

## Analisis Prediksi Kelulusan Mahasiswa Universitas Dinamika Bangsa Menggunakan Metode *Naïve Bayes*

Muhammad Furqan Hakim<sup>1</sup>, Sahril Saputra<sup>2</sup>

<sup>1,2</sup> Pascasarjana, Magister Sistem Informasi, Universitas Dinamika Bangsa, Jambi, Indonesia

Email: <sup>1</sup> [mfurqanhakim139@unama.ac.id](mailto:mfurqanhakim139@unama.ac.id), <sup>2</sup> [sahril.sputra244@gmail.com](mailto:sahril.sputra244@gmail.com)

Email Penulis Korespondensi: [mfurqanhakim139@unama.ac.id](mailto:mfurqanhakim139@unama.ac.id)

### Artikel Info :

Artikel History :

Submitted : 02-10-2024

Accepted : 14-04-2025

Published : 30-04-2025

### Kata Kunci:

Penambangan\_Data,  
Pembelajaran\_Mesin,  
Naïve\_Bayes, Weka,  
Mahasiswa

### Keywords:

Data\_Mining,  
Machine\_Learning,  
Naïve\_Bayes, Weka,  
Students

**Abstrak**– Untuk memfasilitasi proses pembelajaran, Universitas Dinamika Bangsa (UNAMA) memiliki basis data. Setiap tahun, data alumni semakin besar, dan basis data dapat digunakan. Memanfaatkan data alumni melibatkan pengklasifikasian dan analisis periode studi jangka panjang mahasiswa Universitas Dinamika Bangsa (UNAMA) menggunakan metode *naïve bayes*. Hasil klasifikasi *naïve bayes* dalam Mahasiswa program studi Sistem Informasi, dengan akurasi tertinggi diperoleh dengan menggunakan Use Training Set, yang terdiri dari 161 instance yang diklasifikasikan dengan benar dan 39 instance yang diklasifikasikan dengan salah, dengan persentase akurasi 85% untuk instance yang diklasifikasikan dengan benar dan 19,5% untuk instance yang diklasifikasikan dengan salah. Hasil seleksi atribut menggunakan algoritma *Classifier Attribute Evaluation (ClassifierAttributeEval)* menunjukkan bahwa IPK adalah atribut yang memiliki pengaruh terbesar terhadap kecepatan kelulusan. Akurasi dalam model dihitung menggunakan *confusion matrix*, dan pada awal data-data mahasiswa, terdapat banyak *noise*, yang terungkap melalui proses *data cleaning*. Ini adalah proses mengurangi *noise* dalam data menggunakan Microsoft Excel, yang biasanya digunakan oleh penulis untuk menganalisis data dari mahasiswa Sistem Informasi. Akurasi keseluruhan adalah 77,5%, yang merupakan akurasi yang sangat baik saat menganalisis data pelatihan.

**Abstract**– In order to facilitate the learning process, Universitas Dinamika Bangsa (UNAMA) has a database. Every year, the alumni data gets larger, and the database can be used. Utilising alumni data involves classifying and analysing long-term study periods of Universitas Dinamika Bangsa (UNAMA) students using the *naïve bayes* method. The results of the *naïve bayes* classification in the student of information system with the highest accuracy are obtained by using the Use Training Set, which consists of 161 correctly classified instances and 39 incorrectly classified instances, with an accuracy percentage of 85% for correctly classified instances and 19.5% for incorrectly classified instances. The results of attribute selection using the Classifier Attribute Evaluation algorithm (*ClassifierAttributeEval*) indicate that IPK is the attribute that has the greatest influence on kelulusan speed. Akurasi in the model is calculated using a confusion matrix, and at the beginning of the data-data mahasiswa, there is a lot of noise, which is revealed through the data cleaning process. It is the process of reducing noise in data using Microsoft Excel, which the author typically uses to analyse data From of information System students. Overall accuracy is 77.5%, which is a very good accuracy when analysing training data.

## 1. PENDAHULUAN

Saat melakukan penelitian, Universitas Dinamika Bangsa (UNAMA) memiliki basis data. Data dari alumni terus berkembang setiap tahun, dan basis data ini dapat digunakan. Dalam pengaturan yang khas, mahasiswa S1 UNAMA dapat melanjutkan studi mereka selama empat tahun atau delapan semester. Salah satu pertimbangan terpenting dalam proses akreditasi sebuah institusi tinggi adalah kelulusan mahasiswa. BAN-PT memiliki acuan untuk setiap dosen perguruan tinggi. Oleh karena itu, jika seorang dosen memiliki lulusan tepat waktu untuk setiap semester, hal itu dapat membantu dosen tersebut dalam proses akreditasi. Namun, dalam penerapan tingkat kelulusan mahasiswa, kita tidak dapat mengharapkan mereka lulus tepat waktu [1]. Tantangan dalam penelitian ini adalah bagaimana mengkategorikan dan menganalisis periode studi jangka panjang mahasiswa Universitas Dinamika Bangsa (UNAMA) menggunakan metode *naïve bayes*. Solusi untuk masalah yang terjadi adalah dengan menerapkan judul "Algoritma *Naïve Bayes* Untuk Mencari Lama Waktu Studi Mahasiswa (Studi Kasus: Universitas Dinamika Bangsa)" yang, kemungkinan besar, dapat mengklasifikasikan lama waktu studi atau kategori mahasiswa kelulusan yang telah selesai dengan menggunakan data kelulusan sebelumnya.

