

**PEMANFAATAN ALGORITMA NAIVE BAYES UNTUK KLASIFIKASI STATUS
ALUMNI SMK BUSTANUL ULUM AL-GHAZALI
WULUHAN JEMBER**

¹Richy Kusuma Midiarso (1610651035), ²Reni Umilasari, S.Pd, M.Si

Jurusan Teknik Informatika

Fakultas Teknik

Universitas Muhammadiyah Jember

E-mail:¹richykusumamidiarso@gmail.com

ABSTRAK

Data Mining merupakan suatu proses penggalian data atau penyaringan data dengan memanfaatkan kumpulan data dengan ukuran yang cukup besar melalui serangkaian proses untuk mendapatkan informasi yang berharga dari data tersebut. Salah satu teknik data mining yaitu metode Naive Bayes. Naive Bayes merupakan sebuah pengklasifikasian probabilistik sederhana yang menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari dataset yang diberikan.

SMK Bustanul Ulum Al-ghazali merupakan Sekolah Menengah Kejuruan dibawah yayasan Pondok Pesantren Bustanul Ulum Al-ghazali. Selain memiliki SMK, yayasan tersebut juga memiliki perguruan tinggi serta banyak menjalin kerjasama dengan para pengusaha dan beberapa perusahaan. Dengan begitu SMK Bustanul Ulum Al-ghazali melalui guru Bimbingan dan Konseling membutuhkan suatu sistem yang mampu mengklasifikasikan para alumni untuk melanjutkan ke jenjang pendidikan yang lebih tinggi (kuliah) atau jenjang karir (bekerja) secara tepat, konsisten, objektif serta hasil evaluasi dari klasifikasi tersebut dapat diketahui keakuratannya. Untuk mengatasi permasalahan tersebut digunakan algoritma Naive Bayes untuk klasifikasi status alumni SMK Bustanul Ulum Al-ghazali.

Dalam penerapan metode Naive Bayes terhadap data training dari alumni sebanyak 262 record menghasilkan sebanyak 145 alumni masuk dalam kelas bekerja dan 117 alumni masuk dalam kelas kuliah dengan akurasi sebesar 92,37%, presisi sebesar 85,47% dan recall sebesar 97,09%.

Kata Kunci: Data Mining, Algoritma Naive Bayes, SMK Bustanul Ulum Al-ghazali.

I. PENDAHULUAN

Data Mining merupakan suatu proses penggalian data atau penyaringan data dengan memanfaatkan kumpulan data dengan ukuran yang cukup besar melalui serangkaian proses untuk mendapatkan informasi yang berharga dari data tersebut [1]. Salah satu teknik data mining yaitu metode Naive Bayes. Naive Bayes merupakan sebuah pengklasifikasian probabilistik sederhana yang menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari dataset yang diberikan [2].

SMK Bustanul Ulum Al-ghazali merupakan Sekolah Menengah Kejuruan dibawah yayasan Pondok Pesantren Bustanul Ulum Al-ghazali. Selain memiliki SMK, yayasan tersebut juga memiliki perguruan tinggi serta banyak menjalin kerjasama dengan para pengusaha dan beberapa perusahaan. Dengan begitu SMK Bustanul Ulum Al-ghazali membutuhkan suatu sistem yang mampu mengklasifikasikan para alumni untuk melanjutkan ke jenjang pendidikan yang lebih tinggi (kuliah) atau jenjang karir (bekerja) secara tepat, konsisten, objektif serta hasil evaluasi dari klasifikasi tersebut dapat diketahui keakuratannya. Proses klasifikasi tersebut memanfaatkan data alumni yang sebelumnya hanya tersimpan tanpa diolah lebih lanjut, selain itu hasil dari klasifikasi dapat dijadikan sebagai landasan untuk pengambilan kebijakan-kebijakan demi meningkatkan mutu para alumni.

