

# Prediksi Waktu Tunggu Pekerjaan Bagi Lulusan Perguruan Tinggi Menggunakan Algoritma Naïve Bayes

Dyah Retno Utari<sup>1</sup>

Program Studi Manajemen Informatika Fakultas Teknologi Informasi, Universitas Budi Luhur  
Jl. Cileduk Raya 99 Petukangan Utara, Jakarta Selatan DKI Jakarta

<sup>1</sup>dyah.retnoutari.budiluhur.ac.id

*Abstrak— Perguruan Tinggi mempersiapkan para lulusan agar siap menghadapi dunia kerja. Salah satu indikator keberhasilan itu adalah dengan mengetahui waktu tunggu dalam memperoleh kerja pertama kali. Bila waktu tunggu mendekati nol bulan atau negatif (sebelum lulus sudah mendapatkan pekerjaan) maka hal tersebut dinilai baik. Penelitian ini bertujuan untuk memprediksi masa tunggu lulusan dalam memperoleh pekerjaan pertamanya. Permasalahan diselesaikan menggunakan teknik data mining klasifikasi dengan algoritma Naïve bayes. Berdasarkan hasil analisis yang dilakukan menggunakan 199 data training dan 22 data testing didapatkan tingkat akurasi sebesar 90.90%, recall sebesar 90.48%, presisi 100%. Model prediksi ini kemudian berhasil diimplementasikan dalam prototipe aplikasi yang dimanfaatkan oleh pengelola program studi.*

**Kata kunci— prediksi waktu tunggu pekerjaan, lulusan, data mining, naïve bayes**

## I. PENDAHULUAN

Salah satu peran perguruan tinggi adalah berkomitmen menghasilkan lulusan yang berkualitas untuk memenangkan kompetisi dunia kerja. Untuk mencapai kesuksesan kompetisi di dunia kerja diperlukan proses pendidikan terbaik di Perguruan tinggi. Semakin tinggi tingkat pendidikan yang telah ditempuh maka seharusnya semakin berkualitas pula lulusan yang dihasilkan. Salah satu indikator tentang kualitas lulusan adalah bagaimana lulusan ini mampu bersaing di dunia kerja dengan waktu tunggu yang singkat untuk mendapatkan pekerjaannya. Berbagai upaya dilakukan oleh Perguruan Tinggi untuk mempersiapkan output ini agar siap dalam menghadapi dunia kerja salah satunya dengan cara memprediksi masa tunggu mendapatkan pekerjaan. Dengan adanya prediksi dini maka Perguruan tinggi dapat meningkatkan kualitas pendidikan terhadap mahasiswa agar mahasiswa tersebut lebih unggul dan lebih cepat dalam mendapatkan pekerjaan.

Berdasarkan permasalahan tersebut maka studi ini melakukan analisis dan memprediksi waktu tunggu mahasiswa untuk mendapatkan pekerjaan pertamanya. Proses analisis tersebut menggunakan Teknik Data Mining dan akan dianalisa menggunakan algoritma Naïve Bayes.

Beberapa penelitian terdahulu yang membahas tentang data mining menggunakan algoritma Naïve Bayes telah banyak dilakukan, misalnya prediksi penerimaan siswa baru, penelitian ini bertujuan untuk memprediksi penerimaan siswa baru, dengan menganalisis menggunakan beberapa data siswa SMP dianalisis untuk menentukan tingkat penerimaan siswa di SMK dimana bertujuan untuk mengetahui Hasil pengujian algoritma Naïve Bayes dalam prediksi penerimaan siswa baru di Sekolah Menengah Kejuruan [1].

Penelitian lainnya yang berkaitan dengan Algoritma Naïve Bayes yang membahas tentang prediksi kegagalan siswa. Penelitian ini bertujuan untuk meningkatkan keberhasilan siswa dalam studi, agar tidak terjadi lagi siswa yang gagal. Berdasarkan pengujian, algoritma Naïve Bayes menghasilkan tingkat akurasi sebesar 77.97% dari 395 dataset untuk memprediksi kegagalan siswa [2].