Dalam penelitian Hozairi [2], hasil menunjukkan bahwa model K-Nearest Neighbor, Decision Tree, dan Naive Bayes digunakan untuk mengklasifikasikan status kelulusan mahasiswa Teknik Informatika Universitas Islam Madura. Temuan menunjukkan bahwa kinerja Naive Bayes lebih baik dibandingkan dengan K-Nearest Neighbor dan Decision Tree. Dari 35 data uji, Naive Bayes mencapai akurasi 89% dan presisi 88%, sementara K-Nearest Neighbor mencatat akurasi 77% dan presisi 76%, serta Decision Tree dengan akurasi 74% dan presisi 84%. Terdapat juga penelitian sebelumnya yang memanfaatkan Naive Bayes untuk memprediksi berbagai kondisi. Dalam penelitian Tias Mugi Rahayu [3], akurasi data yang terklasifikasi dengan benar mencapai 69,33%.

Sedangkan dalam penelitian Ananda Fiqri Firdaus [4], ditemukan bahwa rasio data training dan testing yang diuji—60:40, 70:30, dan 80:20—menghasilkan akurasi, presisi, dan recall paling optimal dengan rasio 80:20.

Dalam penelitian Supardi Salmu dan Achmad Solichin [5] yang berjudul “Prediksi Tingkat Kelulusan Mahasiswa Tepat Waktu Menggunakan Naïve Bayes: Studi Kasus UIN Syarif Hidayatullah Jakarta,” diidentifikasi masalah sebaran yang tidak seimbang antara jumlah mahasiswa yang masuk dan yang lulus tepat waktu. Jumlah mahasiswa baru lebih banyak, sedangkan yang lulus tepat waktu jauh lebih sedikit. Penelitian ini menggunakan metodologi pengembangan data mining CRISP-DM, dan hasil pengujian menunjukkan bahwa metode klasifikasi Naïve Bayes mencapai akurasi sebesar 80,72% dari 1162 data untuk training dan 587 data untuk testing. Sementara itu, studi lain [6] mengembangkan aplikasi prediksi masa studi mahasiswa dengan mempertimbangkan data akademik, dan algoritma Naïve Bayes yang digunakan dapat memprediksi masa studi dengan akurasi 85,17% pada lima semester data yang diuji. Penelitian [7] lainnya menggunakan Decision Tree berbasis Particle Swarm Optimization untuk memprediksi kelulusan mahasiswa, yang menghasilkan akurasi sangat tinggi sebesar 97,67%..

Selanjutnya, dalam penelitian [8], dilakukan perbandingan antara Algoritma Naive Bayes dan C4.5 untuk menganalisis lama waktu kelulusan mahasiswa. Hasilnya menunjukkan bahwa algoritma C4.5 memiliki akurasi yang lebih baik, yaitu sebesar 84,95%, dengan selisih tidak terlalu signifikan sebesar 1,59% dibandingkan dengan Naive Bayes. Penelitian lain [9] berfokus pada seleksi calon mahasiswa baru menggunakan Algoritma Support Vector Machine (SVM), yang menghasilkan tingkat akurasi sebesar 65,00%. Metode klasifikasi Naive Bayes dipilih karena merupakan metode probabilitas statistik yang sederhana namun mampu memberikan hasil akurat dengan tingkat akurasi di atas 80%. Selain itu, studi [10] mengidentifikasi beberapa faktor yang mempengaruhi kelulusan mahasiswa, seperti jalur masuk perguruan tinggi, jenis kelamin, IP semester awal, daerah asal, dan kondisi ekonomi keluarga.

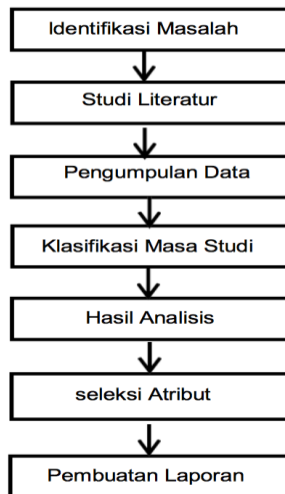
Menunjukkan bahwa penelitian mengenai kelulusan mahasiswa di berbagai institusi, termasuk Universitas Islam Madura, menggarisbawahi pentingnya pemanfaatan algoritma Naive Bayes dalam menganalisis dan mengklasifikasikan lama waktu studi. Dengan data alumni yang terus berkembang, penerapan Naive Bayes diharapkan dapat memberikan gambaran yang lebih jelas mengenai periode studi mahasiswa, serta membantu dalam proses akreditasi dengan memastikan lulusan tepat waktu. Berbagai studi sebelumnya juga mendukung efektivitas metode ini, menunjukkan akurasi yang cukup tinggi dibandingkan dengan metode lainnya, seperti K-Nearest Neighbor dan C4.5.

Selain itu, penelitian-penelitian yang ada menyoroti tantangan yang dihadapi, seperti sebaran yang tidak seimbang antara jumlah mahasiswa yang diterima dan yang lulus tepat waktu. Berbagai faktor yang mempengaruhi kelulusan juga telah diidentifikasi, termasuk jalur masuk, jenis kelamin, dan kondisi ekonomi. Dengan demikian, pendekatan berbasis data mining dan metode klasifikasi yang tepat tidak hanya dapat meningkatkan pemahaman tentang kelulusan mahasiswa, tetapi juga memberikan wawasan yang berguna untuk perbaikan sistem pendidikan di institusi terkait. Penelitian ini menekankan perlunya pendekatan komprehensif dalam analisis kelulusan mahasiswa dengan mempertimbangkan berbagai faktor yang mempengaruhi. Algoritma Naive Bayes terbukti efektif dalam mengklasifikasikan status kelulusan berdasarkan data historis. Dengan memanfaatkan algoritma ini, institusi dapat mengidentifikasi pola dan tren, yang pada gilirannya membantu dalam pengembangan strategi akademik dan administratif yang lebih baik, serta meningkatkan keberhasilan mahasiswa di masa depan.