Dengan menerapkan metode Naive Bayes dimana data tahun 2015 hingga tahun 2018 dapat dimanfaatkan secara maksimal sebagai data *training* dan data testing. Data *training* dan data testing dari alumni SMK Bustanul Ulum al-ghozali dapat diolah dan menghasilkan klasifikasi berupa status alumni secara tepat, konsisten, objektif serta hasil evaluasi dari klasifikasi tersebut dapat diketahui keakuratannya. Seluruh proses hingga hasil dari pengolahan akan disajikan dalam bentuk aplikasi berbasis website agar mudah diakses dan digunakan.

Tujuan dari penelitian ini adalah mengetahui hasil penerapan dan mengetahui nilai evaluasi kinerja metode Naive Bayes untuk klasifikasi status alumni pada SMK Bustanul Ulum Al Ghazali.

II. TINJAUAN PUSTAKA

2.1 Data Mining

Data mining adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Teknik-teknik metode-metode atau algoritma dalam data mining sangat bervariasi. Pemilihan metode atau algoritma yang tepat sangat bergantung pada tujuan dan proses Knowledge Discovery in Database (KDD) secara keseluruhan[3].

2.2 Klasifikasi Data Mining

Klasifikasi adalah salah satu bentuk dari teknik atau metode data mining yang termasuk dalam kategori predictive mining yaitu suatu

teknik yang dapat digunakan untuk meramalkan atau memprediksi kecenderungan data di masa depan. Proses yang terjadi dalam klasifikasi adalah proses penggolongan data ke dalam variabel target atau variabel tujuan dengan membangun sebuah model penyelesaian dengan memperhatikan atribut yang paling berpengaruh. Klasifikasi masuk ke dalam supervised induction, dimana pengujian yang memanfaatkan kumpulan pengujian dari record dan atribut yang terklasifikasi untuk menentukan output dan kelas tambahan. Salah satu contoh algoritmanya adalah Naive Bayes.

2.3 Naive Bayes

Algoritma Naive Bayes merupakan salah satu algoritma yang terdapat pada teknik klasifikasi. Naive Bayes merupakan pengklasifikasian dengan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi peluang di masa depan berdasarkan pengalaman dimasa sebelumnya sehingga dikenal sebagai Teorema Bayes. Teorema tersebut dikombinasikan dengan Naive dimana diasumsikan kondisi antar atribut saling bebas. Klasifikasi Naive Bayes diasumsikan bahwa ada atau tidak ciri tertentu dari sebuah kelas tidak ada hubungannya dengan ciri dari kelas lainnya. Persamaan dari teorema Bayes adalah:

$$P(H | X) = \frac{P(X | H) \cdot P(H)}{P(X)} \quad (2.1)$$

Keterangan:

X : Data dengan class yang belum diketahui

H : Hipotesis data X merupakan suatu class spesifik

$P(H | X)$: Probabilitas hipotesis H berdasar kondisi X (posteriori probability)

$P(H)$: Probabilitas hipotesis H (prior probability)

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

$P(X)$: Probabilitas X

2.4 Confusion Matrix

Confusion Matrix adalah suatu metode yang digunakan untuk melakukan perhitungan akurasi pada konsep data mining. Presisi atau confidence adalah proporsi kasus yang diprediksi positif yang juga positif benar pada data yang sebenarnya. Recall atau sensitivity adalah proporsi kasus positif yang sebenarnya yang diprediksi positif secara benar [4].

Tabel 2.1 Confusion Matrix

Aktual	Classified as	
	+	-
+	True positives (A)	False negatives (B)
-	False positives (C)	True negatives (D)

Perhitungan akurasi dengan tabel confusion matrix adalah sebagai berikut:

$$\text{Akurasi} = \frac{A+D}{A+B+C+D} \quad (2.2)$$

$$\text{Presisi} = \frac{A}{C+A} \quad (2.3)$$

$$\text{Recall} = \frac{A}{A+D} \quad (2.4)$$

2.5 SMK Bustanul Ulum Al-Ghazali

SMK Bustanul Ulum Al-Ghazali merupakan SMK swasta yang berdiri sejak tahun 1993 dibawah naungan yayasan pondok pesantren Bustanul Ulum Al-Ghazali yang terletak di desa Dukuh Dempok kecamatan Wuluhan Kabupaten Jember.

