

ANALISIS PREDIKSI KELULUSAN MAHASISWA FAKULTAS ILMU KOMPUTER UNIVERSITAS NURDIN HAMZAH MENGUNAKAN METODE NAIVE BAYES

Rike Limia Budiarti¹, Novhirtamely Kahar², Rismawati³

^{1,2&3}Program Studi Teknik Informatika, Universitas Nurdin Hamzah, Jambi

Email: ¹rikelimia@gmail.com, ²novmely@gmail.com, ³rismawati.jambi20@gmail.com

Abstract - Universities have an obligation to produce competent graduates. This can be measured from the graduation rate of students. The purpose of the analysis carried out is to predict graduation on time and find out indicators of problems that cause graduation not to be on time. Many factors influence students' late graduation, such as the student's marital status, the status of the student working or not working, the level of student understanding of the course material which can be seen from the student's GPA. Predicting student graduation is very important because this research can make a significant contribution both academically. Understand patterns or factors that influence student success. Provides insights for designing more effective educational interventions, such as academic tutoring. Based on the existing problems, it is necessary to have a system to predict student graduation rates based on existing variables. Attributes used are Name, Name, Gender, School of Origin, Study Program, Year of College Entry, Year of Graduation, Student Status, Marital Status, Age, IPS, GPA and Parent's Occupation. The method used is the Naive Bayes method. Comparison results were carried out using Excel and Rapidminer tools. The resulting accuracy level is 97,96% from 330 training data and 98 test data

Keywords: Analysis, Data Mining, Naive Bayes, Prediction Graduation, Rappidminer

I. PENDAHULUAN

1.1. Latar Belakang

Universitas Nurdin Hamzah merupakan salah satu perguruan tinggi ternama di provinsi Jambi dibawah Naungan Yayasan Dewi Nurdin Hamzah, yang beralamat di Jalan Kolonel Abunjani, RT 25 Selamat Danau Teluk, Kecamatan Telanaipura Kota Jambi, Jambi 36124. Pada Fakultas Ilmu Komputer memiliki beberapa program studi diantaranya, S1 Sistem Informasi, S1 Teknik Informatika, S1 Teknologi Informasi.

Keberhasilan mahasiswa dalam suatu proses belajar merupakan suatu nilai yang bisa diukur dari parameter akademiknya. Apabila nilai mahasiswa baik maka dapat dikatakan mampu menyelesaikan proses belajar mengajar, yang artinya ada suatu kesuksesan dalam transfer pengetahuan dan skil. Hal ini menggambarkan adanya keberhasilan dalam hal pengetahuan, keahlian dan kemampuan mahasiswa.

Pada proses prediksi kelulusan mahasiswa Filkom Universitas Nurdin Hamzah banyak mahasiswa yang mengalami kendala dalam proses kelulusan, dari permasalahan mahasiswa tersebut tidak diketahui apa kendala yang menyebabkan mahasiswa lalai dalam menentukan target kelulusan mahasiswa tepat waktu. Dari kecenderungan mahasiswa dapat dilihat beberapa kendala yang dapat mempengaruhi kelulusan mahasiswa tersebut seperti faktor ekonomi, minat belajar, status mahasiswa, status perkawinan dan masih banyak yang lainnya.

Dalam hal ini sampel data yang diambil dari beberapa sumber Perguruan Tinggi baik itu Universitas, Institut, ataupun Sekolah Tinggi sebagai

penentu dalam hal prediksi tingkat kelulusan mahasiswa tepat waktu, normalnya mahasiswa yang dapat lulus tepat waktu adalah mahasiswa yang masa pendidikannya mampu ditempuh selama empat tahun masa studi, jika lewat dari itu maka mahasiswa tersebut digolongkan mahasiswa tidak tepat waktu.

Berdasarkan penelitian sebelumnya, sudah ada penelitian yang mengangkat tentang prediksi kelulusan mahasiswa. Menurut Hartatik (2020), dalam penelitiannya yang berjudul Optimasi Model Prediksi Kelulusan Mahasiswa Menggunakan Algoritma *Naive Bayes*. Penelitian ini menggunakan variabel-variabel parameter akademis dan parameter sosial seperti, IPK, jenis kelamin dan status tinggal. Dengan hasil akurasi untuk metode *naive bayes* sebesar 75% dan prediksi dengan parameter sosial sebesar 85% dengan selisih akurasi 10%.

Menurut Etriyanti et al (2020), dalam penelitiannya yang berjudul Implementasi Data Mining Menggunakan Algoritma *Naive Bayes Classifier* Dan C4.5 Untuk Memprediksi Kelulusan Mahasiswa. Variabel dalam penelitian ini menggunakan dataset jenis kelamin, asal sekolah, status sekolah dan nilai IP 1-4. Data yang digunakan berjumlah 162 mahasiswa dengan perolehan prediksi tingkat akurasi 79,08% sedangkan metode *naive bayes classifier* hanya 78,46%.

Menurut Qisthiano et al (2021), dalam penelitiannya yang berjudul Pengembangan Model Untuk Prediksi Tingkat Kelulusan Mahasiswa Tepat Waktu Dengan Metode Naive Bayes. Atribut yang digunakan adalah Jurusan, Perguruan Tinggi, Jenis Kelas, Nilai IP Sementara dari semester 1 sampai dengan 4, tahun kelulusan, dan angkatan perguruan tinggi, menggunakan bahasa pemrograman Python 3

dan tools Jupyter Notebook untuk mengolah dataset yang telah disiapkan. Selanjutnya distribusi dataset dibagi dengan 70% untuk data latih dan 30% untuk data uji. Hasil prediksi akurasi yang diperoleh dari bahasa pemrograman Python 3 dan algoritma Naive Bayes adalah sebesar 0,8103.