## 2. METODOLOGI PENELITIAN

### 2.1 Tahapan Penelitian

Langkah-langkah yang diambil sepanjang penelitian ini dimaksudkan untuk memfasilitasi pencapaian hasil, memastikan bahwa penelitian selesai tepat waktu, dan memastikan bahwa proses berjalan sesuai rencana. Ruang lingkup pekerjaan yang dilakukan selama studi dapat dilihat pada Gambar 1.



**Gambar 1.** Tahapan Penelitian

Berdasarkan kerangka kerja yang telah diilustrasikan di atas, berikut dapat disimpulkan dari setiap langkah penelitian:

### 2.1.1 Identifikasi Masalah

Pada tahap ini, penulis mengidentifikasi apa sebenarnya masalah data mining terkait masa studi data MA di UNAMA, sehingga penulis dapat menentukan jadwal kerja dan mengidentifikasi data spesifik yang akan dibutuhkan untuk penelitian ini [11].

### 2.1.2 Pengumpulan Data

Pada tahap ini, penulis mengumpulkan data dan informasi yang dibutuhkan untuk penelitian ini. Pengumpulan data ini dilakukan dengan menggunakan beberapa metode berbeda, yaitu:

#### a. Pengamatan (Observasi)

Metode ini dilakukan melalui pengamatan langsung terhadap objek penelitian. Dalam konteks ini, penulis mengamati pengelolaan data mahasiswa untuk memperoleh softcopy data mahasiswa tahun 2017 dan 2018. Data tersebut diberikan oleh Kepala BAAK UNAMA. Atribut yang diperoleh dari data mahasiswa meliputi nim, nama, program studi, angkatan, jenis kelamin, status kuliah, tanggal lulus, asal sekolah, jurusan asal sekolah, dan IPK.

#### b. Wawancara

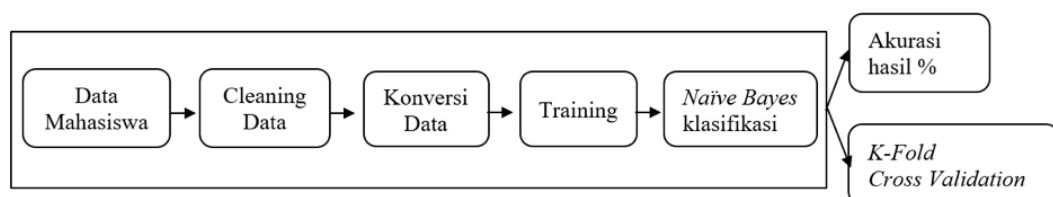
Metode ini dilakukan melalui observasi langsung terhadap objek penelitian. Dalam hal ini, penulis mengamati proses pengelolaan data mahasiswa, sehingga penulis memperoleh softcopy data tersebut.

### 2.1.3 Studi Literatur

Pada tahap ini, penulis mencari dasar-dasar teori yang diambil dari berbagai sumber, seperti buku, jurnal ilmiah, dan referensi lainnya, untuk melengkapi penelitian terkait konsep dan teori, sehingga memiliki acuan yang tepat dan relevan.

### 2.1.4 Klasifikasi Lama Masa Studi Mahasiswa

Proses klasifikasi data mahasiswa dilakukan melalui beberapa tahap. Alur proses klasifikasi ini dapat dilihat pada Gambar 2.