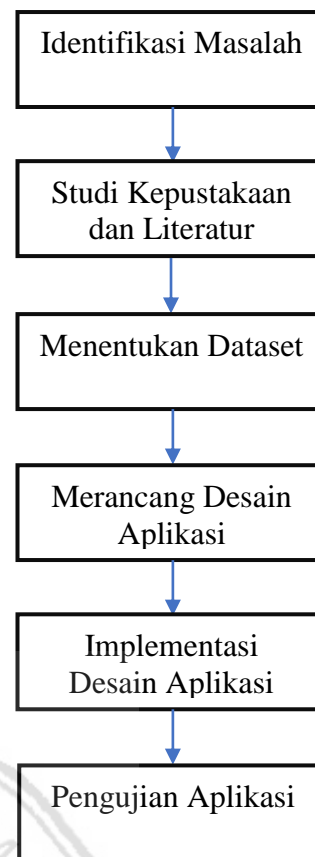
SMK Bustanul Ulum Al-Ghazali adalah SMK dengan jurusan Akuntansi dan memiliki kelas sebanyak dua kelas. Sejak tahun 2014 pemerintah membuat kebijakan dengan menyelenggarakan Ujian Nasional Berbasis Komputer (UNBK) dengan peserta baik dari SMA maupun SMK namun belum menyeluruh, sehingga untuk mempersiapkan UNBK tersebut, SMK Bustanul Ulum Al-Ghazali mulai menerapkan sistem administrasi yang terkomputerisasi (berbasis komputer).

Sistem Administrasi yang berbasis komputer pada SMK Bustanul Ulum Al-Ghazali menghasilkan data-data siswa yaitu berupa biodata siswa, data-data mengenai orang tua siswa yang bersangkutan hingga biodata Alumni mulai tahun 2015 hingga tahun 2018.

III. METODE PENELITIAN

3.1 Rancangan Penelitian

Agar penelitian dapat berjalan dengan lancar dan terstruktur, maka penulis membuat rancangan penelitian, rancangan penelitian tersebut adalah identifikasi masalah, studi kepustakaan dan literatur, menentukan dataset, merancang desain aplikasi, implementasi desain aplikasi dan pengujian aplikasi.



Gambar 3.1 Diagram Rancangan Penelitian

3.2 Penerapan Metode Naive Bayes

Pada penelitian ini menggunakan total data sebanyak 292 record, 262 record untuk data training dan 30 record digunakan sebagai data testing. Data set yang digunakan juga telah dilakukan penghilangan noise, yaitu menghilangkan atribut-atribut yang tidak digunakan dalam perhitungan klasifikasi Naive Bayes, contohnya adalah nama orang tua, alamat dan nama lembaga atau perusahaan selain itu juga dilakukan seleksi data yaitu melakukan eliminasi terhadap data-data yang tidak lengkap sesuai atribut-atribut yang dibutuhkan agar hasil dari klasifikasi Naive Bayes memiliki akurasi yang tinggi.

Atribut-atribut yang digunakan untuk mempertimbangkan status alumni untuk kuliah

atau bekerja adalah jenis kelamin, predikat UN, prestasi akademik, pendidikan orang tua dan pekerjaan orang tua. Sedangkan class yang dijadikan sebagai C1 adalah kuliah, dan C2 adalah bekerja.