Dari penelitian tersebut yang membedakan dalam penelitian saya adalah dataset yang digunakan yaitu data detail mahasiswa mulai dari nama, nim, program studi, asal sekolah, jenis kelamin, usia, tahun masuk kuliah, tahun kelulusan mahasiswa, nilai KHS, IPK, status mahasiswa, status perkawinan dan pekerjaan orang tua. Dengan mengimplementasikan menggunakan *excel dan tools rappidminer*. Output yang dihasilkan berupa Hasil data prediksi kelulusan mahasiswa, *statistics, visualizations, descriptions, performance, simple chart dan distribution table*.

Berdasarkan latar belakang tersebut maka peneliti melakukan penelitian yang berjudul: **“Analisis Prediksi Kelulusan Mahasiswa Filkom UNH Menggunakan Metode Naive Bayes”**

1.2. Rumusan Masalah

Berdasarkan latar belakang masalah diatas, maka peneliti dapat merumuskan masalah dalam penelitian ini adalah bagaimana Menganalisis Prediksi Kelulusan Mahasiswa Filkom Universitas Nurdin Hamzah Dengan Metode *Naive Bayes*.

1.3. Tujuan Penelitian

Adapun tujuan dari penelitian ini adalah

1. Untuk mengetahui hasil dari penerapan metode *naive bayes* dalam prediksi kelulusan mahasiswa.
2. Untuk mengetahui hasil prediksi Kelulusan Mahasiswa Filkom Universitas Nurdin Hamzah pada mahasiswa semester 6 angkatan tahun 2021.
3. Untuk memperkirakan kelulusan tepat waktu dengan masa pendidikan yang ditempuh selama empat tahun.
4. Mengidentifikasi faktor-faktor yang mempengaruhi keberhasilan akademik mahasiswa dalam perkiraan lulus tepat waktu.

II. TINJAUAN PUSTAKA

2.1 Data Mining

Menurut Arhami dan Nasir (2020, hh.1-2), Data mining juga merupakan proses penggalian informasi dan pola yang bermanfaat dari data yang sangat besar. Data mining mencakup pengumpulan data, ekstraksi data, analisis data dan statistik data. Sedangkan menurut Suntoro (2019, hh.2-5), Data mining merupakan suatu proses ekstraksi suatu data

yang sebelumnya tidak diketahui, bersifat *implisit* dan dianggap tidak berguna yang kemudian diolah menjadi sebuah informasi dan pengetahuan dari data yang jumlahnya sangat besar.

2.2 Klasifikasi

Menurut Rahayu et al (2024, h.79), Klasifikasi dalam data mining adalah proses pengelompokan atau kategorisasi data ke dalam kelas atau kelompok berdasarkan karakteristik atau atribut tertentu. Menurut Setiawan et al (2023, h.114), Klasifikasi pada teknik data mining merupakan sebuah proses untuk menemukan kesamaan karakteristik dalam suatu kelas atau kelompok.

2.3 Naive Bayes

Pada penelitian ini menggunakan metode naive bayes. Mengapa tidak menggunakan metode Decision tree atau metode SVM?, karena naive bayes memiliki algoritma yang sederhana dan cepat hanya membutuhkan estimasi probabilitas data. Sangat efisien untuk data besar, dibandingkan dengan Decision Tree yang bisa menjadi lambat saat memecah banyak node, atau SVM yang memiliki waktu komputasi lebih tinggi saat menangani data besar atau dengan dimensi tinggi. Pada dasarnya metode naive bayes lebih cocok untuk berbagai tugas klasifikasi.

Sidiq, Fathonah dan Riza (2020, hh.11-12), *Algoritma Naive Bayes Classifier* merupakan sebuah pengklasifikasian probabilistik sederhana yang menghitung sekumpulan probabilitas dengan menunjukkan jumlah frekuensi dan kombinasi dari dataset yang diberikan. Menurut Hastuti (2023, h.1), *Naive bayes classifier probabilistik* sederhana menghitung kompleksitas dengan sejumlah data kombinasi frekuensi dan nilai dataset yang diberikan.

Adapun persamaan dari teorema bayes dapat dilihat dibawah ini :

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

- Keterangan:
- X :Data *class* yang belum diketahui.
 - H :Hipotesa data menggunakan suatu *class* khusus.
 - P(H|X) :Nilai probabilitas pada hipotesa H berdasarkan kondisi X.
 - P(H) :Nilai probabilitas pada hipotesa H
 - P(X|H) :Nilai probabilitas X berdasarkan dengan kondisi H.
 - P(X) : nilai probabilitas X.

Adapun tahapan dari Teorema *Naive Bayes* adalah sebagai berikut :

1. Baca data training.
2. Hitung jumlah dan probabilitas, jika data numerik maka :

3. Cari nilai mean dan standar deviasi masing-masing parameter yang merupakan data numerik.
4. Cari nilai probabilistik dengan cara menghitung jumlah data yang sesuai dari kategori yang sama dibagi dengan jumlah data pada kategori tersebut.
5. Mendapatkan nilai pada tabel *mean, standar deviasi* dan probabilitas.
6. Solusi yang dihasilkan.