Penelitian lainnya yang berkaitan dengan Algoritma Naïve Bayes yang membahas tentang klasifikasi penerima beasiswa. penelitian ini dilakukan untuk untuk menganalisis kelayakan penerima beasiswa. Dalam penelitian ini dibuktikan dengan kemampuan Naive Bayes Classifier untuk mengklasifikasikan data pendaftar beasiswa PPA. Sehingga menghasilkan model probabilitas klasifikasi untuk penentuan kelas pada pendaftar beasiswa selanjutnya. Dari hasil pengujian akurasi model dari sistem yang dikembangkan, menghasilkan nilai akurasi terkecil sebesar 64% pada pengujian dengan sampel sebanyak 100 data dan menghasilkan nilai akurasi tertinggi sebesar 97,66% [3].

Penelitian ini bertujuan untuk membuat prediksi sejak awal sehingga dapat di lakukan langkah-langkah antisipatif dari pihak kampus agar mahasiswa mendapatkan pekerjaan dengan waktu yang cepat.

## II. LANDASAN TEORI

Data Mining merupakan suatu proses menemukan hubungan yang berarti kecenderungan dengan memeriksa sekumpulan besar data yang tersimpan dalam penyimpanan dengan Teknik pengenalan pola seperti teknik statistik dan matematika. Ketersediaan data yang melimpah, kebutuhan akan informasi atau pengetahuan sebagai pendukung pengambilan keputusan untuk membuat solusi bisnis, dan

dukungan infrastruktur di bidang teknologi informasi merupakan pendukung dari pemrosesan data mining. Data Mining dimanfaatkan untuk mendeteksi kejadian - kejadian yang ganjil seperti berbagai penerapan data mining yakni Analisa pasar dan manajemen, Analisa perusahaan dan manajemen resiko, Telekomunikasi, Keuangan, Asuransi, Olahraga dan Astronomi. Sebagai suatu teknologi yang dapat menghasilkan *knowledge*, data mining melakukan pemrosesan dalam beberapa tahap [4].

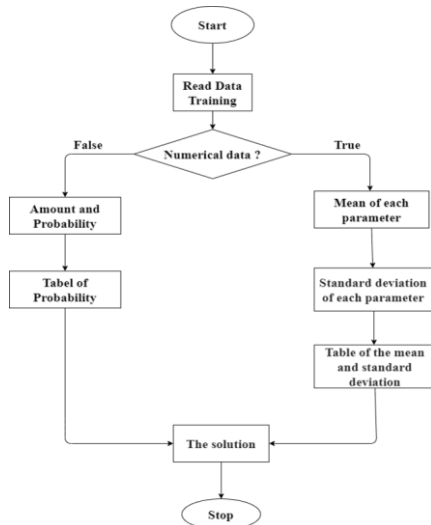
Metode Naïve Bayesian merupakan salah satu teknik data mining yang digunakan untuk memprediksi suatu kejadian pada masa yang akan datang, dengan cara membandingkannya dengan data atau evidence (bukti) yang ada pada masa lampau. Penggunaan probabilitas kata atau token dijadikan sebagai inputan probabilitas dari kejadian. Klasifikasi Naïve Bayesian akan melihat data lama dalam menentukan nilai kemiripan data yang baru. Jadi harus terdapat data lama yang digunakan sebagai data pembanding dalam proses bayes. Persamaan dari teorema Bayes adalah [5]:

$$P(X|H) = \frac{P(X|H)P(H)}{P_X} \quad P(X|H) = \frac{P(X|H)P(H)}{P_X} \quad (1)$$

Dimana:

- X : Data dengan class yang belum diketahui
- H : Hipotesis data merupakan suatu class spesifik
- P(H|X) : Probabilitas hipotesis H berdasarkan kondisi X (*posteriori probabilitas*)
- P(H) : Probabilitas hipotesis H (*prior probabilitas*)
- P(X|H) : Probabilitas X berdasarkan kondisi pada hipotesis H
- P(X) : Probabilitas X

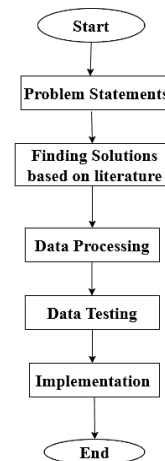
Umumnya, Bayes mudah dihitung untuk fitur bertipe kategoris. Jika fitur memiliki data yang berupa numerik (kontinyu) maka diperlukan perlakuan khusus sebelum dimasukkan dalam *Naïve Bayes*. Adapun tahap dari metode *Naïve Bayes* dapat dilihat dari gambar berikut [6]:



Gambar 1. Tahap Naïve Bayes

### III. METODE PENELITIAN

Penelitian ini terdiri dari beberapa tahapan dimulai dari penentuan rumusan masalah, studi pustaka untuk menemukan alternatif solusi, dan dilanjutkan dengan metode data mining yang terdiri dari pemrosesan data, pelatihan dan dan pengujian data. Tahap akhir studi ini adalah membangun prototipe aplikasi prediksi masa tunggu, sebagai media penghayatan pola atau pengetahuan yang diperoleh dari metode data mining. Alur penelitian ini dapat dilihat pada Gambar 2.



