



### PENERAPAN DATA MINING UNTUK PREDIKSI KELULUSAN SISWA MENGGUNAKAN ALGORITMA NAÏVE BAYES PADA SMK GARUDA

Endah Yaodah Kodratillah<sup>1</sup>, Daririn<sup>2</sup>, Candra Naya<sup>3</sup>

Program Studi Teknik Informatika Fakultas Teknik Universitas Pelita Bangsa

<sup>1</sup>endah.yk@pelitabangsa.ac.id, <sup>3</sup>candranaya@pelitabangsa.ac.id

#### Abstraksi

Kelulusan siswa adalah suatu rangkaian tahapan proses yang harus dinilai oleh setiap siswa, diantaranya harus menyelesaikan sejumlah pembelajaran yang telah ditentukan, dan harus memenuhi syarat-syarat dan ketentuan yang telah ditetapkan oleh pihak sekolah, karena kelulusan menjadi acuan agar bisa melanjutkan pendidikan ke jenjang yang lebih tinggi maupun bekerja di dunia usaha serta dunia industri. Maka dari itu, penelitian ini akan memanfaatkan data tentang kelulusan siswa dengan mengolahnya menggunakan Data Mining untuk mendapatkan informasi berupa prediksi kelulusan siswa. Algoritma yang akan digunakan adalah algoritma Naive Bayes. Atribut yang digunakan dalam memprediksi kelulusan siswa ini adalah Nilai Praktik, US (Ujian Sekolah) dan Perilaku Siswa. Pengujian Data Mining dengan menggunakan algoritma Naive Bayes menghasilkan accuracy 76.25%, precision 93.33%, recall 78.87% dan nilai AUC sebesar 0.751. Dapat disimpulkan bahwa, algoritma Naive Bayes memiliki akurasi yang sangat baik pada penelitian ini.

**Keywords:** *Data mining, Classification, C4.5 Algoritma, Naive Bayes*

#### Abstract

*Student graduation is a series of process stages that must be assessed by each student, including having to complete a predetermined number of lessons, and must meet the terms and conditions set by the school, because graduation is a reference so that they can continue their education to a higher level. and work in the business world and the industrial world. Therefore, this study will utilize data about student graduation by processing it using Data Mining to obtain information in the form of predictions of student graduation. The algorithm that will be used is the Naive Bayes algorithm. The attributes used in predicting student graduation are Practice Score, US (School Examination) and Student Behavior. Data Mining testing using the Naive Bayes algorithm produces an accuracy of 76.25%, precision 93.33%, recall 78.87% and an AUC value of 0.751. It can be concluded that the Naive Bayes algorithm has very good accuracy in this study.*

**Keywords:** *Data mining, Classification, C4.5 Algoritma, Naive Bayes*

#### 1. Pendahuluan

Perkembangan ilmu pengetahuan dan teknologi informasi di segala bidang yang meningkat dengan begitu pesat, teknologi tersebut telah berperan untuk menciptakan kemajuan dan kemudahan pada setiap bidang kehidupan salah satunya dibidang pendidikan[1].

SMK Garuda Nusantara merupakan sekolah menengah kejuruan swasta yang berada di Cikarang Selatan, yang mempunyai visi berbeda untuk maju.

Seiring berkembangnya teknologi, maka kebutuhan akan informasi yang cepat, tepat dan akurat sangat dibutuhkan untuk menunjang pengambilan keputusan. Karena itulah diperlukan adanya suatu metode yang baik agar dapat mengkaji data yang sudah ada untuk dijadikan suatu keputusan yang tepat dan akurat sesuai dengan kebutuhan dalam proses prediksi kelulusan siswa.

Berdasarkan penelitian sebelumnya yang dilakukan, metodologi yang diusulkan dari preprocessing, pemilihan atribut, klasifikasi dan evaluasi kinerja.

Metode data mining yang dipilih karena dapat menghasilkan model dan kriteria yang mudah

dipresentasikan oleh klasifikasi pelatihan data dan pengujian dengan algoritma naïve bayes.