2.4 Analisis

Menurut Taufiq (2023, hh.4-7), Analisis adalah membaca teks yang melokasikan berbagai tanda dan menempatkan tanda-tanda tersebut dalam interaksi yang dinamis dan pesan pesan yang disampaikan. Analisis juga menguraikan suatu (obyek) kedalam komponen-komponen yang diidentifikasi dan dievaluasi terhadap permasalahan, kebutuhan, perintah (keinginan), lalu di carikan berbagai alternatif pemecahan masalah yang dapat memberikan hasil atau kontribusi yang baik. Menurut Pratama et al (2021, h.1), Analisis didefinisikan sebagai metode penelitian yang objektif, sistematis dan bersifat kuantitatif terhadap suatu isi pesan yang manifest. Berdasarkan definisi ini jelas bahwa analisis memiliki karakteristik utama yakni objektif, sistematis dan bersifat kuantitatif.

2.5 Prediksi

Menurut Huda, Awangga dan Fathonah (2020, h.8), Prediksi digunakan untuk memberikan pandangan yang jelas tentang masa depan dan menghubungkan data besar untuk membuat pekerjaan lebih mudah. Prediksi memberikan wawasan dari masa lalu untuk menemukan tren baru dan memprediksi hasil akhir. Menurut Hengky dan Rusman (2022, h.4), Prediksi merupakan proses memperkirakan secara sistematis tentang sesuatu yang paling mungkin terjadi dimasa depan berdasarkan informasi masalalu dan sekarang yang dimiliki. Prediksi tidak harus memberikan jawaban secara pasti kejadian yang akan terjadi namun berusaha untuk memberikan jawaban sedekat mungkin dengan yang akan terjadi.

2.6 Kelulusan

Menurut Zebua et al (2023, h.63), Kelulusan adalah sebagai pernyataan sekolah sebagai suatu lembaga tentang telah diselesaikannya program pendidikan yang diikuti oleh siswa. Setelah seorang siswa selesai mengikuti program pendidikan di suatu sekolah dan berhasil lulus, maka kedepannya diberikan surat keterangan dan sertifikat yang umumnya disebut ijazah atau surat tanda tamat belajar (STTB). Menurut Rusdiana dan Kodir (2022, h.273), Proses kelulusan adalah kegiatan paling

akhir dari manajemen kesiswaan. Kelulusan adalah pernyataan dari sekolah sebagai suatu lembaga tentang telah diselesaikannya program pendidikan yang harus diikuti oleh siswa.

2.7 Mahasiswa

Menurut Fauzi (2021, h.22), Definisi mahasiswa secara sederhana dapat didefinisikan sebagai kelompok masyarakat yang dapat mengenyam pendidikan formal tingkat tinggi pada sebuah instansi perguruan tinggi baik swasta maupun negeri. Menurut Sutrisman (2019, h.115), Mahasiswa adalah orang yang belajar di perguruan tinggi, baik di universitas atau institut. Sebagaimana tercantum dalam undang-undang republik indonesia nomor 12 tahun 2012 tentang pendidikan tinggo pasal 1 ayat (1) yang berbunyi : “Mahasiswa adalah peserta didik pada jenjang pendidikan tinggi.” Terdaftar sebagai mahasiswa disebuah perguruan tinggi syarat administratif menjadi mahasiswa.

III. HASIL DAN PEMBAHASAN

Data *training* merupakan *dataset* pelatihan yang disediakan untuk jadi bahan pembelajaran agar dapat menggeneralisasikan pola data sehingga dapat digunakan untuk prediksi data baru. Data *training* yang digunakan dalam penelitian ini adalah mahasiswa alumni 2016, 2017 dan 2018 yang terdiri dari 330 data.

Tabel 1. Dataset Pelengkap

| No | Variabel | Keterangan |
|----|-----------------|------------------------|
| 1. | Nama | Nama mahasiswa |
| 2. | Nim | Nomor induk mahasiswa |
| 3. | Tahun masuk | 2016, 2017, 2018, 2021 |
| 4. | Tahun kelulusan | 2020, 2021, 2022, 2025 |

Tabel 2. Dataset Yang Digunakan

| No | Variabel | Keterangan |
|----|---------------|---|
| 1. | Asal sekolah | SMA dan SMK |
| 2. | Jenis kelamin | Laki-laki dan perempuan |
| 3. | Program studi | Teknik informatika dan sistem informasi |
| 4. | Usia | Range usia: 1 (20-25 tahun), 2 (26-31 tahun), 3(32-37 tahun), 4(38-43 tahun) |

Pada tabel 1 dan tabel 2 merupakan dataset yang digunakan pada analisis prediksi yang dilakukan pada proses perhitungan manual.

