukti (evidence) E terjadi.
- P(E | H) = Probabilitas sebuah bukti E terjadi akan mempengaruhi hipotesis H .

$P(H)$ = Probabilitas awal hipotesis H terjadi tanpa memandang bukti apapun.

$P(E)$ = Probabilitas awal bukti E terjadi tanpa memandang hipotesis atau bukti yang lain.

Ide dasar dari aturan Bayes adalah bahwa hasil dari hipotesis atau peristiwa (H) dapat diperkirakan pada beberapa bukti (E) yang diamati. Ada beberapa hal penting dari aturan Bayes yaitu :

1. Sebuah probabilitas awal/priori H atau PH adalah probabilitas dari suatu hipotesis sebelum bukti diamati.
2. Sebuah probabilitas akhir H atau (PH) adalah probabilitas dari suatu hipotesis setelah bukti diamati.

Teorema Bayes juga bisa menangani beberapa bukti, misalnya ada $E_1, E_2,$ dan E_3 sehingga akhir untuk hipotesis (H) dapat dihitung dengan cara berikut.

$$P(H|E_1, E_2, E_3) = \frac{P(E_1, E_2, E_3|H) * P(H)}{P(E_1, E_2, E_3)} \dots\dots\dots(2.2)$$

Karena yang digunakan untuk bukti adalah independen, bentuk di atas dapat diubah menjadi.

$$P(H|E_1, E_2, E_3) = \frac{P(E_1|H) * P(E_2|H) * P(E_3|H) * P(H)}{P(E_1) * P(E_2) * P(E_3)} \dots\dots\dots(2.3)$$

2.5.2 Naïve Bayes untuk Klasifikasi

(Prasetyo, E. 2012). Kaitan antara Naïve bayes dengan klasifikasi, korelasi hipotesis, dan bukti dengan klasifikasi adalah bahwa hipotesis dalam teorema Bayes merupakan label kelas yang menjadi target pertama dalam klasifikasi, sehingga bukti-bukti merupakan fitur-fitur yang menjadi masukan dalam model klasifikasi.

Formulasi Naïve Bayes untuk klasifikasi adalah :

$$P(Y | X) = \frac{P(Y) \prod_{i=1}^q P(X_i|Y)}{P(X)} \dots\dots\dots(2.4)$$

Keterangan :

$P(Y | X)$ = probabilitas data dengan vektor X pada kelas Y

$P(Y)$ = probabilitas awal kelas Y

$\prod_{i=1}^q P(X_i|Y)$ = probabilitas independen kelas Y dari semua fitur dalam vector X

Karena $P(X)$ selalu tetap, sehingga dalam perhitungan prediksi nantinya cukup hanya dengan menghitung $P(Y) \prod_{i=1}^q P(X_i|Y)$ dengan memilih yang terbesar sebagai kelas yang di pilih sebagai hasil prediksi. Sementara probabilitas independen $\prod_{i=1}^q P(X_i|Y)$ tersebut merupakan pengaruh semua fitur dari data terhadap setiap kelas Y, yang dinotasikan dengan.

$$P(X|Y=y) = \prod_{i=1}^q P(X_i|Y=y) \dots\dots\dots(2.5)$$

setiap fitur $X = \{X_1, X_2, X_3, \dots, X_q\}$ terdiri atas q atribut (q dimensi).

Umumnya, metode Naïve Bayes ini mudah dihitung untuk fitur bertipe kategoris. Namun untuk tipe numerik (kontinu), ada perlakuan khusus sebelum dimasukkan dalam Naïve Bayes, yaitu :

1. Melakukan diskretisasi pada setiap fitur kontinu dan mengganti nilai fitur kontinu tersebut dengan nilai interval diskret. Pendekatan ini dilakukan dengan mentransformasi fitur kontinu kedalam fitur ordinal.
2. Dari distribusi probabilitas diasumsikan bentuk tertentu untuk fitur kontinu dan memperkirakan parameter distribusi dengan data peralihan. Distribusi Gaussian biasanya dipilih untuk mempresentasikan probabilitas bersyarat dari fitur kontinu pada sebuah kelas $P(X_i|Y)$, sedangkan distribusi Gausaian

dikarakteristikan dengan dua parameter. mean μ , dan varian, σ^2 . Untuk setiap kelas y_j , probabilitas bersyarat kelas y_j untuk fitur X_i adalah

$$P(X_i = x_i | Y = y_j) = \frac{1}{\sqrt{2\pi\sigma_{ij}^2}} \exp \left\{ -\frac{(x_i - \mu_{ij})^2}{2\sigma_{ij}^2} \right\} \dots\dots\dots(2.6)$$