**2. Landasan Pemikiran**

Metodelogi penelitian merupakan suatu usaha dalam menemukan suatu hal yang baru berdasarkan metode yang ilmiah. Tahapan yang akan dibahas yaitu Metode klasifikasi *Naïve Bayes*.

**2.1. Metode Pengumpulan Data**

**a. Studi Literatur**

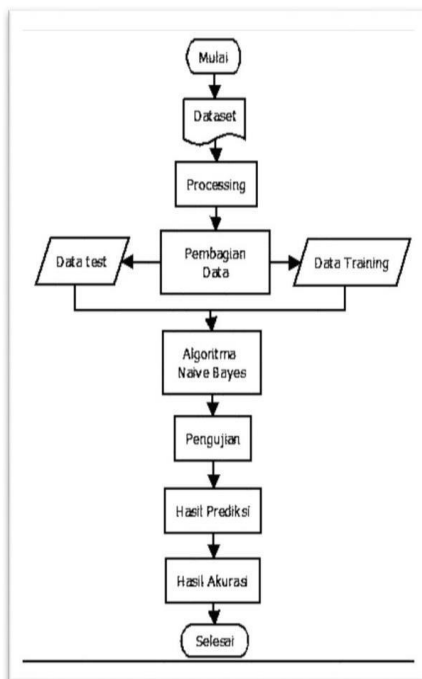
Pada tahapan ini, peneliti mempelajari beberapa referensi berupa buku jurnal yang berkaitan dengan penelitian. Topik yang akan dibahas yaitu Metode Klasifikasi *Naïve Bayes*.

**b. Pengumpulan Data**

Pengumpulan data merupakan bagian terpenting dalam sebuah penelitian. Ketersediaan data akan sangat menentukan dalam proses pengolahan dan Analisa. Pada tahap pengumpulan data, peneliti menggunakan data siswa SMK Garuda Nusantara dari tahun 2020 dengan data sebanyak 200 record.

Metode pengumpulan data yang dilakukan dalam penelitian ini adalah :

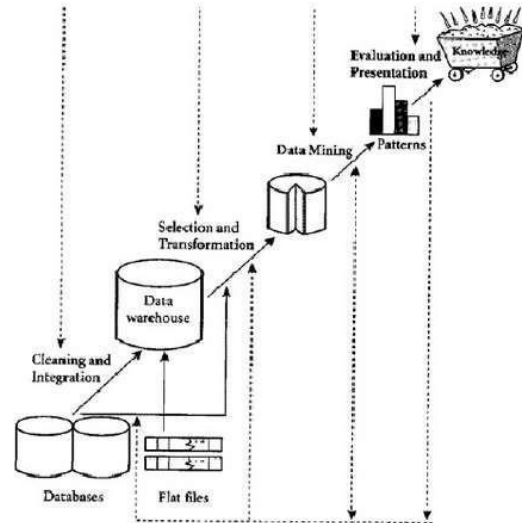
1. Studi Pustaka : Pengumpulan data yaitu berupa literatur buku, jurnal nasional maupun internasional. Kemudian mempelajari referensi lain yang berkaitan dengan masalah yang akan dibahas untuk dapat dijadikan acuan dalam penelitian ini.
2. Wawancara : Kegiatan melakukan tanya jawab dengan Kepala Sekolah secara langsung tentang keterkaitan penelitian.
3. Observasi : Pengamatan langsung maupun tidak langsung proses kegiatan yang ada dengan hasil mendapatkan data penunjang untuk penelitian



**Gambar 1. 1 Tahapan Penelitian**

**2.2. Data Mining**

*Data Mining* dikenal juga sebagai *knowledge discovery in database* (KDD) yaitu proses pencarian pengetahuan yang bermanfaat dari kumpulan data. Proses KDD bersifat interaktif dan iterative, meliputi sejumlah langkah dengan melibatkan pengguna dalam membuat keputusan dan dapat dilakukan pengulangan di antara dua buah langkah[2].