Tabel 3. Data Training

| No (1) | Nama (2) | Nim (3) | Th Masuk (4) | Th Lulus (5) |
|--------|-------------|---------|--------------|--------------|
| 1 | Wira H. | 1602001 | 2016 | 2020 |
| 2 | Anuwar | 1602002 | 2016 | 2020 |
| 3 | Siti M. | 1602003 | 2016 | 2020 |
| 4 | Refliansyah | 1602004 | 2016 | 2020 |

| | | | | |
|-----|-----------|----------|------|------|
| 5 | Aftia D. | 1602005 | 2016 | 2020 |
| ... | | ... | ... | ... |
| 330 | Yulisa F. | 11801081 | 2018 | 2022 |

Tabel 3. Data *Training* (Lanjutan)

| Asal Sekolah (6) | Jenis Kelamin (7) | Program Studi (8) | Usia (9) | Range Usia (10) |
|------------------|-------------------|-------------------|----------|-----------------|
| Smk | L | TI | 25 | 1 |
| Sma | L | TI | 27 | 2 |
| Sma | P | TI | 27 | 2 |
| Smk | L | TI | 26 | 2 |
| Sma | P | TI | 28 | 2 |
| ... | ... | ... | ... | ... |
| Smk | P | SI | 24 | 1 |

Tabel 3. Data *Training* (Lanjutan)

| Nilai Khs (11) | Range Khs (12) | Nilai Ipk (13) | Range Ipk (14) |
|----------------|----------------|----------------|----------------|
| 1,00 | 1 | 2,00 | 1 |
| 2,77 | 2 | 2,74 | 2 |
| 4,00 | 3 | 3,19 | 3 |
| 3,00 | 3 | 3,31 | 3 |
| 4,00 | 3 | 2,00 | 1 |
| ... | ... | ... | ... |
| 4,00 | 3 | 2,92 | 2 |

Tabel 3. Data *Training* (Lanjutan)

| Sts Mhs (15) | Sts Perkawinan (16) | Pekerjaan Orang Tua (17) | Status Kelulusan (18) |
|--------------|---------------------|--------------------------|-----------------------|
| Mhs | Kawin | PNS | Tidak Tepat Waktu |
| Bekerja | Blm Kawin | Petani | Tepat Waktu |
| Mhs | Blm Kawin | Pedagang | Tepat Waktu |
| Mhs | Blm Kawin | Wiraswasta | Tepat Waktu |
| Mhs | Blm Kawin | Guru | Tepat Waktu |
| ... | ... | ... | ... |
| Mhs | Blm Kawin | Mrt | Tepat Waktu |

Tabel 3. Merupakan tabel data *training* untuk dijadikan sampel pengolahan yang nantinya akan dijadikan acuan untuk melakukan testing pada data pediksi. Setelah data *training* di inputkan, maka langkah selanjutnya adalah dengan memasukan rumus *Naive Bayes* untuk menghitung hasil dari probabilitas yang nantinya akan digunakan untuk menghitung data *testing*.

Tabel 4. Data *Testing*

| No (1) | Nama (2) | Nim (3) | Th Masuk (4) | Tahun Kelulusan (5) |
|--------|------------|----------|--------------|---------------------|
| 1 | Lydia M. | 21111001 | 2021 | 2025 |
| 2 | Intan C.A. | 21111002 | 2021 | 2025 |

Tabel 4. Data *Testing* (Lanjutan)

| | | | | |
|-----|------------|----------|------|------|
| 3 | Karina O. | 21111003 | 2021 | 2025 |
| 4 | Riya A.F | 21111004 | 2021 | 2025 |
| 5 | Desvia E.S | 21111005 | 2021 | 2025 |
| ... | ... | ... | ... | ... |
| 98 | Fajri A. | 21111062 | 2021 | 2025 |

Tabel 4. Data *Testing* (Lanjutan)

| Asal Sekolah (6) | Jenis Kelamin (7) | Program Studi (8) | Usia (7) | Range Usia (10) |
|------------------|-------------------|-------------------|----------|-----------------|
| Sma | P | SI | 21 | 1 |
| Sma | P | SI | 21 | 1 |
| Sma | P | SI | 29 | 2 |
| Sma | P | SI | 21 | 1 |
| Sma | P | SI | 22 | 1 |
| ... | ... | ... | ... | ... |
| Smk | L | TI | 25 | 1 |

Tabel 4. Data *Testing* (Lanjutan)

| Nilai Khs (11) | Range Khs (12) | Nilai Ipk (13) | Range Ipk (14) |
|----------------|----------------|----------------|----------------|
| 3,81 | 3 | 3,71 | 3 |
| 4,00 | 3 | 3,76 | 3 |
| 2,62 | 2 | 2,75 | 2 |
| 3,67 | 3 | 3,77 | 3 |
| 3,24 | 3 | 3,65 | 3 |
| ... | ... | ... | ... |
| 4,00 | 3 | 4,00 | 3 |

Tabel 4. Data *Testing* (Lanjutan)

| Sts Mhs (15) | Sts Perkawinan (15) | Pekerjaan Orang Tua (17) |
|--------------|---------------------|--------------------------|
| Mhs | Blm Kawin | PNS |
| Mhs | Blm Kawin | Petani |
| Bekerja | Kawin | Pedagang |
| Mhs | Blm Kawin | Wiraswasta |
| Mhs | Blm Kawin | Guru |
| ... | ... | ... |
| Bekerja | Blm Kawin | Irt |

Tabel 4. Merupakan data *testing* yang sudah didapat dan telah di normalisasikan sesuai dengan

nilai pada setiap atribut. Setelah melakukan *training* dari 330 data dan telah mendapatkan nilai probabilitas atribut. Maka dilakukan dengan menguji data atau *testing* data baru pada mahasiswa semester 6 angkatan tahun 2021.

3.1 Proses Penyelesaian Naive Bayes

Pada tahap ini, dilakukan proses perhitungan manual pada klasifikasi untuk mencari nilai label pada data dengan menggunakan persamaan *naive bayes*.