Gambar 2. Alur Penelitian

Hal yang pertama dilakukan dalam penelitian ini yaitu melakukan studi pustaka yang berkaitan dengan penelitian yang sebelumnya tentang penggunaan metode naïve bayes. Dalam kasus ini peneliti ingin melakukan prediksi tentang masa waktu tunggu mahasiswa dalam mendapatkan pekerjaan dengan menggunakan teknik data mining.

Setelah melakukan studi Pustaka tahap selanjutnya yaitu menentukan metode yang tepat, dalam penelitian ini peneliti memilih algoritma Naïve Bayes karena dapat digunakan untuk menghitung sekumpulan probabilitas dari kombinasi nilai dalam dataset.

Pada penelitian ini yang menjadi objek penelitian adalah alumni Akademi Sekretari Budi luhur di Jakarta. Langkah pada pengolahan data ini meliputi pengumpulan data, penyeleksian data, pembersihan data, dan transformasi data. Setelah pengolahan data maka akan diperoleh dataset yang dapat digunakan untuk proses selanjutnya.

Pada tahap selanjutnya, data di uji dengan metode naïve bayes. Pengujian dilakukan dengan cara perhitungan manual dan membandingkan hasilnya dengan prototipe aplikasi. Dari model prediksi yang dihasilkan selanjutnya dilakukan perhitungan akurasi, presisi dan recall, dengan menggunakan confusion matrix.

#### IV. HASIL DAN PEMBAHASAN

Pada penelitian ini yang menjadi obyek penelitian adalah data historikal alumni Akademi Sekretari Budi Luhur yang telah memperoleh pekerjaan. Langkah atau tahapan pada pengolahan data penelitian ini meliputi pengumpulan data, penyeleksian data, pembersihan data, transformasi data.

Pada tahap pengumpulan data, didapatkan data primer sebanyak 314 *record* untuk dilakukan seleksi data. Seleksi data yaitu proses dimana penyeleksian dari banyaknya atribut yang ada, maka diseleksi berdasarkan kebutuhan dalam penelitian ini, maka penulis menggunakan beberapa atribut untuk dijadikan parameter dalam memprediksi waktu tunggu kerja, atribut yang akan digunakan dapat dilihat pada Tabel 1.

Table 1. Atribut Dataset

Nama Atribut	Keterangan
Nim	Nomor Induk Mahasiswa
Sekolah Asal	Asal SMA/ Sederajat
Asal Kota Sekolah	Asal Kota Sekolah
Radius Sekolah	Radius Sekolah asal menuju kampus
Kategori Sekolah	Negeri / Swasta
Usia Masuk	Usia saat masuk
IPS1	Nilai Indeks Prestasi Semester 1
IPS2	Nilai Indeks Prestasi Semester 2
IPS3	Nilai Indeks Prestasi Semester 3
Waktu Tunggu	Waktu tunggu kerja Positif/Negatif

Pada tahap selanjutnya dilakukan pembersihan data, untuk mengeliminasi data yang masih bersifat *null* agar mempermudah dalam proses perhitungan. Setelah pembersihan data dilakukan maka didapatkan data sebanyak 221 data, yang akan diolah menjadi *data training* dan *data testing*.

Proses transformasi data dilakukan agar data tidak memiliki variasi nilai yang terlalu banyak. Hal ini dilakukan terhadap dua atribut yaitu asal sekolah dan kota sekolah yang akan ditransformasi menjadi jarak sekolah. Perubahan data juga dilakukan pada atribut yang akan dijadikan Label pada penelitian ini yaitu waktu tunggu yang *value-nya* di inialisasikan dapat dilihat pada Tabel 2.

