

## **BAB II**

### **TINJAUAN PUSTAKA**

Dalam bab tinjauan pustaka ini akan menjelaskan tentang seluruh teori, bahan penelitian lain yang diarahkan untuk menyusun konsep yang berkaitan dengan penelitian dan terdiri dari penjelasan studi-studi sebelumnya dan dasar-dasar teori yang digunakan. Teori yang digunakan diambil dari literature dan jurnal.

#### **2.1 Penerimaan Siswa Baru**

Penerimaan Siswa Baru atau PPDB pada sekolah yang dianalisis oleh penulis mempunyai berbagai aspek atau kategori untuk penerimaan siswa baru ini yang dimana nanti aspek-aspek tersebut akan dijadikan acuan untuk perhitungan prediksi yang akan dianalisis oleh penulis antara lain :

1. Nilai Hasil Akhir Ujian Nasional
2. Tes Membaca Al-Qur'an
3. Tes Kepribadian

Dari tiga aspek atau kategori diatas penulis mendapatkan sumber datanya dari pihak Tata Usaha Sekolah Menengah Kejuruan NU 1 Sukodadi, dari tiga kategori yang diatas maka nanti kategori tersebut akan dimasukkan kedalam perhitungan prediksi.

#### **2.2 Penelitian Terdahulu**

Penelitian terdahulu menjadi sebuah acuan dalam penulisan atau penelitian sehingga penulis mempunyai banyak teori untuk digunakan dalam pengkajian penelitian yang akan dilakukan. Penulis juga menemukan berbagai judul yang hampir sama dengan apa yang dianalisis oleh penulis. Ada beberapa penelitian sebagai refrensi dalam memperkaya

bahan kajian pada penelitian kali ini. Berikut merupakan penelitian terdahulu berupa jurnal terkait dengan penelitian yang dilakukan penulis:

Penelitian terdahulu pertama, yang dilakukan oleh Putri dan surahman (2020) dengan judul “Penerapan Model Naïve Bayes Untuk Memprediksi Potensi Pendaftaran Siswa di SMK Taman Siswa Teluk Betung Berbasis Web”. Permasalahan yang dibahas pada penelitian ini adalah adanya persaingan antar SMK di wilayah tersebut serta tidak menutup kemungkinan akan terjadi sebuah kenaikan dan penurunan yang tidak ideal pada SMK Taman Siswa. Dalam penelitian ini penulis mengambil beberapa data terdahulu untuk diprediksi jumlah siswa baru yang akan masuk pada SMK Taman Siswa. Data diambil data siswa pada tahun pelajaran 2017/2018 – 2018/2019. Pada penelitian ini menggunakan metode Naïve Bayes untuk memprediksi data sebelumnya untuk data dimasa depan, penelitian ini juga menggunakan parameter yaitu usia, jenis kelamin, alamat, asal pendidikan, alamat pendidikan, jurusan, agama, prediski potensi pendaftaran. Metode yang digunakan adalah metode naïve bayes yang dimana memiliki kemampuan untuk memprediksi potensi banyaknya siswa yang mendaftar dan tidak mendaftar di SMK Taman Siswa. Sistem ini diuji dan menghasilkan akurasi sebesar 87% untuk prediksi potensi pendaftaran siswa.

Penelitian terdahulu kedua, yang dilakukan oleh Sinaga, Solikhun dan jalalludin (2020) dengan judul “Prediksi jumlah siswa baru pada SMK swasta abdi sejati kerasaan dengan metode backpropagation”. Permasalahan yang dibahas pada penelitian ini adalah tidak stabilnya pendaftaran siswa ditahun yang akan datang atau naik turunnya siswa yang mendaftar disekolah tersebut, yang dapat berpengaruh terhadap kualitas sekolah. Apabila terjadi suatu peningkatan di tahun ayang akan datang maka pihak sekolah akan melakukan antisipasi dalam memenuhi kualitas maupun kuantitas fasilitas sekolah yang tersedia, sedangkan jika ditahun berikutnya ada penurunan siswa maka itu adalah kerugian bagi sekolah karena kurangnya kuota yang ada.

Pada penelitian ini penulis mengumpulkan data siswa dari tahun 2014 – 2018 untuk dijadikan sebuah data *training* yang dimana nantinya akan dihitung normalisasinya untuk mendapatkan nilai atau data siswa yang mendaftar di sekolahn tersebut ditahun berikutnya. Dengan hasil model arsitektur terbaik adalah 3-22-1 dengan tingkat akurasi 75% MSE pelatihan 0,000993661 serta MSE pengujian sebesar 0,146896423.