**Gambar 2. 2 Proses tahapan data mining**

**2.3. Klasifikasi**

Klasifikasi adalah salah satu pembelajaran yang paling umum di *data mining*. Klasifikasi didefinisikan sebagai bentuk analisis data untuk mengekstrak model yang akan digunakan untuk memprediksi label kelas. Kelas dalam klasifikasi merupakan atribut dalam satu set data paling unik yang merupakan variabel bebas dalam statistik. Klasifikasi data terdiri dari dua proses yaitu tahap pembelajaran dan tahap pengklasifikasian. Tahap pembelajaran merupakan tahapan dalam pembentukan model klasifikasi, sedangkan tahap pengklasifikasian merupakan tahapan penggunaan model klasifikasi untuk memprediksi label kelas dari suatu data[3].

**2.4. Prediksi**

Prediksi merupakan bagian awal dari sebuah proses pengambilan suatu keputusan. Sebelum memprediksi, harus diketahui terlebih dahulu persoalan yang sebenarnya dalam pengambilan keputusan. Prediksi adalah pemikiran terhadap besaran misalnya, permintaan terhadap satu atau beberapa produk pada periode yang akan datang[4].

Data masa lampau tersebut, secara sistematis digabungkan melalui metode tertentu dan diolah untuk keadaan pada masa yang akan datang. Tujuan dari peramalan adalah menjadi sebuah pengambil keputusan dan pembuat kebijakan memahami ketidakpastian dan resiko yang mungkin muncul dapat dipertimbangkan sewaktu membuat perencanaan. Dengan melakukan prediksi tersebut para perencana dan pengambil keputusan akan dapat mempertimbangkan alternative lain. Dalam kenyataannya, hasil dari prediksi tidak mutlak tepat, hal

ini dikarenakan keadaan ataupun kejadian dimasa depan tidak menentu. Meskipun demikian, jika semua faktor ditentukan dengan baik, maka hasil prediksi akan mendekati hasil sebenarnya[5].

**2.5. Naïve Bayes Classifier (NBC)**

*Simple Naïve Bayes Classifier* merupakan salah satu metode pengklasifikasian berpeluang sederhana yang berdasarkan pada penerapan Teorema Bayes dengan asumsi antar variabel penjelas saling bebas (independent). Algoritma ini memanfaatkan metode probabilitas di masa depan berdasarkan pengalaman di masa sebelumnya. Dua kelompok peneliti, satu oleh Pantel dan Lin, dan lain oleh *Microsoft Research* memperkenalkan metode statistik Bayesian ini pada teknologi antispam filter. Tetapi yang membuat algoritma *Bayesian filtering* ini populer adalah pendekatan yang dilakukan oleh Paul Graham[6].

Bentuk Umum dari *teorema bayes* :

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

Keterangan :

X : Data dengan kelas yang belum diketahui

H : Hipotesis data x merupakan suatu kelas spesifik

P(H|X) : Probabilitas hipotesis H berdasarkan kondisi X (*postreiorprobability*)

P(H) : Probabilitas H (*priorprobability*)

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

P(X) : Probabilitas X

Dapat dijelaskan dimana (P|HX) itu merupakan probabilitas dari H yang di dalam X atau Bahasa lain dari P(H|X) merupakan presentas dengan banyaknya H yang di dalam X, P(X|H) merupakan probabilitas X di dalam H, P(H) merupakan probabilitas dari H dan P(H) merupakan probabilitas prior dari X

$$P(X|Ci) = \prod_{k=1}^n P(Xk|Ci)$$

Berdasarkan persamaan tersebut maka kelas (label) dari data sampel X adalah kelas label yang memiliki :

$$P(X|Ci) \times P(Ci)$$

yang bernilai maksimum[7].

**3. Metode Penelitian**

Metode Pengumpulan data adalah cara atau Teknik yang dapat digunakan oleh peneliti untuk mengumpulkan data. Untuk memperoleh data yang akurat dalam laporan penelitian ini, maka penelitian ini menggunakan beberapa metode yang dilakukan sebagai berikut :

**3.1 Jenis Data**

Adapun jenis data yang digunakan dalam penelitian ini adalah sebagai berikut :

**1. Data Kualitatif**

Data yang disajikan dalam bentuk kata verbal bukan dalam bentuk angka. Yang termasuk data kualitatif dalam penelitian ini yaitu gambaran umum objek penelitian.