Tabel 2. Inisialisasi Data

Keterangan	Status Waktu Tunggu
Mendapatkan pekerjaan setelah kelulusan	Positif
Mendapatkan pekerjaan sebelum kelulusan	Negatif

Setelah data selesai ditransformasi maka selanjutnya, diimplementasikan analisis dengan formula *Naïve Bayes*. Adapun cara kerja dari proses perhitungan *Naïve Bayes* yaitu diawali dengan mengambil data latih.

Dari data *training* dilakukan perhitungan *Naïve Bayes* menggunakan langkah-langkah sebagai berikut:

- Menghitung jumlah Class / Label  
 $P(Y = \text{Positif}) = 163/201 = 0.81095$   
 “jumlah data Positif dibagi dengan jumlah keseluruhan data”  
 $P(Y = \text{Negatif}) = 38/201 = 0.18905$   
 “jumlah data Negatif dibagi dengan jumlah keseluruhan data”
- Menghitung jumlah kasus yang sama dengan class yang sama, dalam penelitian ini ada 2 tipe atribut yaitu data yang bersifat kategorial dan data yang bersifat numerik. seperti yang dapat dilihat di gambar 1, untuk atribut kategorial dapat langsung dihitung probabilitasnya, tetapi untuk atribut yang bersifat numerik untuk menghitung probabilitasnya harus di cari dahulu nilai mean dan standar deviasinya setelah itu hitung probabilitasnya dengan menggunakan formula densitas gauss yang dapat dilihat pada formula 2 [6].

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

Probabilitas tipe data kategorial:

$$P(\text{Kategori sekolah} = \text{Negeri} | Y = \text{Positif}) = 71/163 = 0.4355828$$

$$P(\text{Kategori sekolah} = \text{Negeri} | Y = \text{Negatif}) = 13/38 = 0.4210526$$

$$P(\text{Kategori sekolah} = \text{Swasta} | Y = \text{Positif}) = 92/163 = 0.5644171$$

$$P(\text{Kategori sekolah} = \text{Swasta} | Y = \text{Negatif}) = 22/38 = 0.5789473$$

Probabilitas tipe data numerik:

Tahap pertama yaitu Menghitung nilai mean, adapun formula untuk menghitung rata-rata (mean) [1], dengan contoh berikut:

$$\mu = \frac{x_1 + x_2 + x_3 + \dots + x_n}{n} \quad (2)$$

Keterangan:

$\mu$  = rata-rata hitung (mean)

$x_i$  = nilai sampel ke-i

$n$  = jumlah sampel

Radius sekolah dengan class Positif

$$= \frac{0+2+2+2+0+0+2+0+2+3+0+\dots+C}{163} = 1.061$$

Radius sekolah class Negatif

$$= \frac{\sqrt{\frac{2+3+0+0+0+2+1+0+3+0+0+\dots+3}{38}}}{\frac{2+3+0+0+0+2+1+0+3+0+0+\dots+3}{38}}$$

$$= 1.079$$

Usia masuk dengan class Positif

$$= \frac{\sqrt{\frac{19+19+20+20+18+18+18+18+\dots+18}{163}}}{\frac{19+19+20+20+18+18+18+18+\dots+18}{163}}$$

$$= 18.436$$

Usia masuk dengan class Negatif

$$= \frac{\sqrt{\frac{18+20+18+17+19+20+18+18+18+\dots+20}{38}}}{\frac{18+20+18+17+19+20+18+18+18+\dots+20}{38}}$$

$$= 18.921$$

IPS1 dengan class Positif

$$= \frac{\sqrt{\frac{3.00+2.52+2.96+3.61+3.61+3.43+3.09+\dots+4.00}{163}}}{\frac{3.00+2.52+2.96+3.61+3.61+3.43+3.09+\dots+4.00}{163}}$$

$$= 3.329$$

IPS1 dengan class Negatif

$$= \frac{\sqrt{\frac{2.61+2.43+2.91+3.22+2.35+2.78+2.43+\dots+3.26}{38}}}{\frac{2.61+2.43+2.91+3.22+2.35+2.78+2.43+\dots+3.26}{38}}$$

$$= 3.141$$

IPS2 dengan class Positif

$$= \frac{\sqrt{\frac{3.42+2.63+2.68+3.67+3.58+3.25+3.63+\dots+2.60}{163}}}{\frac{3.42+2.63+2.68+3.67+3.58+3.25+3.63+\dots+2.60}{163}}$$