1. Menghitung Jumlah Kelas

Langkah pertama yaitu menghitung jumlah kelas kemudian dibagi dengan total data untuk mendapatkan nilai probabilitas kelas. Berikut cara menghitung probabilitas kelas:

$$P(\text{Tepat Waktu/Jumlah Data}) = 298/330 = 0,903$$

$$P(\text{Tidak Tepat Waktu/Jumlah Data}) = 32/330 = 0,097$$

Tabel 5. Tabel Probabilitas Kelas/Label

| Label | Jumlah | Probabilitas |
|------------------|--------|--------------|
| Tepat Waktu | 298 | 0.903 |
| Tidak Tepat Wktu | 32 | 0.097 |

2. Menghitung Probabilitas Masing-Masing Atribut

Untuk mencari nilai probabilitas atribut, dengan menjumlahkan nilai pada tiap kelas dan juga pada tiap atribut pada data *training*. Setelah ditotalkan sesuai atribut yang didapat lalu dibagi dengan jumlah masing-masing kelas.

Berikut cara untuk menghitung nilai probabilitas atribut :

Atribut jenis kelamin dengan kelas tepat waktu dan tidak tepat waktu:

$$P(\text{Tepat Waktu|SMA}) = 114/298 = 0,42$$

$$P(\text{Tepat Waktu|SMK}) = 174/298 = 0,56$$

$$P(\text{Tidak Tepat Waktu|SMA}) = 11/32 = 0,34$$

$$P(\text{Tidak Tepat Waktu|SMK}) = 21/32 = 0,66$$

Tabel 6. Probabilitas Masing Atribut

| | | Probabilitas Atribut | |
|-------------------|-----|----------------------|----------------------|
| | | P(Tepat Waktu) | P(Tidak Tepat Waktu) |
| Asal Sekolah | Sma | 0,42 | 0,34 |
| | Smk | 0,58 | 0,66 |
| Jenis Kelamin | L | 0,56 | 0,88 |
| | P | 0,44 | 0,13 |
| Program Studi | SI | 0,53 | 0,38 |
| | TI | 0,47 | 0,63 |
| Usia (Range Usia) | 1 | 0,60 | 0,69 |
| | 2 | 0,40 | 0,31 |
| | 3 | 0,00 | 0,00 |
| | 4 | 0,00 | 0,00 |

Tabel 6. Probabilitas Atribut (Lanjutan)

| | | | |
|-----------|---|------|------|
| Nilai KHS | 0 | 0,00 | 0,31 |
|-----------|---|------|------|

| | | | |
|-----------------------|-----------------------|------|------|
| (Range KHS) | 1 | 0,02 | 0,56 |
| | 2 | 0,14 | 0,13 |
| | 3 | 0,84 | 0,00 |
| Nilai IPK (Range IPK) | 0 | 0,00 | 0,16 |
| | 1 | 0,02 | 0,72 |
| | 2 | 0,39 | 0,13 |
| | 3 | 0,59 | 0,00 |
| Status Mahasiswa | Mhs | 0,66 | 0,66 |
| | Bekerja | 0,34 | 0,34 |
| Status Perkawinan | Blm Kawin | 0,94 | 0,72 |
| | Kawin | 0,06 | 0,28 |
| Pekerjaan Orang Tua | Wiraswasta | 0,21 | 0,09 |
| | PNS | 0,08 | 0,09 |
| | Pedagang | 0,02 | 0,09 |
| | Kar. Swasta | 0,16 | 0,13 |
| | Petani | 0,24 | 0,16 |
| | Pensiunan | 0,02 | 0,03 |
| | Guru | 0,06 | 0,09 |
| | Kepolisian RI | 0,06 | 0,06 |
| | Karyawan Honorar | 0,02 | 0,06 |
| | Mengurus Rumah Tangga | 0,08 | 0,06 |
| | Buruh Harian Lepas | 0,04 | 0,06 |
| | Tukang Las | 0,02 | 0,06 |

3. Prediksi Data Baru/Data Training

Setelah didapatkan nilai probabilitas dari data *training* dengan rumus *naive bayes*, maka langkah selanjutnya untuk menghitung data baru adalah dengan cara mengkalikan probabilitas setiap label atau kelas pada data *training* sesuai dengan data baru maka siap dilakukan perhitungan menggunakan teori *naive bayes*. Contoh perhitungan dapat dilihat pada berikut ini:

a. Kelas Tepat Waktu:

$$= P(\text{Tepat Waktu/Asal Sekolah}) * P(\text{Tepat Waktu/Jenis Kelamin}) * P(\text{Tepat Waktu/Program Studi}) * P(\text{Tepat Waktu/Range Usia}) * P(\text{Tepat Waktu/Range KHS}) * P(\text{Tepat Waktu/Range IPK}) * P(\text{Tepat Waktu/Status Mahasiswa}) * P(\text{Tepat Waktu/Status Perkawinan}) * P(\text{Tepat Waktu/Pekerjaan Orang Tua}) * P(\text{Status Kelulusan/Tepat Waktu})$$

$$= 0,42 * 0,44 * 0,53 * 0,60 * 0,84 * 0,59 * 0,66 * 0,94 * 0,21 = 0,003363574$$

b. Kelas Tidak Tepat Waktu :

$$= P(\text{Tidak Tepat Waktu/Asal Sekolah}) * P(\text{Tidak Tepat Waktu/Jenis Kelamin}) * P(\text{Tidak Tepat Waktu/Program Studi}) * P(\text{Tidak Tepat Waktu/Range Usia}) * P(\text{Tidak Tepat Waktu/Range KHS}) * P(\text{Tidak Tepat Waktu/Range IPK}) * P(\text{Tidak Tepat Waktu/Status Mahasiswa}) * P(\text{Tidak Tepat Waktu/Status Perkawinan}) * P(\text{Tidak Tepat Waktu/Pekerjaan Orang Tua}) * P(\text{Status Kelulusan/Tidak Tepat Waktu})$$