Penelitian terdahulu ketiga, yang dilakukan oleh Rizal dan Lutfi (2018) dengan judul “Penerapan algoritma Naïve Bayes untuk prediksi Penerimaan Siswa Baru di SMK Al-Amien Wonorejo” Setiap tahunnya, jumlah data siswa yang masuk tidak stabil pada tahun pertama ada penambahan siswa juga ada penurunan siswa. Data tersebut juga tersimpan dalam bentuk file maupun hard file yang dimana data-data tersebut dapat menumpuk dan tersimpan oleh sekolahan. Hal tersebut biasanya disebut sebagai sebuah situasi “*data rich but information poor*” yang dimana itu artinya data yang disimpan sangat banyak atau berlimpah namun sebuah informasinya yang diperoleh sangat kurang. Dengan adanya data yang besar dan dikumpulkan atau disimpan setiap tahunnya jauh melampaui kemampuan seorang manusia untuk menganalisis data tanpa teknik yang tepat. Maka dengan adanya data yang seperti itu menyebabkan data yang dikumpulkan menjadi suatu data yang tidak bermanfaat.

Untuk mengatasi hal tersebut diperlukan suatu teknik dalam *data mining* yang dimana datatersebut bisa memberikan manfaat berupa informasi yang penting. Hasil pengujian perhitungan Naïve Bayes dalam prediksi penerimaan siswa baru di SMK pada 196 data siswa yang diuji dalam penelitian ini, menunjukkan bahwa metode Nive Bayes memiliki tingkat akurasi sebesar 86.22%.

### 2.3 Dasar Teori

Pada dasar teori kali ini penulis akan menjelaskan satu persatu teori-teori yang digunakan untuk penelitian kali ini, yang dimana datanya diambil dari beberapa litelatur.

### 2.3.1 Pengertian Prediksi

Prediksi adalah suatu proses memperkirakan secara sistematis tentang sesuatu yang paling mungkin terjadi di masa depan berdasarkan informasi masa lalu dan sekarang yang dimiliki, agar kesalahannya (selisih antara sesuatu yang terjadi dengan hasil perkiraan) dapat diperkecil. (Manalu E, Sianturi & Manalu R., 2017).

Peramalan atau *forecasting* adalah suatu prosedur untuk membuat informasi factual tentang situasi sosial dimasa depan atas dasar informasi yang telah ada tentang masalah kebijakan.

Proyeksi adalah ramalan yang didasarkan pada ekstrapolasi atas kecenderungan masa lalu maupun masa kini ke masa depan.

Sedangkan sebuah prediksi adalah ramalan yang didasarkan pada asumsi teoritik yang tegas. Asumsi ini dapat berbentuk (hukum teoritis), misalnya hukum berkurangnya nilai uang, (proposisi teoritis) misalnya proposi bahwa pecahan masyarakat sipil diakibatkan oleh kesenjangan antara harapan dan kemampuan, atau (analogi) misalnya analogi antara pertumbuhan organisasi pemerintah dengan pertumbuhan organisme biologis.

### 2.3.2 Metode Naive Bayes

*Naive Bayes* adalah suatu metode pengklasifikasikan probabilistic sederhana yang menghitung sekumpulan data probabilitas dengan cara menjumlahkan frekuensi dan kombinasi sebuah nilai dari dataset yang diberikan. (Harimurti 2017).

Klasifikasi Naive Bayes adalah suatu data yang diprediksi peluang di masa depan berdasarkan pengalaman dimasa sebelumnya sehingga dapat disebut sebagai Teorema Bayes. Klasifikasi Naive Bayes diasumsikan bahwa ada atau tidak ciri tertentu dari sebuah kelas tidak berhubungan dengan ciri dari kelas lain.

Persamaan dari metode Teorema Bayes adalah sebagai berikut :

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

Keterangan :

X : Data kelas yang belum diketahui

H : Hipotesis data X adalah suatu kelas spesifik

P(H|X) : Probabilitas hipotesis data H berdasarkan kondisi X (*posteriori probability*)

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

P(X|H) : Probabilitas X berdasarkan kondisi pada hipotesis H

P(X) : Probabilitas X

Untuk menjelaskan teorem *Naïve Bayes*, perlu diketahui bahwa proses klasifikasi perlu sejumlah petunjuk untuk menentukan kelas yang cocok bagi sampel yang dianalisis. Maka dari itu, teorema bayes di atas disesuaikan seperti berikut:

$$P(C|F1 \dots Fn) = \frac{P(c)P(F1\dots Fn|C)}{P(F1 \dots Fn)} \dots\dots\dots (2.2)$$

Diketahui variable C merepresentasikan kelas, sementara variable F1 ... Fn merepresentasikan karakteristik petunjuk yang dibutuhkan untuk melakukan klasifikasi. Maka dari itu, rumus tersebut menjelaskan bahwa peluang masuknya sampel karakteristik tertentu dalam kelas C (Posterior) adalah peluang munculnya kelas C sebelum masuk sampel (biasanya disebut prior), dikali dengan peluang munculnya karakteristik sampel pada kelas C (biasanya disebut *likelihood*), dibagi dengan peluang munculnya karakteristik sampel secara global (biasanya disebut *evidence*). Maka rumus diatas bisa ditulis dengan sederhana yaitu sebagai berikut:

$$\text{Posterior} = \text{Prior} \times \text{Likelihood} / \text{Evidence} \dots\dots\dots (2.3)$$