$$= 3.484$$

IPS 2 dengan class Negatif

$$= \frac{\sqrt{\frac{3.05+3.13+3.05+3.50+2.93+3.05+2.93+\dots+3.65}{38}}}{\frac{3.05+3.13+3.05+3.50+2.93+3.05+2.93+\dots+3.65}{38}}$$

$$= 3.256$$

IPS3 dengan class Positif

$$= \frac{\sqrt{\frac{3.38+3.26+2.95+3.48+4.00+3.33+3.52+\dots+2.00}{163}}}{\frac{3.38+3.26+2.95+3.48+4.00+3.33+3.52+\dots+2.00}{163}}$$

$$= 3.518$$

IPS3 dengan class Negatif

$$= \frac{\sqrt{\frac{2.76+3.33+3.14+3.48+2.90+3.00+2.76+\dots+3.68}{38}}}{\frac{2.76+3.33+3.14+3.48+2.90+3.00+2.76+\dots+3.68}{38}}$$

$$= 3.331$$

Tahap kedua yaitu Mencari nilai standar deviasi, adapun formula untuk menghitung standar deviasi sebagai berikut [1]:

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (xi - \mu)^2}{n}} \quad \sigma = \sqrt{\frac{\sum_{i=1}^n (xi - \mu)^2}{n}} \quad (3)$$

Keterangan:

$\sigma$  = standar deviasi

$xi$  = nilai x ke-i

$\mu$  = rata-rata hitung

$n$  = jumlah sampel

Radius sekolah dengan class Positif

$$= \frac{\sqrt{\frac{(0-1.061)^2+(2-1.061)^2+(2-1.061)^2+\dots+(0-1.061)^2}{163}}}{\frac{(0-1.061)^2+(2-1.061)^2+(2-1.061)^2+\dots+(0-1.061)^2}{163}}$$

$$= 1.081$$

Radius sekolah class Negatif

$$= \frac{\sqrt{\frac{(2-1.079)^2+(3-1.079)^2+(0-1.079)^2+\dots+(3-1.079)^2}{38}}}{\frac{(2-1.079)^2+(3-1.079)^2+(0-1.079)^2+\dots+(3-1.079)^2}{38}}$$

$$= 1.148$$

Usia masuk class Positif

$$= \frac{\sqrt{\frac{(19-18.436)^2+(19-18.436)^2+(20-18.436)^2+\dots+(19-18.436)^2}{163}}}{\frac{(19-18.436)^2+(19-18.436)^2+(20-18.436)^2+\dots+(19-18.436)^2}{163}}$$

$$= 1.343$$

Usia masuk class Negatif

$$= \frac{\sqrt{\frac{(18-18.921)^2+(20-18.921)^2+(18-18.921)^2+\dots+(20-18.921)^2}{38}}}{\frac{(18-18.921)^2+(20-18.921)^2+(18-18.921)^2+\dots+(20-18.921)^2}{38}}$$

$$= 2.813$$

IPS 1 class Positif

$$= \frac{\sqrt{\frac{(3.00-3.329)^2+(2.52-3.329)^2+(2.96-3.329)^2+\dots+(4.00-3.329)^2}{163}}}{\frac{(3.00-3.329)^2+(2.52-3.329)^2+(2.96-3.329)^2+\dots+(4.00-3.329)^2}{163}}$$

$$= 0.410$$

IPS 1 class Negatif

$$= \frac{\sqrt{\frac{(2.61-3.141)^2+(2.43-3.141)^2+(2.91-3.141)^2+\dots+(3.26-3.141)^2}{38}}}{\frac{(2.61-3.141)^2+(2.43-3.141)^2+(2.91-3.141)^2+\dots+(3.26-3.141)^2}{38}}$$

$$= 0.428$$

IPS 2 class Positif

$$= \frac{\sqrt{\frac{(3.42-3.484)^2+(2.63-3.484)^2+(2.68-3.484)^2+\dots+(2.60-3.484)^2}{163}}}{\frac{(3.42-3.484)^2+(2.63-3.484)^2+(2.68-3.484)^2+\dots+(2.60-3.484)^2}{163}}$$

$$=$$

0.359

IPS 2 class Negatif

$$= \frac{\sqrt{(3.05-3.256)^2+(3.13-3.256)^2+(3.05-3.256)^2+\dots+(3.65-3.256)^2}}{3R}$$