**Gambar 2.** Tahapan proses klasifikasi dan analisis dengan metode naïve bayes

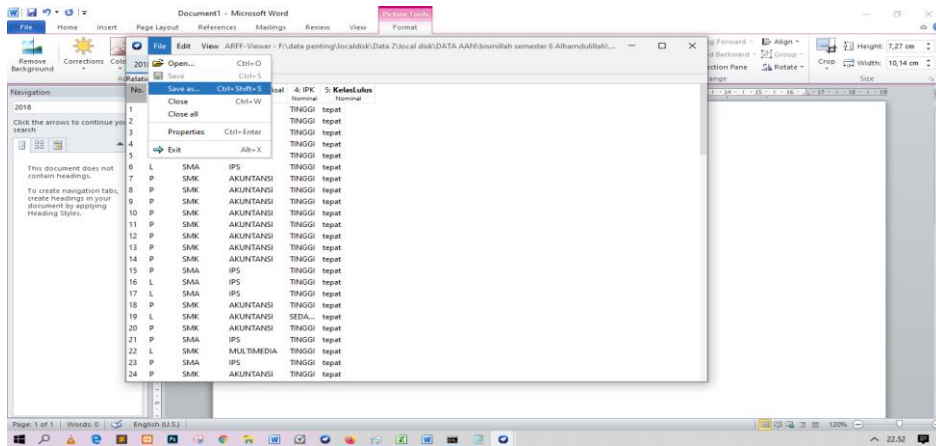
### 3. HASIL DAN PEMBAHASAN

#### 3.1. Hasil

##### 3.1.1. Seleksi Atribut Menggunakan Classifier Attribute evaluator

###### 1. Konversi data dari csv ke arff

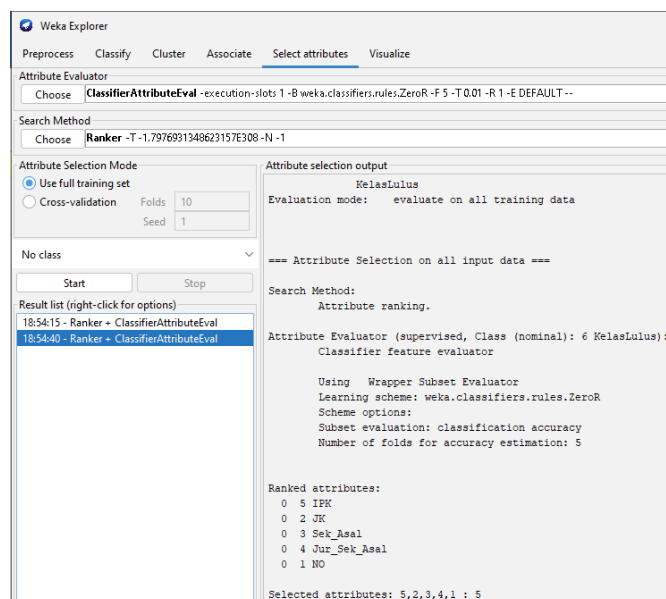
pada menu tools pilih arff viewer, kemudian terbuka windows arff viewer lalu open file yang berextensi csv setelah terbuka, maka kita save as dan pilih extensi arff yang terlihat pada Gambar 4.



Gambar 3 save as pada Arrf Viewer

###### 2. WEKA seleksi atribut

Pilih menu select attributes dan pilih classifierattributeeval sebagai evaluatoryA dan tekan start. Pada gambar 5 dapat dilihat seluruh atribut terdiri dari 4 atribut yang penulis gunakan untuk klasifikasi lama masa studi mahasiswa dan atribut-atribut ini belum diseleksi. Proses seleksi ini dilakukan untuk mengidentifikasi atribut-atribut yang paling berpengaruh terhadap kecepatan kelulusan. Dalam hal ini, IPK terbukti menjadi atribut yang sangat signifikan. Seleksi atribut ini menggunakan algoritma classifier attribute evaluation (classifierattributeeval) pada alat Weka, yang menunjukkan bahwa IPK adalah atribut yang paling berpengaruh terhadap kecepatan kelulusan.



Gambar 4. Select attributes menggunakan ClassifierAttributeEvaluator

### 3.1.2 Hasil Klasifikasi Naïve Bayes Menggunakan Tools WEKA

#### 1. Hasil Klasifikasi Naïve Bayes Menggunakan Tools WEKA (Use Data Training)

Hasil klasifikasi pada Tools WEKA dengan menggunakan data pelatihan menunjukkan bahwa terdapat 161 prediksi yang benar, dengan akurasi mencapai 80,5%. Selain itu, terdapat 39 prediksi yang salah, yang berpersentase 19,5%. Proses klasifikasi ini memerlukan waktu selama 0,02 detik.

```

Classifier output

Time taken to build model: 0 seconds

=== Evaluation on training set ===

Time taken to test model on training data: 0.02 seconds

=== Summary ===

Correctly Classified Instances      161          80.5 %
Incorrectly Classified Instances    39           19.5 %
Kappa statistic                    0.4374
Mean absolute error                 0.2608
Root mean squared error            0.3695
Relative absolute error             71.231 %
Root relative squared error        86.5186 %
Total Number of Instances          200

=== Detailed Accuracy By Class ===

          TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
0,895  0,479  0,855    0,895  0,875    0,440  0,804  0,921  tepat
0,521  0,105  0,610    0,521  0,562    0,440  0,804  0,660  terlambat
Weighted Avg.  0,805  0,389  0,796    0,805  0,800    0,440  0,804  0,858

=== Confusion Matrix ===
  a  b  <-- classified as
136 16 | a = tepat
 23 25 | b = terlambat
    
```

Gambar 5. Klasifikasi Naïve Bayes (Menggunakan Data Training)

#### 2. Hasil klasifikasi dengan naïve bayes menggunakan tools WEKA (10-cross validation)

Hasil klasifikasi dengan Naïve Bayes menggunakan Tools WEKA dengan metode 10-cross validation ditunjukkan pada Gambar 7. Hasil ini menunjukkan 156 prediksi yang benar dengan akurasi sebesar 78% dan 44 prediksi yang salah, dengan persentase 22%. Proses klasifikasi ini memerlukan waktu 0,01 detik.

```

Classifier
Choose NaiveBayes

Test options
Use training set
Supplied test set
Cross-validation Folds 10
Percentage split % 66

(Nom) Kelulusan
Start Stop
Result list (right-click for options)
22.29.00 - bayes.NaiveBayes
23.33.42 - bayes.NaiveBayes

Classifier output
SEMANG 27.0 25.0
BENDAR 3.0 11.0
[total] 185.0 51.0

Time taken to build model: 0.01 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      156          78 %
Incorrectly Classified Instances    44           22 %
Kappa statistic                    0.3792
Mean absolute error                 0.3829
Root mean squared error            0.3966
Relative absolute error             77.2218 %
Root relative squared error        92.8295 %
Total Number of Instances          200

=== Detailed Accuracy By Class ===

          TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
0,868  0,500  0,846    0,868  0,857    0,380  0,752  0,901  tepat
0,500  0,132  0,545    0,500  0,522    0,380  0,752  0,555  terlambat
Weighted Avg.  0,780  0,412  0,774    0,780  0,777    0,380  0,752  0,818

=== Confusion Matrix ===
  a  b  <-- classified as
132 20 | a = tepat
 24 24 | b = terlambat
    
```

