

Prediksi Kinerja Mahasiswa Dalam Perkuliahan Berbasis *Learning Management System* Menggunakan Algoritma *Naïve Bayes*

Asnur Karima^{1*}, Taghfirul Azhima Yoga Siswa²

Program Studi Teknik Informatika, Universitas Muhammadiyah Kalimantan Timur

Jl. Ir. H. Juanda No.15 Sidodadi, Samarinda, Indonesia

*Email Corresponding Author: 1811102441015@umkt.ac.id

Abstract

The Covid-19 pandemic that has hit Indonesia since the beginning of 2020 has had a major impact on the world of education, so that the learning process that was originally carried out face-to-face has turned into online learning. In such a situation, the University of Muhammadiyah East Kalimantan utilizes the Learning Management System (LMS) in an online learning system using the LMS Open Learning platform. The purpose of this study is to find the best attribute values using Correlation Based Featured Selection and to test the performance of the Naïve Bayes algorithm using Confusion Matrix. The attributes used after going through the feature selection are time spent on course, course completed, assignments, mid-semester exams and quizzes. The results of testing 178 data with a ratio of training data schemes and testing data of 70:30 produce an accuracy of 98.14%, 80:20 produces an accuracy of 97.22% and 90:10 produces an accuracy of 94.44%. Thus, the best accuracy is obtained at 70:30 data composition, which is 98.14%.

Keywords: Accuracy level; Naive Bayes; Online learning; Prediction

Abstrak

Pandemi Covid-19 yang melanda Indonesia sejak awal tahun 2020 memberikan dampak besar terhadap dunia pendidikan, sehingga proses pembelajaran yang semula dilakukan secara tatap muka berubah menjadi Pembelajaran Dalam Jaringan (daring). Dalam situasi seperti tersebut, Universitas Muhammadiyah Kalimantan Timur memanfaatkan *Learning Management System* (LMS) dalam sistem pembelajaran daring menggunakan *platform LMS Open Learning*. Tujuan penelitian ini adalah mencari nilai atribut terbaik menggunakan *Correlation Based Featured Selection* dan menguji performa algoritma *Naïve Bayes* menggunakan *Confusion Matrix*. Atribut yang digunakan setelah melalui seleksi fitur adalah *time spent on course*, *course completed*, *nilai penugasan*, *nilai Ujian Tenga Semester* dan *nilai quiz*. Hasil pengujian 178 data dengan rasio skema data *training* dan data *testing* 70:30 menghasilkan akurasi sebesar 98,14%, 80:20 menghasilkan akurasi sebesar 97,22% dan 90:10 menghasilkan akurasi sebesar 94,44%. Dengan demikian, akurasi terbaik diperoleh pada komposisi data 70:30, yaitu sebesar 98,14%.

Kata kunci: Akurasi; Naïve Bayes; Pembelajaran dalam jaringan; Prediksi

1. Pendahuluan

Perkuliahan daring menjadi marak sejak virus *Coronavirus Disease 2019 (Covid-19)* mewabah di Indonesia. *Covid-19* merupakan penyakit yang menyerang sistem pernapasan manusia [1]. Dampak dari *Covid-19* telah mempengaruhi berbagai sektor kehidupan masyarakat, tidak terkecuali pada sektor pendidikan. Oleh karena itu Pemerintah Indonesia melalui Kementerian Pendidikan dan Kebudayaan RI menerapkan kebijakan untuk belajar dan bekerja dari rumah [2]. Hal ini membuat sistem pembelajaran berubah dari pembelajaran konvensional menjadi pembelajaran dalam jaringan (daring).

Universitas Muhammadiyah Kalimantan Timur (UMKT) merupakan universitas swasta terbaik yang telah menerapkan pembelajaran daring dalam perkuliahannya. Pelaksanaan pembelajaran daring di UMKT dilakukan dengan menggunakan *Learning Management System* (LMS) berupa *Open Learning*. *Open Learning* merupakan *platform* penyedia pembelajaran daring

yang didalamnya dapat memasukkan materi pembelajaran. Seperti teks, gambar, video, audio dan fitur diskusi yang dapat memudahkan dosen untuk memberikan materi dan mahasiswa untuk mengakses materi.

Peralihan model pembelajaran konvensional menjadi pembelajaran daring bukanlah suatu hal yang mudah, sehingga butuh kesiapan pihak penyelenggara pendidikan dalam mempersiapkan infrastruktur, juga kesiapan pengajar dan mahasiswa untuk dapat beradaptasi dengan teknologi yang digunakan. Banyak hal yang mampu membuat penerapan pembelajaran daring ini mengalami kendala. Salah satunya adalah mahasiswa sulit menguasai atau mengerti materi yang disampaikan karena kurangnya interaksi antara dosen dan mahasiswa bertemu di dalam kelas. Karena kurangnya interaksi tersebut proses belajar mengajar menjadi kurang efisien. Dibutuhkan sebuah analisis untuk bisa memprediksi bahwa mahasiswa ini benar-benar memiliki kinerja yang baik dalam perkuliahan daring.

Algoritma *Naïve Bayes* merupakan algoritma yang dapat mengklasifikasikan suatu variabel tertentu dengan menggunakan metode probabilitas dan *statistic* [3]. Algoritma *Naïve Bayes* telah digunakan secara meluas dalam melakukan prediksi, seperti dalam bidang pendidikan [4]–[7], bidang penjualan [8][9], bidang kesehatan [10][11], dan bidang bisnis lainnya [12][13][14].