$$= \frac{\sqrt{(3.05-3.256)^2+(3.13-3.256)^2+(3.05-3.256)^2+\dots+(3.65-3.256)^2}}{3R}$$

$$= 0.372$$

IPS 3 class Positif

$$= \frac{\sqrt{(3.38-3.518)^2+(3.26-3.518)^2+(2.95-3.518)^2+\dots+(2.00-3.518)^2}}{163}$$

$$= \frac{\sqrt{(3.38-3.518)^2+(3.26-3.518)^2+(2.95-3.518)^2+\dots+(2.00-3.518)^2}}{163}$$

$$= 0.333$$

IPS 3 class Negatif

$$= \frac{\sqrt{(2.76-3.331)^2+(3.33-3.331)^2+(3.14-3.331)^2+\dots+(3.68-3.331)^2}}{3R}$$

$$= \frac{\sqrt{(2.76-3.331)^2+(3.33-3.331)^2+(3.14-3.331)^2+\dots+(3.68-3.331)^2}}{3R}$$

$$= 0.338$$

Tahap ketiga yaitu menghitung probabilitas dengan menggunakan pendekatan gaussian yang dapat dilihat pada formula 2.

Radius sekolah dengan class Positif

$$= \frac{1}{\sqrt{2\pi(1.081)}} e^{-\frac{(2-1.061)^2}{2(1.081)^2}} \frac{1}{\sqrt{2\pi(1.081)}} e^{-\frac{(2-1.061)^2}{2(1.081)^2}}$$

$$= 0.559696$$

Radius sekolah dengan class Negatif

$$= \frac{1}{\sqrt{2\pi(1.148)}} e^{-\frac{(2-1.079)^2}{2(1.148)^2}} \frac{1}{\sqrt{2\pi(1.148)}} e^{-\frac{(2-1.079)^2}{2(1.148)^2}}$$

$$= 0.513820$$

Usia masuk dengan class Positif

$$= \frac{1}{\sqrt{2\pi(1.343)}} e^{-\frac{(18-18.436)^2}{2(1.343)^2}} \frac{1}{\sqrt{2\pi(1.343)}} e^{-\frac{(18-18.436)^2}{2(1.343)^2}}$$

$$= 0.362968$$

Usia masuk dengan class Negatif

$$= \frac{1}{\sqrt{2\pi(2.813)}} e^{-\frac{(18-18.921)^2}{2(2.813)^2}} \frac{1}{\sqrt{2\pi(2.813)}} e^{-\frac{(18-18.921)^2}{2(2.813)^2}}$$

$$= 0.251022$$

Ips 1 dengan class Positif

$$= \frac{1}{\sqrt{2\pi(0.410)}} e^{-\frac{(3.73-3.329)^2}{2(0.410)^2}} \frac{1}{\sqrt{2\pi(0.410)}} e^{-\frac{(3.73-3.329)^2}{2(0.410)^2}}$$

$$= 1.005418$$

Ips 1 dengan class Negatif

$$= \frac{1}{\sqrt{2\pi(0.428)}} e^{-\frac{(3.73-3.141)^2}{2(0.428)^2}} \frac{1}{\sqrt{2\pi(0.428)}} e^{-\frac{(3.73-3.141)^2}{2(0.428)^2}}$$

$$= 1.572317$$

Ips 2 dengan class Positif

$$= \frac{1}{\sqrt{2\pi(0.359)}} e^{-\frac{(3.71-3.84)^2}{2(0.359)^2}} \frac{1}{\sqrt{2\pi(0.359)}} e^{-\frac{(3.71-3.84)^2}{2(0.359)^2}} = 0.811949$$

Ips 2 dengan class Negatif

$$= \frac{1}{\sqrt{2\pi(0.372)}} e^{-\frac{(3.71-3.256)^2}{2(0.372)^2}} \frac{1}{\sqrt{2\pi(0.372)}} e^{-\frac{(3.71-3.256)^2}{2(0.372)^2}} = 1.3777777$$

Ips 3 dengan class Positif

$$= \frac{1}{\sqrt{2\pi(0.333)}} e^{-\frac{(3.63-3.518)^2}{2(0.333)^2}} \frac{1}{\sqrt{2\pi(0.333)}} e^{-\frac{(3.63-3.518)^2}{2(0.333)^2}} = 0.7317491$$