Gambar 6. Klasifikasi Naïve Bayes (10-cross validation)

## 3.2. Pembahasan

### 3.2.1 Perhitungan Manual Naïve Bayes

“Naïve Bayesian Classification (NBC) adalah metode pengklasifikasian statistik yang digunakan untuk memprediksi probabilitas keanggotaan suatu kelas. Proses klasifikasi teks dengan metode NBC terdiri dari dua tahap: tahap pelatihan dan tahap pengujian/klasifikasi.” [17]. Berikut adalah persamaan Naïve Bayes:

$$P(C_i | X) = \frac{P(X | C_i) * P(C_i)}{P(X)} \quad (1)$$

Keterangan:

X : Kriteria untuk suatu kasus berdasarkan input.

C : Kelas solusi pola ke-i, di mana i menunjukkan jumlah label kelas.

P(C<sub>i</sub>|X) : Probabilitas munculnya label kelas C<sub>i</sub> berdasarkan kriteria input X.

P(X|C<sub>i</sub>) : Probabilitas input X diberikan label kelas C<sub>i</sub>.

P(C<sub>i</sub>) : Probabilitas dari label kelas C<sub>i</sub>.

**Tabel 2. Data Testing**

JK	Sek_Asal	Jur_Sek_Asal	IPK	KelasLulus
P	SMK	AKUNTANSI	TINGGI	?

1. Menghitung jumlah kelas/label

P (KelasLulus = TEPAT) = jumlah seluruh kelas cepat dari total data = **152/200**

P (KelasLulus = LAMBAT) = jumlah seluruh kelas lambat dari total data = **48/200**

2. Menghitung jumlah data per kelas/atribut yang ada

Cara menghitung jumlah data per atribut mirip dengan menghitung jumlah kelas. Pertama, tentukan jumlah mahasiswa laki-laki (L) dari setiap kategori kelas: cepat, tepat, dan lambat. Kemudian, lakukan hal yang sama untuk mahasiswa perempuan (P) di setiap kategori kelas tersebut. Selanjutnya, hitung probabilitas untuk seluruh atribut.

Rumusnya adalah P(X | C<sub>i</sub>), yang berarti P(Jenis Kelamin | Kelas Lulus).

**Tabel 3. P robability Jenis Kelamin diberikan KelasLulus**

JK	KelasLulus=Tepat	KelasLulus=Lambat
L	0,427631579	0,197368421
P	1,8125	0,375

Probability Jenis Kelamin diberikan KelasLulus yaitu :

P(JK=L | KelasLulus = TEPAT) = 65/152 = 0,427631579

P (JK=P | KelasLulus = LAMBAT) = 18/48 = 0,375

P (JK=L | KelasLulus = LAMBAT) = 30/152 = 0,197368421

P (JK=P | KelasLulus = TEPAT) = 87/48 = 1,8125

P(Sekolah Asal | Kelas Lulus)

**Tabel 4. Probability Sekolah Asal diberikan KelasLulus**

Sek_Asal	KelasLulus=Tepat	KelasLulus=Lambat
SMA	0,5	0,75
SMK	0,453947368	0,208333333
MA	0,046052632	0,041666667

Probability Sekolah Asal diberikan KelasLulus yaitu :

P (SEK\_ASAL=SMK | KELASLULUS = TEPAT) = 69/152 = 0,453947368

P (SEK\_ASAL=SMK | KELASLULUS = LAMBAT) = 10/48 = 0,208333333

P (SEK\_ASAL=SMA | KELASLULUS = TEPAT) = 76/152=0,5

$$P(\text{SEK\_ASAL}=\text{SMA} \mid \text{KELASLULUS} = \text{LAMBAT}) = 36/48 = 0,75$$

$$P(\text{SEK\_ASAL}=\text{MA} \mid \text{KELASLULUS} = \text{TEPAT}) = 7/152 = 0,046052632$$

$$P(\text{SEK\_ASAL}=\text{MA} \mid \text{KELASLULUS} = \text{LAMBAT}) = 2/48 = 0,041666667$$

P (Jur\_Sek\_Asal | KelasLulus )

**Tabel 5. Probability Jurusan Sekolah Asal diberikan KelasLulus**

Sek_Asal	KelasLulus=Tepat	KelasLulus=Lambat
IPA	0,157894737	0,229166667
IPS	0,381578947	0,541666667
AKUNTANSI	0,263157895	0,0625
TKJ	0,019736842	0,041666667
MULTIMEDIA	0,105263158	0,0625
ADMINISTRASI	0,026315789	0,020833333
TATABUSANA	0	0,020833333
GRAFIKA	0,006578947	0
BROADCASTING	0,013157895	0
KEAGAMAAN	0,006578947	0,020833333
PEMASARAN	0,019736842	0