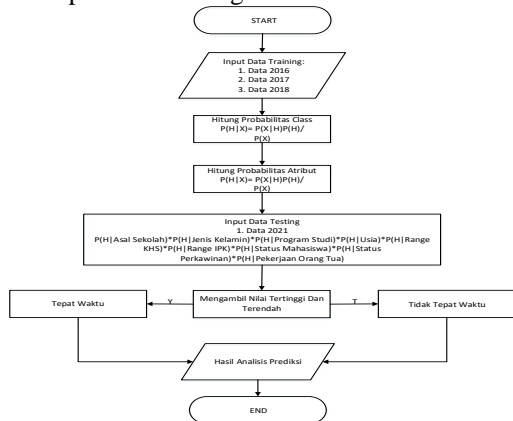
$$\begin{aligned}
 &Waktu/RangeKHS)*P(TidakTepatWaktu/RangeIPK) \\
 &*P(TidakTepatWaktu/StatusMahasiswa)*P(TidakTepatWaktu/StatusPerkawinan)*P(TidakTepatWaktu/P \\
 &ekejaanOrangTua)*P(StatusKelulusan/TidakTepatWaktu) \\
 &=0,34*0,13*0,38*0,69*0,00*0,00*0,66*0,72*0,09 \\
 &= 0
 \end{aligned}$$

Proses perhitungan diatas dilakukan hingga data ke 98. Kemudian untuk melihat hasil akhir adalah dengan cara melihat hasil kelas tepat waktu dan tidak tepat waktu apabila nilai tersebut tinggi diantara salah satunya maka itulah yang menjadi nilai prediksi kelulusan. Setelah diketahui hasil nilai tertinggi dan terendah, maka dapat di ketahui bahwa nilai tertinggi dengan jumlah **0,003363574** dan terendah dengan jumlah **0,000000000**.

Dapat disimpulkan bahwa mahasiswa dengan Nama Lydia Madini, dengan Nim 21111001, tahun masuk 2021, tahun kelulusan 2025, Asal Sekolah SMA, Berjenis Kelamin Perempuan, Program Studi Sistem Informasi, Usia 21 tahun dengan Nilai KHS 3,81 nilai IPK 3,71, Status sebagai Mahasiswa, Belum Kawin dan Pekerjaan Orang Tua Wiraswasta, diprediksi tingkat kelulusan berdasarkan perhitungan metode *naive bayes* adalah lulus tepat waktu.

3.2 Flowchart Naïve Bayes

Proses perhitungan manual menggunakan *excel* dapat dilihat dari gambar 1 dibawah ini:



Gambar 1. Flowchart Naïve Bayes

3.3 Hasil Implementasi Rapidminer

Hasil implementasi berupa data prediksi kelulusan mahasiswa tepat waktu atau tidak tepat waktu dari hasil perbandingan menggunakan *excel* dan *rapidminer*. Hasil implementasi dapat dilihat pada gambar dibawah ini:

1. Implementasi Output Data

Output data menampilkan berupa data prediksi kelulusan mahasiswa tepat waktu atau tidak tepat waktu, menampilkan berbagai kriteria yang digunakan dan menunjukkan nilai *confidence* dari hasil prediksi kelulusan mahasiswa.

Gambar 2. Tampilan Output Data

2. Implementasi Output Statistic

Output Statistic menampilkan tampilan beberapa atribut yang digunakan pada proses prediksi berupa tampilan *statistic*.

Gambar 3. Tampilan Output Statistic

3. Implementasi Output Visualizations

Hasil dari *Output Visualization* merupakan tampilan grafik dari beberapa atribut yang digunakan pada proses prediksi.



Gambar 4. Tampilan Output Visualizations

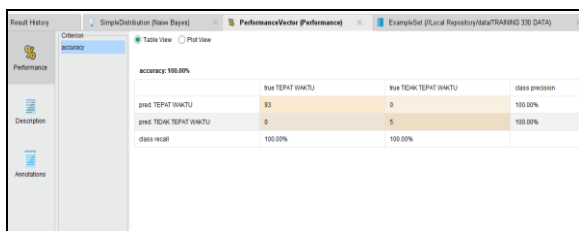
4. Implementasi Output Descriptions

Output Description merupakan penjelasan mengenai atribut yang digunakan yang terdiri dari 9 atribut dan 1 label atribut yaitu status kelulusan dengan probabilitas tidak tepat waktu 0,097 dan tepat waktu 0,903.

Gambar 5. Tampilan Output Descriptions

5. Implementasi Output Performance

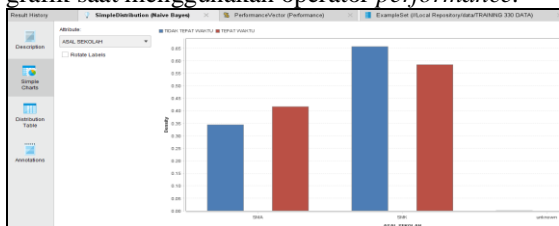
Output Performance merupakan tingkat akurasi dari hasil prediksi terhadap data-data yang digunakan. Untuk hasil tepat waktu terdapat hasil 91 dan tidak tepat waktu terdapat hasil 7. Pada *Performance* ini nilai akurasi yang di peroleh adalah 97,96%. Tingkat akurasi ini merupakan hasil prediksi dari perbandingan *excel* dan *rapidminer*.