Penelitian ini memprediksi kinerja mahasiswa dalam perkuliahan daring melalui pendekatan *data analytic* yaitu *data mining*, dengan menggunakan salah satu teknik *data mining*, yaitu metode klasifikasi berupa algoritma *Naïve bayes*. Penelitian ini mencari nilai atribut terbaik menggunakan *correlation based featured selection* dan menguji performa algoritma *Naïve Bayes* dalam memprediksi kinerja mahasiswa pada perkuliahan daring berbasis *learning management system*.

2. Tinjauan Pustaka

Berbagai penelitian yang telah dilakukan mengenai kinerja mahasiswa dalam perkuliahan diantaranya, efektivitas perkuliahan daring (*online*) pada mahasiswa PGSD di saat pandemi *Covid-19* memberikan gambaran umum mengenai kurang optimalnya mahasiswa dalam memahami materi dan banyaknya tugas diberikan oleh dosen membuat mahasiswa menjadi kurang efektif. Dibuktikan dengan hasil persentase paham materi 15.59%, kurang paham 75.81%, dan tidak paham 8.60% [15]. Analisis Respon Mahasiswa Terhadap Perkuliahan Daring di Prodi Biologi STKIP PI Makassar memberikan gambaran umum mengenai pembelajaran *online* tersebut masih belum efektif dan efisien. Dibuktikan dengan hasil respon yaitu sebanyak 25,3% mahasiswa setuju dan 6% sangat setuju serta merasa efisien dalam pembelajaran *online*. Sedangkan sebanyak 61,4% mahasiswa tidak setuju dan 7,2% sangat tidak setuju serta merasa tidak efisien dalam penerapan pembelajaran *online* [16].

Prediksi kinerja akademik mahasiswa menggunakan *Machine Learning* dengan *Sequential Minimal Optimization* untuk pengelola program studi dengan data yang digunakan berupa data mahasiswa Jurusan Teknik Informatika UNESA berjumlah 330 data. Berdasarkan hasil pengujian diperoleh akurasi 93,94%, presisi 94,7% dan *recall* 94,7% [17]. Memprediksi kinerja mahasiswa memanfaatkan dua model klasifikasi dengan data yang digunakan berupa data mahasiswa berjumlah 40 orang. Hasil dari pengujian data diperoleh akurasi algoritma *Naïve Bayes* yaitu 80% lebih tinggi dibandingkan algoritma C4.5 yang hanya menghasilkan akurasi 60% [18].

Berdasarkan penelitian terkait yang telah dilakukan sebelumnya tentang kinerja mahasiswa dalam perkuliahan daring, berbagai macam metode pernah digunakan dalam proses penyelesaian masalah. Algoritma *Naïve Bayes* pada dasarnya pernah digunakan pada penelitian sebelumnya, namun hanya membandingkan hasil akurasi dengan algoritma lainnya. Berbeda halnya dengan penelitian ini yang menggunakan algoritma *Naïve bayes* memanfaatkan *Correlation Based Featured Selection* (CSF) untuk mencari atribut terbaik dalam proses pengolahan data sehingga hasil akurasi yang didapatkan lebih maksimal.

3. Metodologi

3.1 Pengumpulan Data

Penelitian ini menggunakan data nilai akhir yang diperoleh dari Bagian Akademik Administrasi (BAA) dan data mata kuliah Kewarganegaraan seluruh program studi yang didapatkan melalui *platform OpenLearning*. Data yang digunakan mulai dari tahun akademik 2020/2021 dan 2021/2022 semester ganjil dengan jumlah data sebanyak 2663. Data nilai akhir yang diperoleh terdiri dari 5 atribut diantaranya nomor induk mahasiswa (NIM), nama, nilai akhir, bobot dan simbol. Sedangkan data *OpenLearning* terdiri dari 17 atribut diantaranya yaitu *profile*

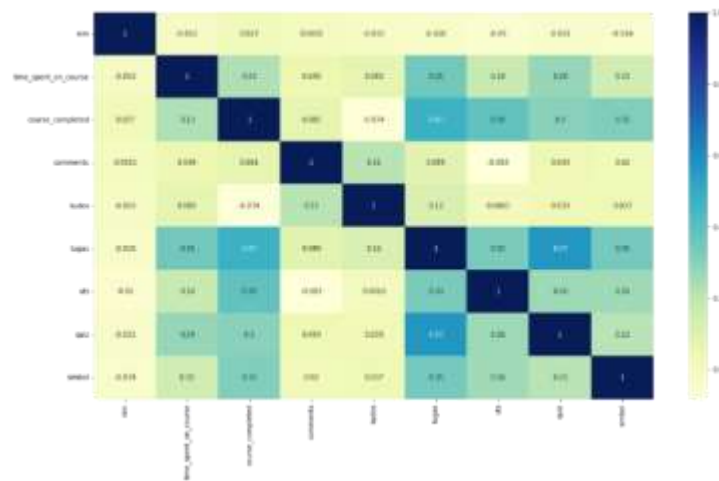
name, learner name, learner email, enrolment ID, institution membership id, enrolment date, completion date, time spent on course, progress, %course completed, certificate, comments, kudos, enrolment cost, nilai penugasan, nilai Ujian Tengah Semester (UTS) dan nilai quiz.