Tabel 3.2 Data Training Alumni

NO	Nama	NIS	Jenis Kelamin	Predikat UN	Prestasi Akademik	Pendidikan Orang Tua	Pekerjaan Orang Tua	Status Alumni
1	SITI HAMIDAH	600019	P	Dibawah Rata-rata	Tidak	SMA	Petani	Bekerja
2	FITRIYATUR ROHANAH	600021	P	Dibawah Rata-rata	Tidak	SMA	Petani	Bekerja
3	Anugrah novita saputri	600022	P	Diatas Rata-rata	Tidak	Sarjana	PNS	Bekerja
4	Novi Nilasari	600023	P	Dibawah Rata-rata	Tidak	SMA	Wiraswasta	Kuliah
5	Marni lestari	600025	P	Dibawah Rata-rata	Tidak	SMA	Petani	Bekerja
6	INDRA PUJI LESTARI	600037	P	Dibawah Rata-rata	Tidak	Sarjana	PNS	Bekerja
7	MOH. ZAINULARIFIN	600041	L	Diatas Rata-rata	Tidak	Tidak tamat SMA	Petani	Bekerja
8	Amelia Alfiani	600049	P	Dibawah Rata-rata	Tidak	Sarjana	Wiraswasta	Kuliah
9	Ramadhan Kunlawan	600069	P	Diatas Rata-rata	Tidak	SMA	Wiraswasta	Kuliah
10	Atika Febri Waluyani	600070	P	Dibawah Rata-rata	Tidak	Tidak tamat SMA	Wiraswasta	Bekerja
11	Inggit kuswardani	600072	P	Diatas Rata-rata	Ada	Tidak tamat SMA	Wiraswasta	Bekerja
12	Munika Heny Marantika	600074	P	Dibawah Rata-rata	Tidak	Tidak tamat SMA	Petani	Bekerja
13	Panji Saiful Rohman	600076	L	Diatas Rata-rata	Ada	Sarjana	PNS	Kuliah
14	LUTFI DWI AGUSTIN	600077	P	Dibawah Rata-rata	Tidak	SMA	Petani	Bekerja
15	RISKI RIDHO	600078	L	Diatas Rata-rata	Tidak	SMA	Wiraswasta	Kuliah
16	Ramadhan Utama Putra	600079	L	Diatas Rata-rata	Tidak	SMA	PNS	Kuliah
17	ROSIDATUL UMAMAH	600083	P	Diatas Rata-rata	Tidak	Sarjana	Wiraswasta	Kuliah
18	Sriningsih	600084	P	Dibawah Rata-rata	Tidak	SMA	Petani	Bekerja
...
262	ELMY FIRDA MUVIDAH	601015	P	Diatas Rata-rata	Tidak	SMA	Wiraswasta	Kuliah

3.2 Tahap-tahap Perhitungan Naive Bayes

1) Baca Data Training

Data training yang disiapkan adalah data yang sudah dihilangkan atribut-atribut yang tidak digunakan seperti alamat, nama orang tua dan nama lembaga atau perusahaan sehingga data menjadi seperti yang tersaji pada tabel 3.2.

2) Hitung Jumlah kelas

Terdapat dua class yaitu C1 dan C2. Hitung jumlah class berdasarkan atribut yaitu seperti yang ditunjukkan pada Tabel 3.3. Kemudian mencari perbandingan class dengan jumlah kasus menggunakan persamaan rumus sebagai berikut:

$$- (C1) = \frac{103}{262} = 0,39312977 \quad (3.6)$$

$$- P(C2) = \frac{159}{262} = 0,60687023$$

Tabel 3.3 Jumlah Kelas

Atribut	Jumlah Kasus	Kuliah		Bekerja	
		C1	C2	C1	C2
Total	262	103	159		
Jenis Kelamin	L	120	38	82	
	P	142	65	77	
Predikat UN	Diatas Rata-rata	121	49	72	
	Dibawah Rata-rata	141	54	87	
Prestasi Akademik	Ada	17	1	16	
	Tidak	245	102	143	
Pendidikan Orang Tua	Sarjana	70	51	19	
	SMA	126	50	76	
	Tidak Tamat SMA	66	2	64	
Pekerjaan Orang Tua	PNS	46	30	16	
	Wiraswasta	86	72	14	
	Petani	130	1	129	

Tabel 3.4 Perbandingan Class Dengan Jumlah Kasus

Atribut	Jumlah Kasus	Kuliah		Bekerja		P	
		C1	C2	C1	C2	C1	C2
Total	262	103	159	0,39312977	0,60687023		

3) Hitung jumlah kasus yang sama dengan class yang sama

Untuk P(C1) dan P(C2) telah dihitung pada langkah sebelumnya, selanjutnya hitung $P(X | C_i)$ seperti berikut:

$$- P(\text{jenis Kelamin} = "L", \text{Status Alumni} = "Kuliah") = \frac{38}{103} = 0,3689$$

$$- P(\text{jenis Kelamin} = "L", \text{Status Alumni} = "Bekerja") = \frac{82}{159} = 0,5157$$

$$- P(\text{jenis Kelamin} = "P", \text{Status Alumni} = "Kuliah") = \frac{65}{103} = 0,6311$$

$$- P(\text{jenis Kelamin} = "P", \text{Status Alumni} = "Bekerja") = \frac{77}{159} = 0,4843$$

Berlaku untuk semua kriteria pada tiap atribut.