Ips 3 dengan class Negatif

$$= \frac{1}{\sqrt{2\pi(0.338)}} e^{-\frac{(3.63-3.331)^2}{2(0.338)^2}} \frac{1}{\sqrt{2\pi(0.338)}} e^{-\frac{(3.63-3.331)^2}{2(0.338)^2}} = 1.0150540$$

Perkalian untuk setiap kelas

$$(P|\text{Positif})=P(\text{Kategori\_sekolah}=\text{Swasta}|\text{Positif}) \cdot P(\text{Radius\_sekolah}|\text{Positif}) \cdot P(\text{Usia\_masuk}|\text{Positif}) \cdot P(\text{Ips\_1}|\text{Positif}) \cdot P(\text{Ips\_2}|\text{Positif}) \cdot P(\text{Ips\_3}|\text{Positif}) \cdot P(Y=\text{Positif})$$

$$= 0.5644171 \cdot 0.57812 \cdot 0.362968 \cdot 1.005418 \cdot 0.811949 \cdot 0.7317491 \cdot 0.81095$$

$$= 0.05554611$$

$$(P|\text{Negatif})=P(\text{Radius\_sekolah}|\text{Negatif}) \cdot P(\text{Usia\_masuk}|\text{Negatif}) \cdot P(\text{Kategori\_sekolah}=\text{Swasta}|\text{Negatif}) \cdot P(\text{Ips\_1}|\text{Negatif}) \cdot P(\text{Ips\_2}|\text{Negatif}) \cdot P(\text{Ips\_3}|\text{Negatif}) \cdot P(Y=\text{Negatif})$$

$$= 0.513820 \cdot 0.5789473 \cdot 0.251022 \cdot 1.572317 \cdot 1.3777777 \cdot 1.0150540 \cdot 0.18905$$

$$= 0.03104187$$

Tahap akhir adalah membandingkan hasil class Positif dan Negatif. Dari hasil diatas, terlihat bahwa nilai probabilitas tertinggi ada pada kelas (P | Positif) sehingga dapat disimpulkan bahwa mahasiswa tersebut diprediksi memiliki masa waktu tunggu “**Positif**” yang artinya mahasiswa tersebut di prediksi akan mendapat pekerjaan setelah dinyatakan lulus.

Setelah model prediksi berupa aturan didapatkan, maka dapat aturan-aturan tersebut diimplementasikan menjadi aplikasi prediksi waktu tunggu mahasiswa untuk mendapatkan pekerjaan. Tampilan layar terlihat pada Gambar 3 hingga Gambar 6.

Gambar 3. Proses Pelatihan Data

Pada halaman data training digunakan untuk mengimpor data training. Setelah proses import data, maka semua data di proses untuk perhitungan naïve bayes, perhitungan tersebut dapat dilihat pada Gambar 4.

Gambar 4. Proses Pelatihan Data

Pada halaman layar selanjutnya dilakukan proses pengujian. Pengujian dilakukan terhadap 22 data latih yang merupakan 10% dari seluruh data terkumpul untuk penelitian ini. Hasil pengujian terlihat pada Gambar 5.

Gambar 5. Hasil Pengujian Data

Pada bagian ini data testing dihitung menggunakan perhitungan Naïve Bayes berdasarkan aturan dari hasil pemodelan menggunakan data training. Hasil pengujian terlihat pada Gambar 6.

Gambar 6. Parameter Hasil Pengujian

Setelah dilakukan pengolahan data training maka diperoleh akurasi pada data tersebut. Akurasi dihitung dengan menggunakan *Confusion matrix*. Berikut ini dijabarkan confusion matrix pada Tabel 3.

Tabel 3. Nilai *Confusion Matrix*

Klasifikasi	Nilai sebenarnya	
	Positif	Negatif
Positif	TP = 19	FP = 0
Negatif	FN = 2	TN = 1

Berdasarkan hasil confusion matrix terlihat pada class Positif terdapat 19 mahasiswa yang diprediksi tepat sebagai class Positif, sementara ditemukan 2 mahasiswa pada class Negatif diprediksi tidak tepat sebagai class Negatif, serta ditemukan 1 mahasiswa yang diprediksikan pada class Negatif tepat sebagai class Negatif.