3.2 Data Preparation

a. Data Selection

Tahap seleksi dilakukan setelah proses pengumpulan data, dimana data yang telah dikumpulkan akan diseleksi atribut yang digunakan. Proses seleksi terbagi menjadi dua tahapan yaitu menyeleksi atribut yang tidak diperlukan dan melihat nilai korelasi pada tiap-tiap atribut. Pada tahap pertama menyeleksi atribut yang tidak diperlukan dengan menghapus atribut data BAA UMKT yaitu Nama, Nilai akhir, dan Bobot. Sedangkan atribut yang dihapus pada data *OpenLearning* yaitu *Profile name, Learner name, Learner email, Enrolment ID, Institution Membership ID, Enrolement Date, Progress, dan Certificate ID.*

Tahap kedua yaitu melihat nilai korelasi antar atribut dengan menggunakan *feature selection*. Adapun atribut yang akan digunakan untuk proses ini yaitu *nim, time spent on course, course completed, comments, kudos, nilai penugasan, nilai UTS, nilai quiz* dan nilai akhir terhadap target atau simbol.



Gambar 1. Pencarian Nilai Korelasi

Pencarian nilai korelasi pada setiap atribut terhadap target ini menggunakan *python* yang dapat dilihat pada gambar 2. Dari penentuan korelasi diatas diperoleh atribut yang digunakan yaitu *time spent on course = 0.15, course completed = 0.32, tugas = 0.35, uts = 0.26, quiz = 0.22.*

b. Data Integration

Tahap integrasi dilakukan untuk menggabungkan data nilai akhir dengan atribut simbol dan data *OpenLearning* berupa atribut *time spent on course, course completed, nilai penugasan, nilai UTS, dan nilai quiz.*

c. Data Transformation

Tahap transformasi dilakukan untuk merubah isi data atau tipe data sebelum masuk pada pemodelan. Atribut yang akan ditransformasi adalah *time spent on course* yang merupakan lama waktu mahasiswa menyelesaikan kursus. Adapun hasil transformasi dapat dilihat pada Tabel 1.

Tabel 1. Hasil Transformasi Atribut *Time Spent on Course*

<i>Time spent on course</i> (Sebelum di Transformasi)	<i>Time spent on course</i> (Setelah di Transformasi)
4 Hrs 56 Mins	296
5 Hrs 45 Mins	345
17 Hrs 47 Mins	1067
:	:
4 Hrs 42 Mins	228

Kemudian atribut yang ditransformasi selanjutnya adalah simbol, simbol ini akan digunakan sebagai kelas target pada pemodelan. Hasil transformasi atribut simbol dapat dilihat pada Tabel 2.

Tabel 2. Hasil Transformasi Atribut Simbol

Simbol (Sebelum di Transformasi)	Simbol (Setelah di Transformasi)
A	Baik
AB	Baik
B	Baik
BC	Buruk
C	Buruk
D	Buruk
E	Buruk
T	Buruk

d. Data Reduction dan Data Cleaning

Tahapan ini dilakukan untuk menghindari dataset yang tidak seimbang dengan mengurangi data mayoritas. Data mayoritas merupakan data kelas BAIK yang berjumlah 2536 sehingga dikurangi sejumlah data minoritas BURUK yang berjumlah 127 data. Jadi total data kelas baik berjumlah 127 yang diambil secara random. Setelah itu dilakukan tahapan *data cleaning* atau pembersihan data terhadap data yang tidak konsisten. Oleh karena itu pada tahapan reduksi ini dilakukan proses pembersihan data yang tidak konsisten. Sehingga jumlah data yang digunakan pada penelitian ini adalah 178 data.

3.3 Algoritma Naïve Bayes

Algoritma *Naïve Bayes* merupakan metode yang menggunakan *probability* untuk membuat model prediksi klasifikasi yang memanfaatkan data mengenai kejadian masa lampau, model ini juga dapat menghitung *probability* suatu kejadian dan dapat berubah jika ada informasi pendukung tambahan yang disediakan [19]. Adapun langkah-langkah dalam penerapan *Naïve Bayes* adalah sebagai berikut [20]:

1. Membaca *Data Training*
2. Menghitung Jumlah dan Probabilitas
 - a. Bila terdapat data numerik, maka temukan nilai *mean* dan standar deviasi dari masing-masing parameter yang menggambarkan data angka. Berikut rumus yang digunakan untuk menghitung nilai *mean*:

$$\mu = \sum_{i=1}^n x_i \text{ atau } \mu = \frac{x_1+x_2+x_3+\dots+x_n}{n} \dots \dots \dots (1)$$

Penjelasan:

μ = rata-rata hitung (mean)

X_i = nilai sampel ke-i

n = jumlah sampel

Dan berikut rumus untuk menghitung nilai standar deviasi:

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \mu)^2}{n-1}} \dots \dots \dots (2)$$

Penjelasan:

σ = Standar devisi

X_i = nilai x ke -i

μ = rata-rata hitung

n = jumlah sampel

- b. Bila data berupa kategori, maka menghitung nilai probabilitas untuk fitur *data testing* yang sama dengan kelas yang sama lalu dibagi dengan data pada kelas tersebut
3. Nilai Probabilitas Fitur Setiap Kelas
Mengetahui nilai probabilitas pada setiap fitur dengan menghitung jumlah data yang sesuai dari kategori yang sama lalu dibagi dengan jumlah data pada kategori tersebut.
4. Nilai Distribusi *Gaussian*
Selanjutnya menghitung nilai probabilitas untuk fitur *data testing* yang memiliki data numerik.

$$P = (X_i = x_i | Y=y_j) = \frac{1}{\sqrt{2\pi\sigma_{ij}}} \times e^{-\frac{(x_i-\mu_{ij})^2}{2\sigma_{ij}^2}} \dots\dots\dots (3)$$

5. Probabilitas Akhir Setiap Kelas
Menghitung probabilitas akhir untuk setiap kelas dengan memasukkan semua data nilai *distribusi gaussian* yang ada kedalam satu kelas yang sama.