**2. Data Kuantitatif**

Data yang didapat diukur atau dihitung secara langsung, yang berupa informasi atau penjelasan yang dinyatakan dengan bilangan atau bentuk angka. Dalam hal ini data kuantitatif yang diperlukan adalah data yang digunakan sebagai pengujian berupa angka meliputi : jenis kelamin, nilai praktik, nilai us, nilai perilaku dalam bentuk angka sebagai acuan dalam penentuan parameter pada penelitian.

**3.2 Sumber Data**

Data yang digunakan dalam penelitian ini adalah data *training* dan data *testing*. Data tersebut merupakan dari data nilai siswa yang diperoleh dari pihak akademik SMK Garuda Nusantara. Di bawah ini sample data nilai siswa sebelum dilakukan pengolahan data.

Tabel 3. 1 Data Nilai Siswa

No.	NIS	Jenis Kelamin	Nama Siswa	Jurusan	Nilai Praktek	Nilai Ujian Sekolah	Perilaku Siswa
1	31016944	P	ALFINA FATIM SUNDARI	AK	A	A	BAIK
2	37610706	P	AMANI SUHAIMAH	AK	A	A	BAIK
3	39965437	P	ANIS SATUTOYIBAH	AK	A	A	BAIK
4	24984597	P	ANISSA OCTAVIANI	AK	A	A	BAIK
5	37610467	P	AULIA FAZA	AK	A	A	BAIK
6	34788715	P	AVILJA PINGKY NABILA	AK	A	A	KURANG BAIK
7	37169619	P	CICIH	AK	A	A	BAIK
8	38124146	P	DWI ASTUTI BAROKAH	AK	A	A	BAIK
9	20320168	P	FITRIANI	AK	A	A	BAIK
10	38381935	P	LIFYA AULIYANI	AK	A	A	BAIK
11	32905129	P	MUTIARA BAITUL ZANAH	AK	A	A	BAIK
12	34680276	P	NOPIASARI	AK	A	A	BAIK
13	22575993	P	OKTAFIANI	AK	A	A	BAIK
14	31016802	P	SHELLA NURMALA	AK	A	A	BAIK

(Sumber, SMK Garuda Nusantara 2020)

**3.3 Data yang akan digunakan**

Pada penelitian ini data yang digunakan adalah data nilai siswa tahun 2020 SMK Garuda Nusantara Cikarang Selatan. Data raw yang berisi atribut nama, jenis kelamin, nilai praktik, nilai ujian sekolah, nilai perilaku siswa dan keterangan. Kemudian data yang akan digali yaitu nilai dari setiap kriteria tersebut menjadi patokan untuk prediksi kelulusan siswa.

Dengan mengamati persoalan diatas *data mining* dengan metode *naïve bayes* sangat tepat digunakan untuk menghasilkan *Knowledge* kelompok prioritas. Adapun data yang akan dijadikan sebagai data *training* yaitu sebanyak 120 record data, data *training* digunakan untuk melakukan prose pelatihan terhadap dataset untuk menghasilkan suatu klasifikasi. Sedangkan data *testing* digunakan untuk melakukan pengujian terhadap proses klasifikasi yang digunakan algoritma terhadap dataset.

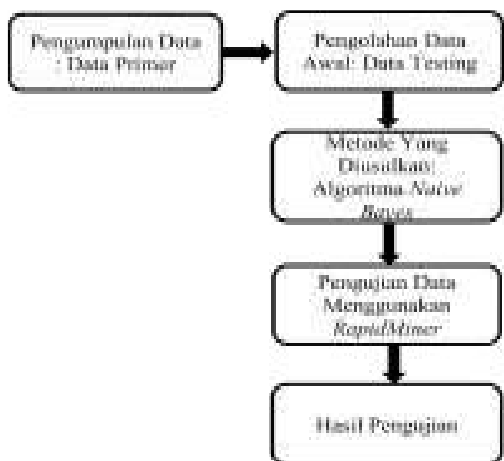
**3.4 Pengolahan Data Awal**

Pada tahap ini menjelaskan tentang tahap awal data mining. Data yang telah didapatkan akan diolah ke format yang dibutuhkan, pengelompokkan dan penentuan atribut, variabel. Dalam melakukan pengolahan data awal, akan di *training* atau di hitung dan dilakukan beberapa tahapan agar didapatkan data yang digunakan untuk tahap selanjutnya.