Probability Jurusan Sekolah Asal diberikan KelasLulus yaitu:

$$P(\text{Jur\_Sek\_Asal}=\text{IPA} \mid \text{KelasLulus} = \text{TEPAT}) = 24/152 = 0,157894737$$

$$P(\text{Jur\_Sek\_Asal}=\text{IPS} \mid \text{KelasLulus} = \text{TEPAT}) = 58/152 = 0,381578947$$

$$P(\text{Jur\_Sek\_Asal}=\text{AKUNTANSI} \mid \text{KelasLulus} = \text{TEPAT}) = 40/152 = 0,263157895$$

$$P(\text{Jur\_Sek\_Asal}=\text{TKJ} \mid \text{KelasLulus} = \text{TEPAT}) = 3/152 = 0,019736842$$

$$P(\text{Jur\_Sek\_Asal}=\text{MULTIMEDIA} \mid \text{KelasLulus} = \text{TEPAT}) = 16/152 = 0,105263158$$

$$P(\text{Jur\_Sek\_Asal}=\text{ADMINISTRASI} \mid \text{KelasLulus} = \text{TEPAT}) = 4/152 = 0,026315789$$

$$P(\text{Jur\_Sek\_Asal}=\text{TATABUSANA} \mid \text{KelasLulus} = \text{TEPAT}) = 0/152 = 0$$

$$P(\text{Jur\_Sek\_Asal}=\text{GRAFIKA} \mid \text{KelasLulus} = \text{TEPAT}) = 1/152 = 0,006578947$$

$$P(\text{Jur\_Sek\_Asal}=\text{BROADCASTING} \mid \text{KelasLulus} = \text{TEPAT}) = 2/152 = 0,013157895$$

$$P(\text{Jur\_Sek\_Asal}=\text{KEAGAMAAN} \mid \text{KelasLulus} = \text{TEPAT}) = 1/152 = 0,006578947$$

$$P(\text{Jur\_Sek\_Asal}=\text{PEMASARAN} \mid \text{KelasLulus} = \text{TEPAT}) = 3/152 = 0,019736842$$

$$P(\text{Jur\_Sek\_Asal}=\text{IPA} \mid \text{KelasLulus} = \text{LAMBAT}) = 11/48 = 0,229166667$$

$$P(\text{Jur\_Sek\_Asal}=\text{IPS} \mid \text{KelasLulus} = \text{LAMBAT}) = 26/48 = 0,541666667$$

$$P(\text{Jur\_Sek\_Asal}=\text{AKUNTANSI} \mid \text{KelasLulus} = \text{LAMBAT}) = 3/48 = 0,0625$$

$$P(\text{Jur\_Sek\_Asal}=\text{TKJ} \mid \text{KelasLulus} = \text{LAMBAT}) = 2/48 = 0,041666667$$

$$P(\text{Jur\_Sek\_Asal}=\text{MULTIMEDIA} \mid \text{KelasLulus} = \text{LAMBAT}) = 3/48 = 0,0625$$

$$P(\text{Jur\_Sek\_Asal}=\text{ADMINISTRASI} \mid \text{KelasLulus} = \text{LAMBAT}) = 3/48 = 0,020833333$$

$$P(\text{Jur\_Sek\_Asal}=\text{TATABUSANA} \mid \text{KelasLulus} = \text{LAMBAT}) = 1/48 = 0,020833333$$

$$P(\text{Jur\_Sek\_Asal}=\text{GRAFIKA} \mid \text{KelasLulus} = \text{LAMBAT}) = 0/48 = 0$$

$$P(\text{Jur\_Sek\_Asal}=\text{BROADCASTING} \mid \text{KelasLulus} = \text{LAMBAT}) = 0/48$$

$$P(\text{Jur\_Sek\_Asal}=\text{KEAGAMAAN} \mid \text{KelasLulus} = \text{LAMBAT}) = 1/48 = 0,020833333$$

$$P(\text{Jur\_Sek\_Asal}=\text{PEMASARAN} \mid \text{KelasLulus} = \text{LAMBAT}) = 0/48 = 0$$

P (IPK | KelasLulus)

**Tabel 6. Probability IPK diberikan KelasLulus**

IPK	KelasLulus=Tepat	KelasLulus=Lambat
RENDAH	0,01316	0,20833
SEDANG	0,17105	0,45833
TINGGI	0,81579	0,33333

Probability Sekolah Asal diberikan KelasLulus yaitu :

$$\begin{aligned}
 P(\text{IPK}=\text{RENDAH} \mid \text{KELASLULUS} = \text{TEPAT}) &= 69/152= 0,453947368 \\
 P(\text{IPK}=\text{RENDAH} \mid \text{KELASLULUS} = \text{LAMBAT}) &= 10/48=0,208333333 \\
 P(\text{IPK}=\text{SEDANG} \mid \text{KELASLULUS} = \text{LAMBAT}) &= 36/48=0,75 \\
 P(\text{IPK}=\text{SEDANG} \mid \text{KELASLULUS} = \text{TEPAT}) &= 7/152=0,046052632 \\
 P(\text{IPK}=\text{TINGGI} \mid \text{KELASLULUS} = \text{LAMBAT}) &= 36/48=0,75 \\
 P(\text{IPK}=\text{TINGGI} \mid \text{KELASLULUS} = \text{TEPAT}) &= 7/152=0,046052632
 \end{aligned}$$

3. Mengkalikan semua variabel kelas untuk mencari KelasLulus

Menggunakan rumus Naïve bayes untuk mencari kelas lulus dengan variable kelas lulus sebagai variable yang ingin dicari Kelas Lulusnya

$$\text{KelasLulus}=(\text{JK}=\text{P},\text{Sek\_Asal}=\text{SMK},\text{Jur\_Sek\_Asal}=\text{AKUNTANSI},\text{IPK}=\text{TINGGI}) (2)$$

4. Perkalian variabel kelas untuk KelasLulus TEPAT

Pada persamaan dibawah adalah untuk mencari KelasLulus Tepat dengan *probability* kolom P yang TEPAT, kolom; SMK yang TEPAT, kolom AKUNTANSI yang TEPAT dan kolom TINGGI yang TEPAT  

$$\text{PKelasLulus}=\text{PTEPAT}*\text{PTEPAT}*\text{PTEPAT}*\text{PTEPAT} =1,8125* 0,453947368* 0,263157895*0,81579 =0,176635624796386940409225 \quad (3)$$