6. Probabilitas Akhir
Probabilitas akhir diperoleh dari perhitungan nilai probabilitas akhir kelas kedalam rumus *Naïve Bayes Classifier*.

$$P (Kelas |X) = P(Kelas) \times P(X) \dots\dots\dots (4)$$

Kemudian setelah mendapatkan probabilitas akhir, langkah terakhir adalah melakukan normalisasi dengan cara membagikan nilai probabilitas satu kategori dengan jumlah nilai semua kategori.

$$P (Kelas) = \frac{P(Kelas|X)}{P(X|Kelas)} + P(X|Kelas) \dots\dots\dots(5)$$

3.4 Evaluasi Confusion Matrix

Evaluasi data dilakukan setelah proses prediksi pada algoritma *Naïve Bayes*. Adapun metode evaluasi yang digunakan adalah *confusion matrix* dengan melihat nilai akurasi. Tabel 3 merupakan tabel *confusion matrix*.

Tabel 3. Evaluasi *Confusion Matrix*

Class	Actual = Yes	Actual =No
Predicted = Yes	TP	FP
Predicted = No	FN	TN

- a. TP (*True Positive*) berarti data aktual bernilai positif yang diprediksi positif
- b. TN (*True Negative*) berarti data aktual bernilai negatif yang diprediksi negatif
- c. FP (*False Positive*) merupakan data aktual yang bernilai negatif namun diprediksi positif
- d. FN (*False Negative*) merupakan data aktual yang bernilai positif namun diprediksi negatif.

Kemudian untuk mencari nilai akurasi dapat dihitung dengan melihat perbandingan jumlah item yang diprediksi positif pada seluruh prediksi yang dilakukan. Berikut rumus perhitungannya.

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} \dots\dots\dots(6)$$

4. Hasil dan Pembahasan

4.1 Modelling

Modelling pada penelitian ini menggunakan algoritma *Naïve Bayes* dengan pembagian uji data 70:30, 80:20 dan 90:10.

a. Dataset

Tabel 4. Dataset

No	Time Spent on Course	Course Completed	Tugas	Quiz	UTS	Simbol
1	349	98	80,8	91	80	Baik
2	321	81	58,4	54	32	Buruk
3	435	98	80,5	50	84	Baik
4	252	63	39,4	0	66	Buruk
5	131	29	27	0	62	Buruk
:	:	:	:	:	:	:
174	517	99	83,9	64	75	Baik
175	280	99	85	86	63	Baik
176	683	98	80,2	78	87	Baik
177	443	98	80,1	86	87	Baik
178	304	81	80,2	70	60	Baik

b. Pengujian Data

Pengujian data pada penelitian ini menggunakan tiga pembagian data yaitu pembagian data 70:30 yang terdapat data training sebanyak 124 data dan data testing sebanyak 54 data.

Kemudian pada pembagian data 80:20 terdapat data training sebanyak 142 data dan data testing sebanyak 36 data. Selanjutnya pada pembagian data 90:10 terdapat data training sebanyak 160 data dan data testing 18 data.

c. Perhitungan Algoritma *Naïve Bayes*

Pada perhitungan algoritma ini hanya mengambil salah satu pembagian data yaitu 80% *data training* dan 20% *data testing*. Dari 6 atribut yang telah ditetapkan seluruh datanya berjenis numerik, oleh karena itu menggunakan perhitungan *Gaussian Naive Bayes*.

Tabel 5. *Data Training 80:20*

No	Time Spent on Course	Course Completed	Tugas	Quiz	UTS	Target
1	443	82	82,6	52	66	Baik
2	1347	82	86,3	86	80	Baik
3	880	98	76,7	86	71	Baik
4	280	99	85	86	63	Baik
5	73	77	0	0	0	Buruk
:	:	:	:	:	:	:
142	328	81	80	66	81	Baik

Tabel 6. *Data Testing 80:20*

No	Time Spent on Course	Course Completed	Tugas	Quiz	UTS	Target
1	143	89	19	0	0	Buruk
2	561	97	75,5	78	66	Baik
3	187	85	41,9	22	0	Buruk
4	151	86	22,2	62	53	Buruk
5	799	97	35	86	60	Baik
:	:	:	:	:	:	:
36	120	83	27	14	0	Buruk

1) Mencari Nilai Probabilitas

Langkah pertama yang dilakukan adalah menghitung nilai probabilitas dari kelas target "baik" dan "buruk". Dimana jumlah kelas "baik" sebanyak 73 data dan jumlah kelas "buruk" sebanyak 69 data.

$$P(\text{Target} = \text{Baik}) = \frac{73}{142} = 0,514$$

$$P(\text{Target} = \text{Buruk}) = \frac{69}{142} = 0,485$$

2) Mencari Nilai Mean

Langkah kedua adalah menghitung nilai mean pada data testing dari masing-masing atribut.