4) Kalikan semua nilai hasil sesuai dengan data yang dicari kelasnya

Selanjutnya adalah dengan cara mengalikan nilai yang telah dihasilkan dengan data yang akan dicari, dimisalkan disini pada Tabel 3.6 adalah record pertama yang akan dihitung.

Tabel 3.6 Record Yang Akan Dihitung

NO	Nama	Jenis Kelamin	Predikat UN	Prestasi Akademik	Pendidikan Orang Tua	Pekerjaan Orang Tua
1	SITI HAMIDAH	P	Dibawah Rata-rata	Tidak	SMA	Petani

Untuk C1

- Jenis Kelamin = P = 0,6311
- Predikat UN = Dibawah rata – rata = 0,5243
- Prestasi akademik = Tidak = 0,9903
- Pendidikan orang tua = SMA = 0,4854
- Pekerjaan orang tua = Petani = 0,0097

Sehingga bila dikalikan adalah 0,00154282

a. Untuk C2

- Jenis Kelamin = P = 0,4843
- Predikat UN = Dibawah rata – rata = 0,5472
- Prestasi akademik = Tidak = 0,8994
- Pendidikan orang tua = SMA = 0,4780
- Pekerjaan orang tua = Petani = 0,8113

Sehingga bila dikalikan adalah 0,09243210

Langkah selanjutnya adalah menghitung P(C1) dan P(C2) dengan persamaan sebagai berikut:

$$P(C1) = C1 \times C1(Total) = 0,00154282 \times 0,39312977 = 0,0006065300$$

$$P(C2) = C2 \times C2(Total) = 0,09243210 \times 0,60687023 = 0,0560942889$$

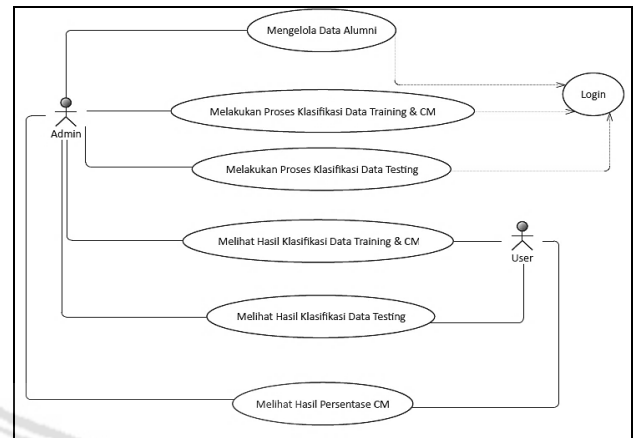
Hasil akhir diputuskan dengan membandingkan P(C1) dengan P(C2). Jika P(C1) lebih besar daripada P(C2) maka hasil akhir dari alumni yang bersangkutan adalah “Kuliah” sedangkan jika P(C2) lebih besar daripada P(C1) maka hasil akhir adalah “Bekerja”.

3.4 Confusion Matrix

Confusion Matrix bekerja dengan cara membandingkan hasil di lapangan dengan hasil dari suatu klasifikasi. Selanjutnya akan dihitung kriteria. Beberapa kriteria yang dimaksud adalah *True Positive*, *False Positive*, *True Negative* dan *False Negatif*. Jika masing-masing kriteria telah diketahui jumlahnya, selanjutnya adalah mencari tingkat kualitas hasil klasifikasi dengan menghitung akurasi, presisi dan *recall*.

3.5 Desain Aplikasi

Pada desain aplikasi ini, penulis menggunakan pendekatan desain aplikasi menggunakan obyek atau biasa disebut UML (*Unified Modeling Language*).