Keterangan :

μ_{ij} : mean sampel X_i (\bar{X}) dari semua data latih yang menjadi milik kelas y_j

σ_{ij}^2 : varian sampel (S^2) dari data latih.

2.6. Penelitian Sebelumnya

Naïve Bayes merupakan metode populer yang banyak di gunakan untuk klasifikasi. Berbagai riset yang telah dilakukan berkaitan dengan kasus prediksi yang menggunakan metode Naïve Bayes, antara lain :

Amir Hamzah, lulusan Institut Sains dan Teknologi AKPRID Tahun 2012. Penelitiannya yang berjudul : “*Klasifikasi teks dengan menggunakan Naïve Bayes Classifier (NBC) untuk mengklompokan teks berita dan abstrak akademis*”. Dalam keakurasian Algoritma NBC ini, memiliki kinerja yang cukup tinggi untuk klasifikasi dokumen teks, baik dokumen berita maupun dokumen akademik. Pada klasifikasi dokumen berita didapatkan akurasi yang lebih tinggi (maksimal 91%) dibandingkan dengan dokumen akademik (maksimal 82%).

Nur Indah Sari, lulusan Universitas Muhammadiyah Gresik Tahun 2013. Penelitiannya berjudul : “*klasifikasi kecendrungan penyelesaian studi mahasiswa baru dengan menggunakan metode Naïve Bayes*”. Atribut yang dilakukan sebagai data latih sistem adalah NP (Nomor Pegawai), nama, umur, jenis kelamin, pendidikan, agama, level, lama kerja , kesehatan dan jadwal. Atribut jadwal digunakan sebagai kelas data. Adapun data yang diambil dalam penelitian ini adalah sampel dari 40 *record* dengan kelas “Lama” dan “Tidak Lama” masing-masing berjumlah 19 untuk kelas lama dan 21 untuk kelas tidak lama yang akan dibagi menjadi datalatih data uji. Dan menggunakan 30 data latih, kelas lama 19

dan kelas tidak lama 21. 10 data uji kelas lama 5 dan kelas tidak lama 5. Hasil dari 10 data uji, terdapat 3 data yang hasil klasifikasi dengan mengaunkan metode naïve bayes yang tidak sama dengan hasil data yang sebenarnya. Keakurasian ketepatan hasil klasifikasi dengan naïve bayes untuk mengetahui tingkat kecenderungan penyelesaian studi mahasiswa sistem memiliki akurasi kebenaran sistem sebesar 70% dan nilai error sebesar 30%.

Penelitian ketiga di lakukan oleh Meinggalan Vilian Sari, lulusan Universitas Muhammadiyah Gresik Tahun 2014. penelitian untuk “*memprediksi prestasi (IPK) mahasiswa berdasarkan latar belakang sekolah asal dan atribut mahasiswa ketika awal masuk kuliah menggunakan Naïve Baye*”. Adapun data yang diambil dalam penelitian ini adalah sampel dari 103 *record* dengan kelas “Tinggi” dan “Rendah” masing-masing berjumlah 69 dan 34 yang akan dibagi menjadi data latih data uji. Dan menggunakan 6 variable, adapun variabel yang dipakai : Instansi Sekolah, Setatus Sekolah, Jurusan Sekolah, Motivasi Pilihan Kuliah, Status Kerja, Nilai Danem. Dan keakurasian hasil penelitian menggunakan metode Naïve Bayes ini menunjukkan akurasi tertinggi pada pengujian pertama adalah 84.62%. dan pada pengujian keempat, ketiga percobaan menghasilkan nilai sensitivitas 100%, yang artinya semua data uji kelas tinggi diprediksi secara benar mempunyai kelas tinggi.