Atribut "Time Spent on Course"

$$\mu_{\text{target}} = \text{Baik} = \frac{443 + 1347 + 880 + 280 + 644 + \dots + 328}{73} = 504,219$$

$$\mu_{\text{target}} = \text{Buruk} = \frac{73 + 85 + 52 + 278 + 38 + \dots + 128}{69} = 192,233$$

Atribut "Course Completed"

$$\mu_{\text{target}} = \text{Baik} = \frac{82 + 82 + 98 + 99 + 99 + \dots + 91}{73} = 90,383$$

$$\mu_{\text{target}} = \text{Buruk} = \frac{77 + 77 + 16 + 57 + 50 + \dots + 88}{69} = 57,188$$

Atribut "Tugas"

$$\mu_{\text{target}} = \text{Baik} = \frac{82,6 + 86,3 + 76,7 + 85 + 68,7 + \dots + 80}{73} = 73,602$$

$$\mu_{\text{target}} = \text{Buruk} = \frac{0 + 19 + 6 + 45,3 + 19,4 \dots + 0}{69} = 19,486$$

Atribut "UTS"

$$\mu_{\text{target}} = \text{Baik} = \frac{52 + 86 + 86 + 86 + 82 + \dots + 30}{73} = 72,561$$

$$\mu_{target} = Buruk = \frac{0 + 30 + 0 + 54 + 0 + \dots + 62}{69} = 23,362$$

Atribut "Quiz"

$$\mu_{target} = Baik = \frac{66 + 80 + 71 + 63 + 83 + \dots + 81}{73} = 69,767$$

$$\mu_{target} = Buruk = \frac{0 + 0 + 0 + 0 + 0 + \dots + 0}{69} = 3$$

3) Mencari Nilai Standar Deviasi

Setelah menghitung nilai *mean* pada setiap atribut, langkah selanjutnya adalah menghitung nilai standar deviasi pada setiap atribut.

Atribut "Time Spent on Course"

$$\sigma_{target} = Baik = \sqrt{\frac{\sum(443 - 504,219)^2 + \dots + (328 - 504,219)^2}{73 - 1}} = 328,454$$

$$\sigma_{target} = Buruk = \sqrt{\frac{\sum(73 - 161,710)^2 + \dots + (128 - 161,710)^2}{69 - 1}} = 145,956$$

Atribut "Course Completed"

$$\sigma_{target} = Baik = \sqrt{\frac{\sum(82 - 90,383)^2 + \dots + (81 - 90,383)^2}{73 - 1}} = 8,572$$

$$\sigma_{target} = Buruk = \sqrt{\frac{\sum(77 - 57,188)^2 + \dots + (88 - 57,188)^2}{69 - 1}} = 30,483$$

Atribut "Tugas"

$$\sigma_{target} = Baik = \sqrt{\frac{\sum(82,6 - 73,602)^2 + \dots + (80 - 73,602)^2}{73 - 1}} = 13,975$$

$$\sigma_{target} = Buruk = \sqrt{\frac{\sum(0 - 19,486)^2 + \dots + (0 - 19,486)^2}{69 - 1}} = 19,844$$

Atribut "UTS"

$$\sigma_{target} = Baik = \sqrt{\frac{\sum(52 - 72,561)^2 + \dots + (66 - 72,561)^2}{73 - 1}} = 13,491$$

$$\sigma_{target} = Buruk = \sqrt{\frac{\sum(0 - 23,362)^2 + \dots + (30 - 23,362)^2}{69 - 1}} = 22,991$$

Atribut "Quiz"

$$\sigma_{target} = Baik = \sqrt{\frac{\sum(62 - 69,767)^2 + \dots + (0 - 69,767)^2}{73 - 1}} = 16,379$$

$$\sigma_{target} = Buruk = \sqrt{\frac{\sum(66 - 3)^2 + \dots + (81 - 3)^2}{103 - 1}} = 11,473$$

4) Mencari Nilai Distribusi Gaussian

Langkah selanjutnya adalah menghitung nilai *gaussian* pada data *testing* di setiap atribut. Data *testing* yang dihitung pada penelitian ini mengambil satu sampel yaitu data *testing* 1 (satu).

Atribut "Time Spent on Course" = 143

$$g_{time} = Baik = \frac{1}{\sqrt{2 \times 3,14 \times 328,454}} \exp\left\{-\frac{(143-504,219)^2}{2(328,454)^2}\right\} = 0,01202$$

$$g_{time} = Buruk = \frac{1}{\sqrt{2 \times 3,14 \times 145,956}} \exp \frac{-(143-161,710)^2}{2(145,956)^2} = 0,03275$$

Atribut "Course Completed" = 89

$$g_{course} = Baik = \frac{1}{\sqrt{2 \times 3,14 \times 8,572}} \exp \frac{-(89-90,383)^2}{2(8,572)^2} = 0,13452$$

$$g_{course} = Buruk = \frac{1}{\sqrt{2 \times 3,14 \times 30,483}} \exp \frac{-(89-57,188)^2}{2(30,483)^2} = 0,04192$$

Atribut "Tugas" = 19

$$g_{tugas} = Baik = \frac{1}{\sqrt{2 \times 3,14 \times 13,975}} \exp \frac{-(62-73,602)^2}{2(13,975)^2} = 5,1732E - 05$$

$$g_{tugas} = Buruk = \frac{1}{\sqrt{2 \times 3,14 \times 18,469}} \exp \frac{-(19-19,486)^2}{2(18,469)^2} = 0,09282$$

Atribut "UTS" = 0

$$g_{UTS} = Baik = \frac{1}{\sqrt{2 \times 3,14 \times 13,491}} \exp \frac{-(0-72,561)^2}{2(13,491)^2} = 5,6878E - 08$$

$$g_{UTS} = Buruk = \frac{1}{\sqrt{2 \times 3,14 \times 22,991}} \exp \frac{-(0-23,362)^2}{2(22,991)^2} = 0,04966$$

Atribut "Quiz" = 0

$$g_{quiz} = Baik = \frac{1}{\sqrt{2 \times 3,14 \times 16,379}} \exp \frac{-(0-69,767)^2}{2(16,379)^2} = 1,1332E - 05$$

$$g_{quiz} = Buruk = \frac{1}{\sqrt{2 \times 3,14 \times 11,473}} \exp \frac{-(0-3)^2}{2(11,473)^2} = 0,11384$$