**3.5 Pembagian Dataset**

Untuk memperoleh dataset dengan jumlah atribut yang lebih sedikit tetapi bersifat informative. Maka dataset dibagi menjadi data *training* dan data *testing*. Data *training* diperoleh dari pembagian sumber data 200 *record*, dimana 120 data dialokasikan untuk *training* data dan sisanya 80 *record* untuk *testing* data. *Training* data sebagai penentuan prediksi kelulusan siswa “LULUS” atau “TIDAK LULUS”, data tersebut dilakukan *cleaning* data yaitu dengan menambahkan nilai pada atribut yang kosong dengan nilai rata-rata serta menghapus atribut apa saja yang tidak diperlukan karena tidak berpengaruh terhadap kelulusan siswa.

### 3.6 Metode yang digunakan



Gambar 3. 1 tahapan penelitian

Metode yang digunakan dalam pengumpulan data-data yang dibutuhkan pada penelitian adalah sebagai berikut;

#### 1. Metode Observasi

Observasi adalah suatu pengamatan sistematis yang dilakukan secara langsung terhadap objek di tempat penelitian. Observasi bertujuan untuk mendapatkan data tentang suatu masalah, sehingga diperoleh informasi sebagai alat pembuktian terhadap informasi yang diperoleh sebelumnya. Dalam arti yang luas observasi dilakukan dengan cara mengamati langsung lokasi yang menjadi sumber informasi memperoleh informasi yang dibutuhkan dalam penelitian ini. Beberapa informasi yang diperoleh dari hasil observasi adalah tempat, pelaku, kegiatan, objek, perbuatan, kejadian atau peristiwa waktu dan perasaan.

#### 2. Metode Wawancara

Metode wawancara merupakan metode pengumpulan data dengan cara mengajukan beberapa pertanyaan langsung kepada narasumber. Wawancara dilakukan untuk mengetahui secara langsung oleh yang bersangkutan. Dalam kegiatan ini diajukan pernyataan lisan dalam usaha untuk melengkapi data yang akan diperoleh. Wawancara dilakukan di SMK Garuda Nusantara.

### 3. Metode Studi Pustaka

Metode studi pustaka ini dilakukan dengan cara mempelajari jurnal, buku-buku atau studi literatur terdahulu yang berkaitan dengan kelulusan siswa, *data mining* dan *naïve bayes classifier*. Setelah data dikumpulkan maka dilakukan pembelajaran mengenai data yang akan digunakan. pembelajaran data diharapkan dapat membiasakan peneliti dengan data-data yang sudah dikumpulkan serta dapat menemukan wawasan awal mengenai informasi apa saja yang bisa didapatkan didalamnya.

## 4. Pembahasan

### 4.1 Dataset

Pada penelitian ini jumlah dataset yang digunakan sebanyak 200 *record*. Data Nilai siswa diambil dari tahun 2020. Jumlah data yang diperoleh dari proses tersebut sebanyak 200 *record* yang terdiri dari atribut “LULUS” sebanyak 176 *record*, dan “TIDAK LULUS” sebanyak 24 *record*.

#### 1. Perhitungan Naïve Bayes

Sumber : Peneliti, 2021

Tahap 3  $P(X|Ci) * P(Ci)$

$$P(X | \text{Kelulusan} = \text{Lulus}) = P(\text{Jenis Kelamin} = \text{Laki-laki} | \text{Kelulusan} = \text{Lulus}) * P(\text{Jurusan} = \text{TKR} | \text{Kelulusan} = \text{Lulus}) * P(\text{Nilai Praktik} = \text{A} | \text{Kelulusan} = \text{Lulus}) * P(\text{Nilai US} = \text{A} | \text{Kelulusan} = \text{Lulus}) * P(\text{Perilaku Siswa} = \text{Baik} | \text{Kelulusan} = \text{Lulus}) = 0,54285714 * 0,45714286 * 0,93333333 * 0,97142857 * 0,95238095 = 0,2142870086$$