5. Perkalian variabel kelas untuk KelasLulus LAMBAT

Persamaan di bawah ini digunakan untuk menghitung probabilitas Kelas Lulus Tepat, dengan mempertimbangkan atribut P (yang TEPAT), SMK (yang TEPAT), AKUNTANSI (yang TEPAT), dan TINGGI (yang TEPAT).  

$$\text{PKelasLulus}=\text{PLAMBAT}*\text{PLAMBAT}*\text{PLAMBAT}*\text{PLAMBAT} =0,375* 0,208333333* 0,0625* 0,33333 = 0,001627587888020859375 (4)$$

6. Membandingkan hasil per kelas

Berdasarkan hasil perhitungan manual dengan menggunakan sampel data pengujian yang telah disebutkan sebelumnya, berikut adalah data testing yang digunakan:

**Tabel 6. Perbandingan Hasil Perhitungan**

KelasLulus	JK,Sek_Asal,Jur_Sek_Asal,IPK
TEPAT	0,176635624796386940409225
LAMBAT	0,001627587888020859375

Karena nilai hasil perkalian dari seluruh probabilitas tertinggi adalah kelas TEPAT, dapat disimpulkan bahwa hasil kelas lulus untuk JK=P, Sek\_Asal=SMK, Jur\_Sek\_Asal=AKUNTANSI, dan IPK=TINGGI adalah TEPAT.

**3.2.2. Hasil Evaluasi Akurasi Dari Data Kelulusan Mahasiswa**

Hasil klasifikasi menggunakan Use Training Set dan 10-Fold Cross Validation menunjukkan akurasi masing-masing 80,5% dan 78% untuk prediksi yang benar. Perhitungan akurasi model dilanjutkan dengan confusion matrix setelah data pelatihan diolah.

**Tabel 7. Evaluasi Akurasi**

MODEL EVALUASI	Akurasi	Jumlah Kelas	Persentasi	Satuan
Use Training Set	Correctly Classified Instances	161	80.5	%
	Incorrectly Classified Instances	39	19.5	%
10 Fold Cross	Correctly Classified Instances	156	78	%
	Incorrectly Classified Instances	44	22	%



Validation

1. Akurasi confusion matrix pada use training set

$$\begin{aligned} \text{Presentasi Akurasi} &= ((TP + TN)/(TP + TN + FP + FN)) * 100\% \\ &= ((135 + 27) / (135 + 27 + 17 + 21)) * 100\% \\ &= 0,78 \text{ (78\%)} \quad (5) \end{aligned}$$

2. Akurasi confusion matrix pada 10 fold cross validation

$$\begin{aligned} \text{Presentasi Akurasi} &= ((TP + TN)/(TP + TN + FP + FN)) * 100\% \\ &= ((132 + 22) / (132 + 22 + 20 + 26)) * 100\% \\ &= 0,77 \text{ (77\%)} \quad (6) \end{aligned}$$

Presentasi Akurasi dihitung dari *Confusion Matrix* dari *use training set* yaitu TP atau *True Positive* adalah 135, FP atau *False Positive* adalah 17, TN atau *True Negative* adalah 27, FN atau *False Negative* adalah 21, kemudian dari Akurasi pada *10 fold cross validation* yaitu TP atau *True Positive* adalah 132, FP atau *False Positive* adalah 20, TN atau *True Negative* adalah 22, FN atau *False Negative* adalah 26, Tabel dari *Confusion Matrix* untuk metode *Use Training Set* dan *10-Fold Cross Validation* dapat dilihat pada Tabel 4.15 dan Tabel 4.16.

**Tabel 8.** *Confusion Matrix use training set*

	Positive	Negative
Positive	132 (TP)	20 (FP)
Negative	26 (FN)	22 (TN)

**Tabel 9.** *Confusion Matrix 10 fold cross validation*

	Positive	Negative
Positive	135 (TP)	17 (FP)
Negative	21 (FN)	27 (TN)

#### 4. KESIMPULAN

Penelitian ini bertujuan untuk memprediksi kecepatan kelulusan mahasiswa Universitas Dinamika Bangsa (UNAMA) angkatan 2017–2018 menggunakan algoritma *Naïve Bayes*. Data yang digunakan berjumlah 200 setelah melalui proses data cleaning untuk menghilangkan noise. Model dikembangkan dengan dua metode evaluasi: *Use Training Set*, yang menghasilkan akurasi 80,5%, dan *10-Fold Cross Validation*, yang mencapai akurasi 78%. Analisis atribut menggunakan *Classifier Attribute Evaluation* mengidentifikasi bahwa *Indeks Prestasi Kumulatif (IPK)* merupakan faktor paling signifikan yang mempengaruhi kelulusan mahasiswa tepat waktu.

Lebih lanjut, distribusi atribut seperti jenis kelamin, asal sekolah, jurusan asal sekolah, dan *IPK* dianalisis dalam kaitannya dengan dua kategori kelulusan: "Tepat Waktu" dan "Lambat". Hasil klasifikasi manual dan otomatis menunjukkan konsistensi dalam identifikasi pola kelulusan mahasiswa berdasarkan data historis. Evaluasi model melalui *confusion matrix* juga memperkuat keakuratan prediksi dengan akurasi keseluruhan mencapai 77,5%.