5) Probabilitas Akhir Setiap Kelas

Dalam menentukan nilai akhir langkah pertama yang harus dilakukan adalah mengalikan semua atribut pada kelas baik dan kelas buruk.

$$g_{Baik} = 0,01202 \times 0,13452 \times 5,1732E - 05 \times 5,6878E - 08 \times 1,1332E - 05$$

$$= 5,39474E - 20$$

$$g_{Buruk} = 0,03275 \times 0,04192 \times 0,09282 \times 0,04966 \times 0,11384$$

$$= 7,20849E - 07$$

6) Probabilitas Akhir

Setelah mengalikan seluruh kelas "BAIK" dan kelas "BURUK", langkah selanjutnya adalah mengalikan kelas "BAIK" dan "BURUK" pada probabilitas yang sudah dihitung sebelumnya.

$$P(Target = Baik|X) = 5,39474E - 20 \times 0,514 = 2,77335E - 07$$

$$P(Target = Buruk|X) = 7,20849E - 07 \times 0,485 = 3,50272E - 07$$

Selanjutnya yaitu membandingkan nilai antara kedua kelas tersebut. Diperoleh hasil nilai kelas "BURUK" lebih besar dibandingkan nilai "BAIK". Oleh karena itu hasil prediksi dari data testing ini adalah "BURUK".

Tabel 7. Hasil Klasifikasi

No	Time Spent on Course	Course Completed	Tugas	Quiz	UTS	Aktual	Prediksi
1	143	89	19.0	0	0	Buruk	Buruk
2	561	97	75.5	78	66	Baik	Baik
3	187	85	41.9	22	0	Buruk	Buruk
4	151	86	22.2	62	53	Buruk	Baik
5	799	97	35	86	60	Baik	Baik
:	:	:	:	:	:	:	:
36	120	83	27.0	14	0	Buruk	Buruk

Berdasarkan hasil klasifikasi yang terdapat pada Tabel 7, dapat dilihat bahwa penentuan target kinerja baik maupun buruk dapat dilihat dari nilai probabilitas akhir dari kelas baik dan kelas buruk. Jika nilai probabilitas akhir kelas baik lebih besar dari kelas buruk maka hasil yang diperoleh adalah kelas baik, dan bila nilai probabilitas akhir kelas buruk lebih besar dari kelas baik maka hasil yang diperoleh adalah kelas buruk.

4.2 Evaluation

Evaluasi data dilakukan setelah proses prediksi pada algoritma *Naïve Bayes*. Adapun metode evaluasi yang digunakan adalah *confusion matrix* yang akan menghasilkan nilai *accuracy*. Pada Gambar 2 dibawah ini merupakan proses pencarian *confusion matrix*.

Gambar 2. Proses *Confusion Matrix* 70:30, 80:20 dan 90:10

Confusion Matrix 70:30			
No	Aktual	Prediksi	Confusion Matrix
1	BURUK	BURUK	TN
2	BAIK	BAIK	TP
3	BURUK	BURUK	TN
4	BURUK	BAIK	FP
5	BAIK	BAIK	TP
6	BAIK	BAIK	TP
7	BURUK	BURUK	TN
8	BAIK	BAIK	TP
9	BAIK	BAIK	TP
10	BAIK	BAIK	TP
11	BAIK	BAIK	TP
12	BAIK	BAIK	TP
13	BURUK	BURUK	TN
14	BAIK	BAIK	TP
15	BAIK	BAIK	TP
16	BURUK	BURUK	TN
17	BURUK	BURUK	TN
18	BURUK	BURUK	TN
19	BURUK	BURUK	TN
20	BURUK	BURUK	TN
21	BURUK	BURUK	TN
22	BURUK	BURUK	TN
23	BURUK	BURUK	TN
24	BAIK	BAIK	TP
25	BURUK	BURUK	TN
26	BAIK	BAIK	TP
27	BAIK	BAIK	TP
28	BURUK	BURUK	TN
29	BAIK	BAIK	TP
30	BAIK	BAIK	TP
31	BURUK	BURUK	TN
32	BURUK	BURUK	TN
33	BAIK	BAIK	TP
34	BURUK	BURUK	TN
35	BURUK	BURUK	TN
36	BURUK	BURUK	TN
37	BAIK	BAIK	TP
38	BAIK	BAIK	TP
39	BAIK	BAIK	TP
40	BAIK	BAIK	TP
41	BURUK	BURUK	TN
42	BURUK	BURUK	TN
43	BAIK	BAIK	TP
44	BAIK	BAIK	TP
45	BAIK	BAIK	TP
46	BURUK	BURUK	TN
47	BAIK	BAIK	TP
48	BURUK	BURUK	TN
49	BAIK	BAIK	TP
50	BURUK	BURUK	TN
51	BAIK	BAIK	TP
52	BAIK	BAIK	TP
53	BURUK	BURUK	TN
54	BAIK	BAIK	TP

Confusion Matrix 80:20			
No	Aktual	Prediksi	Confusion Matrix
1	Buruk	Buruk	TN
2	Baik	Baik	TP
3	Buruk	Buruk	TN
4	Buruk	Baik	FP
5	Baik	Baik	TP
6	Baik	Baik	TP
7	Buruk	Buruk	TN
8	Baik	Baik	TP
9	Baik	Baik	TP
10	Baik	Baik	TP
11	Baik	Baik	TP
12	Baik	Baik	TP
13	Buruk	Buruk	TN
14	Baik	Baik	TP
15	Baik	Baik	TP
16	Buruk	Buruk	TN
17	Buruk	Buruk	TN
18	Buruk	Buruk	TN
19	Buruk	Buruk	TN
20	Buruk	Buruk	TN
21	Buruk	Buruk	TN
22	Buruk	Buruk	TN
23	Buruk	Buruk	TN
24	Baik	Baik	TP
25	Buruk	Buruk	TN
26	Baik	Baik	TP
27	Baik	Baik	TP
28	Buruk	Buruk	TN
29	Baik	Baik	TP
30	Baik	Baik	TP
31	Buruk	Buruk	TN
32	Buruk	Buruk	TN
33	Baik	Baik	TP
34	Buruk	Buruk	TN
35	Buruk	Buruk	TN
36	Buruk	Buruk	TN