Gambar 6. Tampilan Output Performance

6. Implemtasi Output Simple Chart

Output Simple Chart merupakan tampilan grafik saat menggunakan operator performance.



Gambar 7. Tampilan Output Simple Chart

7. Tingkat Akurasi

Untuk hasil tingkat akurasi pengujian analisis prediksi kelulusan mahasiswa menggunakan perhitungan manual excel maka diperoleh hasil akurasi yang sama menggunakan rapidminer. Mahasiswa tepat waktu 91 dan tidak tepat waktu 7 mahasiswa, untuk tingkat akurasi di dapatkan dengan nilai 97,96%.

| | true TEPAT WAKTU | true TIDAK TEPAT WAKTU | class precision |
|-------------------------|------------------|------------------------|-----------------|
| pred. TEPAT WAKTU | 91 | 2 | 97.85% |
| pred. TIDAK TEPAT WAKTU | 0 | 5 | 100.00% |
| class recall | 100.00% | 71.43% | |

Gambar 8. Tampiln Tingkat Akurasi

IV. PENUTUP

4.1 Kesimpulan

Pada panelitian ini dataset diperoleh dari Universitas Nurdin Hamzah diolah menggunakan Algoritma Naive Bayes. Maka dengan ini peneliti dapat mengambil suatu kesimpulan sebagai berikut:

1. Variabel kriteria yang digunakan sebagai atribut dataset pelengkap yang digunakan pada penelitian ini adalah nama, nim, tahun masuk kuliah, tahun kelulusan mahasiswa. Sedangkan variabel kriteria atribut dataset yang digunakan adalah asal sekolah, program studi, nilai KHS, IPK, jenis kelamin, usia, status mahasiwa, status perkawinan dan pekerjaan orang tua.
2. Untuk melakukan proses perhitungan analisis prediksi kelulusan mahasiswa dengan menggunakan metode naive bayes.
3. Perhitungan manual dilakukan dengan mengguankan excel dan sebagai bahan perbandingan tingkat akurasi yang dihasilkan adalah menggunakan tools rapidminer.

4. Output yang dihasilkan berupa keterangan lulus tepat waktu dan tidak tepat waktu.
5. Hasil prediksi kelulusan yang diperoleh dengan output mahasiswa tepat waktu adalah 91 mahasiswa dan 7 mahasiswa tidak tepat waktu dalam bentuk descriptions, statistik, grafik dan visualizations.

4.2 Saran

Adapun saran-saran tersebut antara lain sebagai berikut :

1. Dalam penelitian selanjutnya, diharapkan adanya pengembangan metode Naive Bayes dapat dikembangkan dengan baik lagi, sehingga didapatkan hasil yang lebih maksimal dan kompleks.
2. Diharapkan bisa dikombain dengan berbagai penerapan metode lainnya sehingga bisa memperkuat tingkat akurasi yang dihasilkan.
3. Memperluas sampel atau data penelitian dan mendalami setiap-setiap kriteria sehingga analisis prediksi kelulusan ini dapat melakukan klasifikasi yang lebih akurat.
4. Analisis prediksi menggunakan naive bayes ini juga perlu dikembangkan ke versi website agar sistem ini bisa berkembang mengikuti kemajuan teknologi saat ini.

DAFTAR REFERENSI

- [1] Arhami, M & Nasir, M 2020, Data Mining Algoritma dan Implementasi, Andi Offset, Surabaya
https://www.google.co.id/books/edition/Data_Mining_Algoritma_danImplementasi/AtcEAAAQBAJ?hl=id&gbpv=1&dq=definisi+d+ata+mining&printsec=frontcover
- [2] Etriyanti, E, Syamsuar, D dan Kunang, Y, N, “Implementasi Data Mining Menggunakan Algoritme Naive Bayes Classifier dan C4.5 untuk Memprediksi Kelulusan Mahasiswa”, <https://pdfs.semanticscholar.org/60f9/b246d1fc0f9e7a63900e457c17d1c36acf55.pdf>
- [3] Fauzi, H, Tempat Ketiga Bagi Mahasiswa, Cv Budi Utama, 2022, Yogyakarta
https://www.google.co.id/books/edition/Tempat_Ketiga_Bagi_Mahasiswa/TxjjeaaqbaJ?hl=Id&gbpv=1&Dq=Definisi+Mahasiswa&Pg=Pa22&Printsec=Frontcover
- [4] Hartatik, Optimasi Model Prediksi Kelulusan Mahasiswa Menggunakan Algoritma Naive Bayes, Indonesia Journal Of Applied Informatics, Vol. 5 No.1 Tahun 2020
<https://jurnal.uns.ac.id/ijai/article/view/44379>
- [5] Hastuti, Y, Multinomial Naves Network Status Gizi Balita, Cv.Azka Pustaka, 2023, Sumatra Barat
https://www.google.co.id/books/edition/Multinomial_Bayes_Network_Status_Gizi_Ba_DirkeaaaqbaJ?hl=Id&gbpv=1&Dq=Definisi

+Metode+Naive+Bayes&Pg=Pa69&Printsec=Frontcover

[6] Hengki, H,K Dan Rusman, A, D, P, Model Prediksi Stunting, 2020
https://www.google.co.id/books/edition/Model_Prediksi_Stunting/Igimeaaaqbj?hl=Id&Gbpv=1&Dq=Definisi+Prediksi&Pg=Pr6&Printsec=Frontcover