Nilai *Evidence* akan tetap untuk setiap kelas pada satu sampel. Nilai dari posterior tersebut akan dibandingkan dengan nilai-nilai posterior kelas lainnya untuk dapat menentukan ke kelas suatu sampel yang akan

diklasifikasikan. Penjelasan lebih lanjut dalam rumus Bayes akan dilakukan dengan menjabarkan  $(C|F_1...F_n)$  menggunakan aturan perkalian sebagai berikut:

$$\begin{aligned}
 (C|F_1...F_n) &= (C)(F_1...F_n|C,F_1) \\
 &= (C)P(F_1|C)P(C)P(F_2...F_n|C,F_1) \\
 &= P(C)P(F_1|C)P(F_2|C,F_1)P(F_3,...,F_n|C,F_1,F_2) \\
 &= P(C)P(F_1|C)P(F_2|C,F_1)P(F_3|C,F_1,F_2)P(F_4,...,F_n|C,F_1,F_2,F_3) \\
 &= P(C)P(F_1|C)P(F_2|C,F_1)P(F_3|C,F_1,F_2)...P(F_n|C,F_1,F_2,F_3,...,F_{n-1}).... \quad (2.4)
 \end{aligned}$$

Dapat diketahui bahwa hasil penjabaran tersebut menyebabkan semakin banyak dan semakin kompleks faktor-faktor syarat yang mempengaruhi nilai probabilitas, hampir tidak memungkinkan untuk dilakukan analisis satu persatu. Maka dari itu, perhitungan tersebut menjadi sangat sulit untuk dilakukan perhitungan. Digunakan asumsi independensi yang sangat tinggi (naif), bahwa setiap petunjuk  $F_1, 2 \dots F_n$  saling bebas (independen) satu dengan lainnya. Dengan asumsi tersebut, maka berlaku suatu kesamaan seperti berikut.

$$P(P_i|F_j) = P(F_i \cap F_j)P(F_j) = P(F_i)P(F_j)P(F_j) = P(F_i).$$

Untuk  $i \neq j$ , sehingga didapatkan sebagai berikut:

$$(P_i|C, F_j) = (P_i|C).....(2.5)$$

Persamaan diatas adalah model dari teorema Naïve Bayes yang selanjutnya akan digunakan dalam suatu proses klasifikasi. Klasifikasi dengan data lanjutan, maka digunakan rumus *Densitas Gauss* sebagai berikut:

$$P = (X_i = x_i | Y_i = y_i) = \frac{1}{\sqrt{2\pi\sigma_{ij}}} e^{-\frac{(x_i - \mu_{ij})^2}{2\sigma_{ij}^2}} .....(2.6)$$

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
- $u$  : Mean, menyatakan rata-rata dari seluruh atribut
- $\sigma$  : Deviasi standar, menyatakan varian dari seluruh atribut

Setelah itu ada alur dari metode Naïve Bayes yaitu sebagai berikut:

1. Baca data Training
2. Hitung jumlah dan probabilitas, jika data numerik maka harus dilakukan:
  - a. Cari nilai mean dan standar deviasi dari masing-masing parameter data numerik.
  - b. Cari nilai probabilistic dengan cara menghitung jumlah data yang sesuai dari kategori yang sama dibagi dengan jumlah data pada kategori tersebut.
3. Menghasilkan nilai dalam table mean, standar devisi, dan probabilitas
4. Solusi dapat dihasilkan.

### 2.3.3 Pengertian Web

World Wide Web adalah suatu program yang ditemukan oleh Tim Berners-Lee pada tahun 1980an dan diresmikan pada tahun 1991. Web pada dasarnya adalah sebuah kumpulan halaman pada suatu domain di internet yang dibuat dengan tujuan tertentu. Dengan adanya program ini, dapat mempermudah untuk mendapatkan sebuah informasi. (Hidayatullah & Kawistara 2017)

Web akhir-akhir ini sering digunakan oleh para user untuk melakukan sebuah pekerjaan seperti web Pribadi, Web bisnis dan Blog. Pada kesempatan ini penulis ingin membuat sistem prediksi penerimaan siswa baru menggunakan Web dan penelitian ini termasuk pada Web Blog.

#### 2.3.4 Pengertian Xampp

Xampp adalah perangkat lunak atau software yang berfungsi untuk menjalankan website dengan Bahasa Pemrograman PHP dan menggunakan pengolahan data MySQL di komputer lokal. Xampp berperan sebagai server *Virtual*, yang bisa membantu melakukan *preview* sehingga dapat mengerjakan Web tanpa tersambung dengan Internet.

#### 2.3.5 Pengertian MySQL

MySQL ialah salah satu program yang ada pada Xampp, MySQL merupakan database management sistem yang menggunakan perintah dasar SQL (*Structured Query Language*) yang cukup terkenal oleh pengguna Bahasa Pemrograman. MySQL masuk kedalam jenis RDBMS (*Relation Database Management System*). Maka dari itu istilah seperti baris, kolom, tabel dipakai pada MySQL. Salah satu contohnya didalam MySQL sebuah database terdapat satu atau beberapa tabel. (Hidayatullah & Kawistara) SQL merupakan bahasa pemrograman yang biasa dipakai dalam pengambilan data pada relation database.