Confusion Matrix 90:10			
No	Aktual	Prediksi	Confusion Matrix
1	BURUK	BURUK	TN
2	BAIK	BAIK	TP
3	BURUK	BURUK	TN
4	BURUK	BAIK	FP
5	BAIK	BAIK	TP
6	BAIK	BAIK	TP
7	BURUK	BURUK	TN
8	BAIK	BAIK	TP
9	BAIK	BAIK	TP
10	BAIK	BAIK	TP
11	BAIK	BAIK	TP
12	BAIK	BAIK	TP
13	BURUK	BURUK	TN
14	BAIK	BAIK	TP
15	BAIK	BAIK	TP
16	BURUK	BURUK	TN
17	BURUK	BURUK	TN
18	BURUK	BURUK	TN

a. Perhitungan *confusion matrix* pengujian data 70:30

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} = \frac{28+25}{28+25+0+1} = \frac{53}{54} = 98.14\%$$

b. Perhitungan *confusion matrix* pengujian data 80:20

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} = \frac{16+19}{16+19+0+1} = \frac{35}{36} = 97.22\%$$

c. Perhitungan *confusion matrix* pengujian data 90:10

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} = \frac{10+7}{14+7+0+1} = \frac{21}{26} = 94.44\%$$

Tabel 8. Hasil Pengujian Data

Pengujian Data	TP	TN	FP	FN	Hasil Akurasi
70:30	28	25	1	0	98,14%
80:20	16	19	1	0	97,22%
90:10	10	7	1	0	94,44%

Pada Tabel 8 diatas merupakan hasil pengujian data algoritma *Naïve Bayes*. Dimana TP (*True Positive*) berarti data aktual bernilai positif yang diprediksi positif, sedangkan TN (*True Negative*) berarti data aktual bernilai negatif yang diprediksi negatif. Kemudian FP (*False Positive*) merupakan data aktual yang bernilai negatif namun diprediksi positif, sedangkan FN (*False Negative*) merupakan data aktual yang bernilai positif namun diprediksi negatif.

Dari pengujian data tersebut diperoleh hasil evaluasi dan hasil *accuracy* algoritma *Naive Bayes*. Pada pembagian data 70:30 menghasilkan *true positive* 28 data, *true negative* 25 data, *false positive* 0 data dan *false negative* 1 data yang menghasilkan akurasi sebesar 98,14%. Kemudian pada pembagian data 80:20 memiliki *true positive* 16 data, *true negative* 19 data, *false positive* 0 data dan *false negative* 1 data yang menghasilkan akurasi sebesar 97,22%. Dan yang terakhir pada pembagian data 90:10 memiliki *true positive* 10 data, *true negative* 7 data, *false positive* 0 data dan *false negative* 1 data yang menghasilkan akurasi 94,44%. Dari hasil pengujian data diperoleh pembagian data 70:30 memiliki akurasi terbaik yaitu 98,14%.

Hasil akurasi yang diperoleh pada penelitian ini lebih tinggi dibandingkan penelitian [17] yang hanya menghasilkan akurasi 93,94%. Pada penelitian tersebut menggunakan 330 data, berbeda dengan penelitian ini yang menggunakan 178 data. Selain jumlah data, atribut yang digunakan juga berbeda yaitu: NIPD, jenis kelamin, IPS1 sampai IPS 13, dan jumlah SKS, sehingga mempengaruhi hasil akurasi. Penelitian ini berhasil memprediksi kinerja mahasiswa dalam pembelajaran daring menggunakan *platform LMS Open Learning* dengan hasil akurasi yang lebih baik dari penelitian sebelumnya.

5. Simpulan

Dengan menerapkan algoritma *Naïve Bayes* dapat memprediksi apakah mahasiswa tersebut memiliki kinerja yang baik atau buruk. Hal ini dapat berguna dalam memberikan informasi serta masukan untuk pihak perguruan tinggi dalam kebijakan kedepannya. Tingkat keberhasilan dari algoritma *Naïve Bayes* ditentukan oleh indikator *time spent on course*, *course completed*, *nilai penugasan*, *nilai UTS* dan *nilai quiz* yang menghasilkan akurasi terbaik sebesar 98,14%. Rekomendasi untuk penelitian selanjutnya yaitu data *training* yang digunakan pada penelitian ini dijadikan sebagai rujukan untuk pembuatan aplikasi sistem prediksi kinerja mahasiswa.