$$P(X | \text{Kelulusan} = \text{Tidak Lulus}) = P(\text{Jenis Kelamin} = \text{Laki-laki} | \text{Kelulusan} = \text{Tidak Lulus}) * P(\text{Jurusan} = \text{TKR} | \text{Kelulusan} = \text{Tidak Lulus}) * P(\text{Nilai Praktik} = \text{A} | \text{Kelulusan} = \text{Tidak Lulus}) * P(\text{Nilai US} = \text{A} | \text{Kelulusan} = \text{Tidak Lulus}) * P(\text{Perilaku Siswa} = \text{Baik} | \text{Kelulusan} = \text{Tidak Lulus}) = 0,73333333 * 0,8 * 0,46666667 * 0,2 * 0,33333333 = 0,0182518517$$

Hasil probabilitas prior berdasarkan persamaan menghasilkan nilai untuk kelas lulus sebesar 0,02142870086 dan untuk kelas tidak lulus sebesar 0,0182518517.

#### 4.2 Implementasi Klasifikasi Naïve Bayes

Setelah melakukan perhitungan manual selanjutnya dengan menggunakan Rapid Miner. Pengujian ini untuk mengetahui nilai akurasi dan prediksi data. Pengujian dilakukan dengan algoritma naïve bayes dan data yang digunakan sebanyak 120 *record* untuk data *training* dan 80 *record* untuk data *testing*. Berikut tahap pengujiannya :

#### 1. Penerapan Rapid Miner dengan Metode Naïve Bayes

Rapid Miner Studio 9.9 adalah salah satu tool untuk melakukan prediksi dan Analisa *data mining*.

#### 2. Proses Import Data

Pada tahapan ini data nilai siswa yang sudah diklasifikasi, *integration* dan *selection* diimport

kedalam tool Rapid Miner 9.9 yaitu terdiri dari data training dan data testing, data training ini digunakan untuk latihan membentuk sebuah model classifier untuk prediksi kelas data baru yang belum pernah ada.

Row No.	Keterangan	Jenis Kelamin	Nama Siswa	Jurusan	Nilai Praktek	Nilai Ujian S...	Perilaku Sis...
1	LULUS	P	ALFIRA FATI...	AK	A	A	BAIK
2	LULUS	P	AMANI SUHAI...	AK	A	A	BAIK
3	LULUS	P	ANIS SAIYUT...	AK	A	A	BAIK
4	TIDAK LULUS	P	ANISSA OCT...	AK	A	A	BAIK
5	LULUS	P	AULIA FAZA	AK	A	A	BAIK
6	LULUS	P	AVILIA PRINGK...	AK	A	A	KURANG BAIK
7	LULUS	P	CICIH	AK	B	A	BAIK
8	LULUS	P	DWI ASTUTI...	AK	A	A	BAIK
9	LULUS	P	FITRIANI	AK	A	A	BAIK
10	LULUS	P	LYA AULIYANI	AK	A	A	BAIK
11	LULUS	P	MUTHARA BAL...	AK	A	A	BAIK
12	LULUS	P	NOPTIASARI	AK	B	A	BAIK

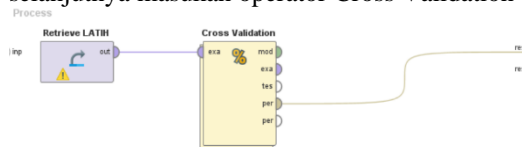
Gambar 4. 1 Import data training Data testing ini digunakan untuk mengukur sejauh mana prediction berhasil melakukan prediksi kualitas produk dengan benar.