Secara keseluruhan, algoritma *Naïve Bayes* terbukti efektif dan akurat dalam mengklasifikasikan data kelulusan mahasiswa. Model ini dapat digunakan sebagai dasar dalam pengambilan keputusan akademik, penyusunan strategi pembinaan, serta peningkatan mutu layanan pendidikan di lingkungan UNAMA.

## REFERENCES

- [1] Y. Apridiansyah, et al., "Prediksi Kelulusan Mahasiswa Fakultas Teknik Informatika Universitas Muhammadiyah Bengkulu Menggunakan Metode Naive Bayes," *JSAI (Journal Sci. Appl. Informatics)*, vol. 4, no. 2, pp. 236–247, 2021, doi: 10.36085/jsai.v4i2.1701.
- [2] H. Hozairi, et al., "Implementasi Orange Data Mining Untuk Klasifikasi Kelulusan Mahasiswa Dengan Model K-Nearest Neighbor, Decision Tree Serta Naive Bayes," *Netw. Eng. Res. Oper.*, vol. 6, no. 2, p. 133, 2021, doi: 10.21107/nero.v6i2.237.
- [3] T. M. Rahayu, B. A. Ningsi, Isnurani, and I. Arofah, "KLASIFIKASI KETEPATAN WAKTU KELULUSAN MAHASISWA DENGAN METODE NAÏVE BAYES Oleh," vol. 15, no. 11, pp. 5651–5658, 2021.
- [4] A. F. Firdaus, et al., "Implementasi Metode Klasifikasi Naive Bayes Implementation of Naive Bayes Classification Method in Predicting," vol. 8, no. 5, pp. 9274–9279, 2021.
- [5] D. Salmu, S. and A. Solichin, "Prediksi Tingkat Kelulusan Mahasiswa Tepat Waktu Menggunakan Naive Bayes : Studi Kasus UIN Syarif Hidayatullah Jakarta Prediction of Timeliness Graduation of Students Using Naive Bayes : A Case Study at Islamic State University Syarif Hidayatullah Jakarta," *Pros. Semin. Nas. Multidisiplin Ilmu*, no. April, pp. 701–709, 2017.
- [6] Amelia, Mongan Winny, Lumenta, Arie S. M., & Jacobus, Agustinus. (2017). Prediksi Masa Studi Mahasiswa dengan Menggunakan Algoritma Naive Bayes. *E-Journal Teknik Informatika*. Vol 11, No.1 ISSN : 2301 – 8364
- [7] Rakhman, Arif. (2017), Prediksi Ketepatan Kelulusan Mahasiswa Menggunakan Metode Decision Tree Berbasis Particle Swarm Optimization. *Smart Comp*. Volume 6, No 1.
- [8] Risqiati & Ismanto, Bambang. (2017), Analisis Komparasi Algoritma Naive Bayes Dan C4-5 untuk Waktu Kelulusan Mahasiswa. *IC-Tech*. Volume XII, No. 1
- [9] Saifudin, Aries. (2018). Metode Data Mining untuk Seleksi Calon Mahasiswa pada Penerimaan Mahasiswa Baru di Universitas Pamulang. *Jurnal Teknologi*. Volume 10, No. 1
- [10] Darwin, D., & Zurimi, S. (2019). Analisis model aplikatif Multivariate Adaptive Regression Spline (MARS) terhadap klasifikasi faktor yang mempengaruhi masa studi mahasiswa FKIP Universitas Darussalam Ambon. *Jurnal Simetrik*, 9(2), 250–255. <http://ejournal-polnam.ac.id/index.php/JurnalSimetrik/artic/viewFile/426/278>
- [11] Duei Putri, D., Nama, G.F., & Sulistiono, W.E. (2022). Analisis Sentimen Kinerja Dewan Perwakilan Rakyat (DPR) Pada Twitter Menggunakan Metode Naive Bayes Classifier. *Jurnal Informatika dan Teknik Elektro Terapan*.
- [12] Al Maeni, N., Suarna, N., & Prihartono, W. (2024). ANALISIS SENTIMEN PENGGUNA SHOPEE BERDASARKAN DATA TWEET DARI TWITTER MENGGUNAKAN METODE NAIVE BAYES CLASSIFIER. *JATI (Jurnal Mahasiswa Teknik Informatika)*.
- [13] Harun, R., Ishak, R., & Panna, S. (2023). Analisis Sentimen Opini Publik Pengguna Twitter Terhadap Kenaikan Harga BBM Menggunakan Algoritma Naive Bayes. *Jurnal Ilmiah Ilmu Komputer Banthayo Lo Komputer*.
- [14] Natasuwarna, A.P. (2019). ANALISIS SENTIMEN KEPUTUSAN PEMINDAHAN IBUKOTA NEGARA MENGGUNAKAN KLASIFIKASI NAIVE BAYES.
- [15] Chandra, D., Indrawan, G.S., & Sukaraja, I.N. (2016). Klasifikasi Berita Lokal Radar Malang Menggunakan Metode Naive Bayes Dengan Fitur N-Gram.
- [16] Adyatma Subagja, R., Widiastiwi, Y., & Chamidah, N. (2021). Klasifikasi Ulasan Aplikasi Jenius pada Google Play Store Menggunakan Algoritma Naive Bayes. *Informatik : Jurnal Ilmu Komputer*.
- [17] Puji Astuti, A., Alam, S., & Jaelani, I. (2022). Komparasi Algoritma Support Vector Machine dengan Naive Bayes Untuk Analisis Sentimen Pada Aplikasi BRImo. *Jurnal Bangkit Indonesia*.