Daftar Referensi

- [1] WHO, "Pertanyaan dan Jawaban Terkait Coronavirus," 2022. <https://www.who.int/indonesia/news/novel-coronavirus/qa/qa-for-public> (diakses Feb 12, 2022).
- [2] Kemendikbud, "Kemendikbud Terbitkan Pedoman Penyelenggaraan Belajar dari Rumah," 2020. <https://www.kemdikbud.go.id/main/blog/2020/05/kemendikbud-terbitkan-pedoman-penyelenggaraan-belajar-dari-rumah> (diakses Feb 12, 2022).
- [3] Y. I. Kurniawan, "Perbandingan Algoritma Naive Bayes dan C.45 dalam Klasifikasi Data Mining," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 4, hal. 455, 2018, doi: 10.25126/jtiik.201854803.
- [4] D. Anggreani, Herman, dan W. Astuti, "Kinerja Metode Naive Bayes dalam Prediksi Lama Studi Mahasiswa Fakultas Ilmu Komputer," *Semin. Nas. Ilmu Komput. dan Teknol. Inf.*, vol. 3, no. 2, hal. 107–111, 2018, [Daring]. Tersedia pada: <http://e-journals.unmul.ac.id/index.php/SAKTI/article/view/1843>.
- [5] Y. Apridiansyah, N. D. M. Veronika, dan E. D. Putra, "Prediksi Kelulusan Mahasiswa Fakultas Teknik Informatika Universitas Muhammadiyah Bengkulu Menggunakan Metode Naive Bayes," *JSAI (Journal Sci. Appl. Informatics)*, vol. 4, no. 2, hal. 236–247, 2021, doi: 10.36085/jsai.v4i2.1701.
- [6] A. F. Firdaus, R. Saedudin, R. Andeswari, dan U. Telkom, "Implementasi Metode Klasifikasi Naive Bayes Implementation of Naive Bayes Classification Method in Predicting," vol. 8, no. 5, hal. 9274–9279, 2021.
- [7] P.S.C. Moonallika, K.Q. Fredlina, and I.K. Sudiarmika "Penerapan Data Mining Untuk Memprediksi Kelulusan Mahasiswa Menggunakan Algoritma Naive Bayes Classifier (Studi Kasus STMIK Primakara)". *Progresif: Jurnal Ilmiah Komputer*, vol. 16, no. 1, pp. 47-56, 2020.
- [8] N. Nosieli, S. Sriyanto, dan F. Maylani, "Perbandingan Teknik Data Mining Untuk Prediksi Penjualan Pada UMKM Gerabah," *Pros. Semin. Nas. Darmajaya*, vol. 1, no. 0, hal. 72–86, 2021.
- [9] W. Ananda, M. Safii, dan M. Fauzan, "Prediksi Jumlah Hasil Panen Sawit Menggunakan Algoritma Naive Bayes," *TIN Terap. Inform. Nusant. Vol*, vol. 1, no. 10, hal. 513–519, 2021.
- [10] N. Maulidah, R. Supriyadi, D. Y. Utami, F. N. Hasan, A. Fauzi, dan A. Christian, "Prediksi Penyakit Diabetes Melitus Menggunakan Metode Support Vector Machine dan Naive Bayes," *Indones. J. Softw. Eng.*, vol. 7, no. 1, hal. 63–68, 2021, doi: 10.31294/ijse.v7i1.10279.
- [11] Rumini dan A. Nasruddin, "Prediksi Awal Penyakit Diabetes Mellitus Menggunakan Algoritma Naive Bayes," *J. ICT Inf. Commun. Technol.*, vol. 20, no. 2, hal. 246–253, 2021, doi: 10.36054/jict-ikmi.v20i2.376.
- [12] H. Nalatissifa, W. Gata, S. Diantika, dan K. Nisa, "Perbandingan Kinerja Algoritma Klasifikasi Naive Bayes, Support Vector Machine (SVM), dan Random Forest untuk Prediksi Ketidakhadiran di Tempat Kerja," *J. Inform. Univ. Pamulang*, vol. 5, no. 4, hal. 578, 2021, doi: 10.32493/informatika.v5i4.7575.
- [13] R. Alfiani, "Prediksi Kepuasan Customer Terhadap Performance Terapis Baby Massage Menggunakan Algoritma Naive Bayes," vol. 9, no. 1, hal. 83–92, 2022.
- [14] T. Taufiq and Y. Yudihartanti, "Penerapan Theorema Bayes Pada Penilaian Kelayakan Angkutan Kota". *Jutisi: Jurnal Ilmiah Teknik Informatika dan Sistem Informasi*, vol. 10, no. 1, pp. 111-122, 2021.
- [15] A. Widiyono, "Efektifitas Perkuliahan Daring (Online) pada Mahasiswa PGSD di Saat Pandemi Covid 19," *J. Pendidik.*, vol. 8, no. 2, hal. 169–177, 2020, doi: 10.36232/pendidikan.v8i2.458.
- [16] N. A. Akhmad, "Analisis Respon Mahasiswa Terhadap Perkuliahan Daring di Prodi Biologi STKIP PI Makassar," *J. Pendidik. Fis. dan Ter.*, vol. 3, no. 2, hal. 62–65, 2021.
- [17] A. I. Nurhidayat dan D. Fatrianto, "Prediksi Kinerja Akademik Mahasiswa Menggunakan Machine Learning dengan Sequential Minimal Optimization untuk Pengelola Program Studi," vol. 05, hal. 84–91, 2021.
- [18] E. Sabna, "Data Mining Dengan 2 (Dua) Model Klasifikasi Untuk Prediksi Kinerja Mahasiswa," *J. Ilmu Komput.*, vol. 10, no. 2, hal. 128–131, 2021, doi: 10.33060/jik/2021/vol10.iss2.229.

- [19] D. Kurniawan, *Pengenalan Machine Learning Python*. Jakarta: PT Alex Media Komputindo, 2020.
- [20] M. Guntur, J. Santony, dan Y. Yuhandri, "Prediksi Harga Emas dengan Menggunakan Metode Naïve Bayes dalam Investasi untuk Meminimalisasi Resiko," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 2, no. 1, hal. 354–360, 2018, doi: 10.29207/resti.v2i1.276.