[7] Huda, A, S, Awangga, R, M dan Fathonah, R, N, S, Prediksi Penerimaan Pegawai Baru Dengan Metode Nave Bayes, Kreatif Industri Nusantara, 2020, Bandung
https://www.google.co.id/books/edition/Prediksi_Penerimaan_Pegawai_Baru_Dengan/aawGEAAAQBAJ?hl=id&gbpv=1&dq=prediksi+adalah&pg=PA8&printsec=frontcover

[8] Qisthiano, M. R., Kurniawan, T. B., Negara, E. S., & Akbar, M. (2021). Pengembangan Model Untuk Prediksi Tingkat Kelulusan Mahasiswa Tepat Waktu dengan Metode Naïve Bayes. *Jurnal Media Informatika Budidarma*, 5(3), 987-994. Dilihat pada tanggal 20 Agustus 2024.
<http://stmikbudidarma.ac.id/ejurnal/index.php/mib/article/view/3030>

[9] Rahayu, P, W, Sudipa, I, G, I dan Suryani, Surachman, A, Ridwan, A, Darmawiguna, I, G, M, Sutoyo, M, N, Slamet, I, Harlina, S, Maysanjaya, I, M, D, Buku Ajar Data Mining, PT Sonpedia Publishing Indonesia, 2024, Jambi
https://www.google.co.id/books/edition/Buku_Ajar_Data_Mining/vCruEAAAQBAJ?hl=id&gbpv=1&dq=Buku+Ajar+Data+Mining&pg=PR2&printsec=frontcover

[10] Rusdiana Dan Kodir, A, Pengelolaan Madrasah Diniyah Kontemporer, Yayasan Darul Hikam, 2022, Bandung
https://www.google.co.id/books/edition/Pengelolaan_Madrasah_Diniyah_Kontemporer/Dr1meaaaqbj?hl=Id&Gbpv=1&Dq=Definisi+Kelulusan+Dalam+Suatu+Pendidikan+Adalah&Pg=Pa273&Printsec=Frontcover

[11] Setiawan, Z, Fajar, M, Priyatno, A, M, Putri A, Y, P, Aryuni, M, Yuliyanti, S, Widiputra, H, Meilani, B, D, Ibrahim, R, N, Azdy, R, A, Junaidi, S, Wijaya, A, Buku Ajar Data Mining, 2023
https://www.google.co.id/books/edition/BUKU_AJAR_DATA_MINING/1nLVEAAAQBAJ?hl=id&gbpv=1&dq=Buku+Ajar+Data+M

[12] Sidiq, Y, N, S, Fatonah, R, N, S, Riza, N, Metode Pengklasifikasian Menentukan Kenaikan Level Umkm Bandung Timur Dengan Algoritma Naïve Bayes Pada Sistem Juragan Berbasis Komunitas, Kreatif Industri Nusantara, 2020, Bandung
https://www.google.co.id/books/edition/Metode_Klasifikasi_Menentukan_Kenaikan_Level/1Biv9dwaaaqbj?hl=Id&Gbpv=1&Dq=Definis

i+Metode+Naive+Bayes&Pg=Pa11&Printsec=Frontcover

[13] Suntoro, J 2019, Data Mining Algoritma dan Implementasi dengan Pemrograman PHP, PT Elex Media Komputindo
https://www.google.co.id/books/edition/Data_Mining_Algoritma_dan_Implementasi_d/8NKdDwAAQBAJ?hl=id&gbpv=1&dq=definisi+data+mining&pg=PA2&printsec=frontcover

[14] Sutrisman, D, Pendidikan Politik, Persepsi, Kepemimpinan Dan Mahasiswa, 2019
https://www.google.co.id/books/edition/Pendidikan_Politik_Persepsi_Kepemimpinan/OaEDwAAQBAJ?hl=id&gbpv=1&dq=definisi+mahasiswa+menurut+sutrisman&printsec=frontcover

[15] Taufiq, Analisis Sistem Informasi (Konsep, Metodologi dan Pendekatan), CV Ananta Vidya, 2023
https://www.google.co.id/books/edition/Analisis_Sistem_Informasi_Konsep_Metodologi/Axdqeaqbaaj?hl=Id&Gbpv=1&Dq=Definisi+Analisis&Pg=Pa8&Printsec=Frontcover

IDENTITAS PENULIS

Nama : Rike Limia Budiarti, M.Kom
 NIDN/NIK : -/006128802
 TTL : Jambi, 06 Desember 1988
 Gol / Pangkat : IIC
 Jab. Fungsional : Lektor
 Alamat Rumah : Kenali
 Telp. : 0822-8103-7072
 Email : Rikelimia@gmail.com

Nama : Novhirtamely Kahar ST.,M.Kom
 NIDN/NIK : 06.025/1015118101
 TTL : Jambi, 15 November 1981
 Gol / Pangkat : IIC
 Jab. Fungsional : Lektor
 Alamat Rumah : -
 Telp. : 0823-7825-6664
 Email : novmely@gmail.com.

Nama : Rismawati
 NIDN/NIK : 2002059
 TTL : Palembang, 13 Mei 2000
 Gol / Pangkat : -
 Jab. Fungsional : -
 Alamat Rumah : Kumpeh
 Telp. : 0813-1151-1411
 Email : rismawati.jambi20@gmail.com