Gambar 3.2 Use Case Diagram

IV. IMPLEMENTASI DAN PENGUJIAN

Sesuai dengan metode kegiatan yang telah dirancang sebelumnya, tahap selanjutnya adalah implementasi aplikasi Naive Bayes dan dilanjutkan dengan pengujian aplikasi.

4.1 Implementasi Aplikasi

1) Baca Data Training

Gambar 4.18 Form Input

Pada tahap ini, membaca data training sekaligus memasukkan data training ke

dalam aplikasi menggunakan form seperti yang ditunjukkan pada Gambar 4.18.

2) Hitung Jumlah Kelas

Kemudian dilakukan proses perhitungan jumlah kelas, pada aplikasi yang ditandai dengan kotak hijau pada Gambar 4.19.

Bayesian Classification

Tabel Atribut

JUMLAH KASUS

Total Kasus	Kasus C1	Kasus C2	P (C1)	P (C2)
262	103	159	0.39312977	0.60687023

Showing 1 to 1 of 1 entries

ATRIBUT JENIS KELAMIN

ID	Jenis Kelamin	Jumlah Kasus	Kuliah (C1)	Bekerja (C2)
L	Laki-laki	120	38	82
P	Perempuan	142	65	77

Showing 1 to 2 of 2 entries

ATRIBUT PREDIKAT UN

ID	Predikat UN	Jumlah Kasus	Kuliah (C1)	Bekerja (C2)
atas_rata2	Diatas Rata-rata	121	49	72
bawah_rata2	Dibawah Rata-rata	141	54	87

Showing 1 to 2 of 2 entries

ATRIBUT PRESTASI AKADEMIK

ID	Prestasi Akademik	Jumlah Kasus	Kuliah (C1)	Bekerja (C2)
Ada	Ada	17	1	16
Tidak	Tidak	245	102	143

Showing 1 to 2 of 2 entries

ATRIBUT PENDIDIKAN TERAKHIR ORANG TUA

ID	Pendidikan Orang Tua	Jumlah Kasus	Kuliah (C1)	Bekerja (C2)
dibawah_sma	Tidak tamat SMA	66	2	64
sarjana	Sarjana	70	51	19
SMA	SMA	126	50	76

Showing 1 to 3 of 3 entries

ATRIBUT PEKERJAAN ORANG TUA

ID	Pekerjaan Orang Tua	Jumlah Kasus	Kuliah (C1)	Bekerja (C2)
Petani	Petani	130	1	129
PNS	PNS	46	30	16
Wiraswasta	Wiraswasta	86	72	14

Showing 1 to 3 of 3 entries

Gambar 4.19 Perhitungan Jumlah Kelas

3) Hitung Jumlah Kasus yang sama dengan Kelas yang Sama

Selanjutnya adalah menghitung jumlah kasus yang sama dengan kelas yang sama, hal ini berfungsi untuk mengetahui masing-masing probabilitas untuk setiap kategori pada atribut serta mengetahui masing-masing probabilitas dari jumlah kasus. Perhitungan probabilitas ini nanti yang akan dijadikan sebagai acuan untuk menentukan kelas pada

masing-masing record pada dataset. Hasil perhitungan seperti yang ditunjukkan pada Gambar 4.20.

Bayesian Classification

Tabel Atribut

JUMLAH KASUS

Total Kasus	Kasus C1	Kasus C2	P (C1)	P (C2)
262	103	159	0.39312977	0.60687023

Showing 1 to 1 of 1 entries

ATRIBUT JENIS KELAMIN

ID	Jenis Kelamin	Jumlah Kasus	Kuliah (C1)	Bekerja (C2)	P (C1)	P (C2)
L	Laki-laki	120	38	82	0.3669	0.5187
P	Perempuan	142	65	77	0.6311	0.4813

Showing 1 to 2 of 2 entries

ATRIBUT PREDIKAT UN

ID	Predikat UN	Jumlah Kasus	Kuliah (C1)	Bekerja (C2)	P (C1)	P (C2)
atas_rata2	Diatas Rata-rata	121	49	72	0.4797	0.4328
bawah_rata2	Dibawah Rata-rata	141	54	87	0.5203	0.5672