$$\begin{aligned}
 \text{Akurasi} &= \frac{TP+TN}{TP+TN+FP+FN} = \frac{TP+TN}{TP+TN+FP+FN} \\
 &= \frac{19+1}{19+1+0+2} = \frac{20}{22} = 0,9090 = 90,90\%
 \end{aligned}$$

$$\begin{aligned}
 \text{Presisi} &= \frac{TP}{FP+TP} = \frac{TP}{FP+TP} \\
 &= \frac{19}{0+19} = \frac{19}{19} = 1 = 100\%
 \end{aligned}$$

$$\begin{aligned}
 \text{Recall} &= \frac{TP}{FN+TP} = \frac{TP}{FN+TP} \\
 &= \frac{19}{2+19} = \frac{19}{21} = 0,90476 = 90,48\%
 \end{aligned}$$

$$\begin{aligned}
 \text{Recall} &= \frac{TP}{FN+TP} = \frac{TP}{FN+TP} \\
 &= \frac{19}{2+19} = \frac{19}{21} = 0,90476 = 90,48\%
 \end{aligned}$$

$$\begin{aligned}
 \text{Recall} &= \frac{TP}{FN+TP} = \frac{TP}{FN+TP} \\
 &= \frac{19}{2+19} = \frac{19}{21} = 0,90476 = 90,48\%
 \end{aligned}$$

$$\begin{aligned}
 \text{Recall} &= \frac{TP}{FN+TP} = \frac{TP}{FN+TP} \\
 &= \frac{19}{2+19} = \frac{19}{21} = 0,90476 = 90,48\%
 \end{aligned}$$

$$\begin{aligned}
 \text{Recall} &= \frac{TP}{FN+TP} = \frac{TP}{FN+TP} \\
 &= \frac{19}{2+19} = \frac{19}{21} = 0,90476 = 90,48\%
 \end{aligned}$$

Dari perhitungan di atas, dapat dilihat bahwa tingkat akurasi yang diperoleh dari model prediksi ini adalah sebesar 90.90%. Nilai *presisi* sebesar 100%, dan nilai *recall* sebesar 90.48%.

## V. KESIMPULAN

Berdasarkan hasil penelitian maka dapat diambil kesimpulan bahwa metode Naïve Bayes dapat digunakan untuk memprediksi waktu tunggu mahasiswa Akademi Sekretari Budi Luhur dalam mendapatkan pekerjaan dengan menggunakan 199 data training dan 22 data testing, yang terdiri dari 1 atribut kategorial dan 5 atribut numerik

Dari proses pengujian, diperoleh confusion matrix dengan nilai akurasi sebesar 90.90%, nilai *presisi* sebesar 100%, dan nilai *recall* sebesar 90.48%. Kesimpulan akhir studi ini menunjukkan bahwa lebih banyak mahasiswa yang masa tenggunya diprediksi dengan hasil Positif atau mendapat pekerjaan setelah lulus.

## REFERENSI

- [1] Rizal, S., dan Lutfi, M., "Penerapan algoritma naive bayes untuk penerimaan siswa baru di SMK AL-AMIEN Wonorejo," *Jurnal Explore IT*, vol.10, no.1, pp.14-21, 2018.
- [2] Rumini dan Norhikmah, "Prediksi kegagalan siswa dalam data mining menggunakan metode naive bayes," *Jurnal mantik penusa*, vol.3, no.3, p p.42-46, 2019.
- [3] Adi, S., "Implementasi Algoritma Naive Bayes Classifier Untuk Klasifikasi Penerima Beasiswa PPA Di Universitas Amikom Yogyakarta," *Jurnal Mantik Penusa*, vol.22, no.1, pp.11-16, 2018.
- [4] Rifai, M., Jatnika dan Valentino, B., "Penerapan Algoritma Naïve Bayes Pada Sistem Prediksi Tingkat Kelulusan Peserta Sertifikasi Microsoft Office Specialist (MOS)," *Petir*, vol.12, pp.131-144, 2019.
- [5] Ryan, P.D., Supianto, A.A., dan Setiawan, N.Y., "Prediksi Kelulusan Mahasiswa Berdasarkan Kinerja Akademik Menggunakan Pendekatan Data Mining Pada Program Studi Sistem Informasi Fakultas Ilmu Komputer Universitas Brawijaya," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer* 2196, vol.3, no.3, pp.2194-2200, 2019.
- [6] Saleh, A., "Implementasi Metode Klasifikasi Naive Bayes Dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga," *Creative Information Technology Journal*, vol.2, no.3, pp.207-217, 2015