Row No.	Keterangan	Jenis Kelamin	Nama Siswa	Jurusan	Nilai Praktek	Nilai Ujian S...	Perilaku Sis...
1	LULUS	L	YOYON	TKR	A	A	BAIK
2	LULUS	L	TIOHO	TKR	B	B	BAIK
3	LULUS	L	TARA	TKR	B	B	BAIK
4	LULUS	L	NUR ADI	TKR	A	A	BAIK
5	LULUS	L	TUHADI	TKR	A	A	BAIK
6	LULUS	L	ENTASLI	TKR	A	A	BAIK
7	LULUS	P	MUNA	TKR	A	A	BAIK
8	LULUS	L	BENI	TKJ	A	B	BAIK
9	LULUS	L	SUMNI	TKJ	A	B	BAIK

Gambar 4. 2 Import data testing

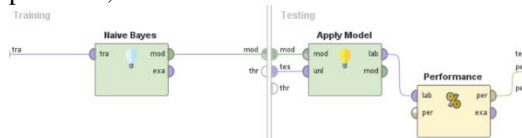
3. Proses Pengujian

Proses klasifikasi pada Rapid Miner Menggunakan metode Naïve Bayes untuk membandingkan data training dan data testing yang sudah diketahui rule-ruhnya sebelumnya. Berikut langkah-langkahnya Pada tampilan proses masukan operator read excel input masing-masing data training dan data testing, selanjutnya masukan operator Cross Validation



Gambar 4. 3 Cross Validation

Pada pemodelan cross validation digunakan untuk memberikan pelatihan di dalamnya terdapat 3 bagian yaitu, pada bagian training digunakan algoritma naïve bayes, untuk bagian testing menggunakan fitur Apply Model untuk mengaplikasikan model pada data data testing dan untuk menampilkan confusion table yang digunakan untuk menampilkan hasil accuracy, precision, recall.



Gambar 4. 4 Desain Pengujian Metode Naïve Bayes

4.3 Analisa Hasil Pengujian

Analisa dilakukan secara deskriptif dalam rangka memperoleh gambaran mengenai prediksi kelulusan siswa. Berdasarkan hasil penelitian yang telah diperoleh, disertai dengan data yang ada, maka selanjutnya peneliti akan melakukan analisis terhadap hasil penelitian yang telah dipaparkan

sebelumnya, untuk mengetahui dari analisis data tersebut dapat diketahui dengan menghitung presentase dari data analisis.

1. Performance Vector

Performance Vector: accuracy: 94.17% +/- 6.86% (micro average: 94.17%)

Confusion Matrix:

True: LULUS TIDAK LULUS

LULUS: 101 3

TIDAK LULUS: 4 12

precision: 76.67% +/- 34.43% (micro average: 75.00%) (positive class: TIDAK LULUS)

Confusion Matrix:

True: LULUS TIDAK LULUS

LULUS: 101 3

TIDAK LULUS: 4 12

recall: 80.00% +/- 34.96% (micro average: 80.00%) (positive class: TIDAK LULUS)

Confusion Matrix:

True: LULUS TIDAK LULUS

LULUS: 101 3

TIDAK LULUS: 4 12

AUC (optimistic): 0.940 +/- 0.171 (micro average: 0.940) (positive class: TIDAK LULUS)

AUC: 0.931 +/- 0.200 (micro average: 0.931) (positive class: TIDAK LULUS)

AUC (pessimistic): 0.922 +/- 0.229 (micro average: 0.922) (positive class: TIDAK LULUS)

2. Accuracy

Perhitungan dari algoritma naïve Bayes, akurasi dilakukan dengan cara jumlah TP + TN dibagi jumlah total data yang diuji. Dengan mengetahui jumlah data yang diklasifikasikan secara benar maka dapat diketahui akurasi hasil prediksi yaitu 94.17%.

	true LULUS	true TIDAK LULUS	class precision
pred LULUS	56	4	93.33%
pred TIDAK LULUS	15	5	25.00%
class recall	78.87%	55.56%	

Gambar 4. 5 accuracy

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} = \frac{56 + 5}{56 + 4 + 15 + 5} = 76.25\%$$

3. Precision

Precision adalah jumlah data yang true positive ( jumlah data positif yang dikenali secara benar sebagai positif ) dibagi dengan jumlah data dikenali sebagai positif. Dari hasil pengujian nilai precision yaitu 93.33 % untuk class Layak dan 25.00% untuk class Tidak Layak.

	true LULUS	true TIDAK LULUS	class precision
pred LULUS	56	4	93.33%
pred TIDAK LULUS	15	5	25.00%
class recall	78.87%	55.56%	

Gambar 4. 6 Precision

$$Precision = \frac{Tp}{Tp + FP} = \frac{56}{56 + 4} = 93.33\%$$

4. Recall

Recall adalah jumlah data yang true positive dibagi dengan jumlah data yang sebenarnya positive (true positive – true negative). Untuk nilai Recall yaitu 78.87 % untuk class Layak dan 55.56% untuk

recall: 55.56% (positive class: TIDAK LULUS)			
	true LULUS	true TIDAK LULUS	class precision
pred LULUS	56	4	93.33%
pred TIDAK LULUS	15	5	25.00%
class recall	78.87%	55.56%	

Gambar 4. 7 Recall

$$Recall = \frac{TP}{TP+FN} = \frac{56}{56 + 15} = 78.87\%$$

Setelah dilakukan pengujian dengan menggunakan algoritma naive bayes maka didapatkan nilai rata – rata akurasi sebesar 76.25 % dengan hasil ini maka Penggunaan klasifikasi naive bayes untuk prediksi kelulusan siswa di SMK Garuda Nusantara cukup bagus dan dapat dipakai sebagai acuan untuk penelitian selanjutnya.

## 5 Penutup

Berdasarkan hasil penelitian yang telah dilakukan mengenai prediksi kelulusan siswa dengan metode naive bayes maka dapat diambil beberapa kesimpulan sebagai berikut :

Penerapan metode naive bayes memanfaatkan data training untuk menghasilkan probabilitas setiap kriteria untuk class yang berbeda, sehingga nilai-nilai probabilitas dari kriteria tersebut dapat dioptimalkan untuk menentukan prediksi kelulusan siswa “LULUS “dan “TIDAK LULUS“ dengan cepat dan efisien berdasarkan klasifikasi yang dilakukan oleh metode naive bayes.

Dari hasil pengujian yang telah dilakukan maka hasil penelitian dengan metode naive bayes pada Rapid Miner maka didapatkan sebuah hasil bahwa nilai akurasi adalah 76.25%. Sehingga hasil ini adalah termasuk akurasi yang sangat baik. Dengan begitu metode ini bisa diterapkan dalam memprediksi kelulusan siswa.

## Daftar Pustaka

[1] M. F. Arif Siswandi, “Implementasi Data

Mining Dengan Metode Klasifikasi Naive Bayes Unruk Prediksi Stok bahan jadi,” *Teknologi Pelita Bangsa*, vol. 10, p. 1, 2019.

[2] R. P. T. Selvia Lorena Br Ginting, “Teknik Data Mining Menggunakan Metode Bayes Classifier Untuk Optimasi Pencarian Pada Aplikasi Perpustakaan,” *Univ. Pas. Bandung*, pp. 1–3, 2020.

[3] Dewi Sartika, “Perbandingan Algoritma Klasifikasi Naive Bayes, Nearest Neighbour, dan Decision Tree pada Studi Kasus Pengambilan Keputusan Pemilihan Pola Pakaian,” *Jatiji*, p. 153, 2017.

[4] E. H. Ermanto, “Penerapan Data Mining Menggunakan Metode Algoritma Naive Bayes Untuk Menentukan Kelayakan Kredit Rumah Bersubsidi,” *Teknologi Pelita Bangsa*, vol. 10, no. 4, p. 120, 2020.

[5] Weneda Mitha Harani, “Prediksi Penjualan Barang Pada Alfamart rembang Menggunakan Exponential Smoothing,” *Media Infotama*, p. 133, 2016.

[6] M.Ridwan Effendi, “Akurasi Data Mining Untuk Menghasilkan pola Kelulusan Mahasiswa dengan Metode Naive Bayes,” *Univ. Muhammad Husni Thamrin Jakarta*, 2017.

[7] R. H. Muhtajudin Danny, “Penerapan Algoritma Naive Bayes Dalam Mendignosa penyakit ISPA,” *Teknologi Pelita Bangsa*, vol. 10, no. 3, p. 3, 2020.