Showing 1 to 2 of 2 entries

ATRIBUT PRESTASI AKADEMIK

ID	Prestasi Akademik	Jumlah Kasus	Kuliah (C1)	Bekerja (C2)	P (C1)	P (C2)
Ada	Ada	17	1	16	0.0097	0.1006
Tidak	Tidak	245	102	143	0.3903	0.6094

Showing 1 to 2 of 2 entries

ATRIBUT PENDIDIKAN TERAKHIR ORANG TUA

ID	Pendidikan Orang Tua	Jumlah Kasus	Kuliah (C1)	Bekerja (C2)	P (C1)	P (C2)
dibawah_sma	Tidak tamat SMA	66	2	64	0.0184	0.4025
sarjana	Sarjana	70	51	19	0.4991	0.1799
SMA	SMA	126	50	76	0.4854	0.4180

Showing 1 to 3 of 3 entries

ATRIBUT PEKERJAAN ORANG TUA

ID	Pekerjaan Orang Tua	Jumlah Kasus	Kuliah (C1)	Bekerja (C2)	P (C1)	P (C2)
Petani	Petani	130	1	129	0.0087	0.6112
PNS	PNS	46	30	16	0.2913	0.1000
Wiraswasta	Wiraswasta	86	72	14	0.6990	0.0887

Showing 1 to 3 of 3 entries

Gambar 4.20 Hasil Perhitungan Jumlah Kasus yang Sama

4) Kalikan Semua Nilai Hasil

Tahap selanjutnya setelah diketahui probabilitas C1 dan probabilitas C2 untuk masing-masing atribut serta probabilitas C1 dan probabilitas C2 untuk total jumlah kasus keseluruhan, langkah berikutnya adalah mengalikan semua nilai hasil dari dataset yang dicari kelasnya, dimisalkan data yang dicari adalah seperti yang ditunjukkan pada Gambar 4.21.

DATA TRAINING ALUMNI

10 records per page

Nama	NIS	Jenis Kelamin	Predikat UN	Prestasi Akademik	Pendidikan Orang Tua	Pekerjaan Orang Tua	Status Alumni
Amelia Alfiani	600049	P	Dibawah Rata-rata	Tidak	Sarjana	Wiraswasta	Kuliah

Gambar 4.21 Data *Training* Alumni selanjutnya adalah dicari nilai masing-masing atribut untuk C1 maupun C2 dari data tersebut dan mengalikannya untuk

masing-masing kelas C1 dan C2 seperti yang disajikan pada Gambar 4.22

Bayesian Classification

Admin | 2017-2018, 2018 & 2019

Detail Record - Amelia Alfiani (600049)

DATA ATRIBUT C1

No	Atribut	Kategori	Nilai
1	Jenis Kelamin	Perempuan	0.6311
2	Predikat UN	Dibawah Rata-rata	0.5243
3	Prestasi Akademik	Tidak	0.9903
4	Pendidikan Terakhir Orang Tua	Sarjana	0.4951
5	Pekerjaan Orang Tua	Wiraswasta	0.6990

Hasil Kali masing-masing atribut untuk kelas C1 adalah: 0.13340049

DATA ATRIBUT C2

No	Atribut	Kategori	Nilai
1	Jenis Kelamin	Perempuan	0.4843
2	Predikat UN	Dibawah Rata-rata	0.5472
3	Prestasi Akademik	Tidak	0.8994
4	Pendidikan Terakhir Orang Tua	Sarjana	0.1995
5	Pekerjaan Orang Tua	Wiraswasta	0.0881

Hasil Kali masing-masing atribut untuk kelas C2 adalah: 0.00250933

Gambar 4.22 Hasil Kali Masing-masing Atribut

Setelah didapat perkalian untuk masing-masing kelas C1 dan kelas C2, proses selanjutnya adalah mengalikan hasil perkalian tersebut terhadap probabilitas C1 dan C2 dari keseluruhan jumlah kasus seperti yang ditunjukkan pada Gambar 4.23.

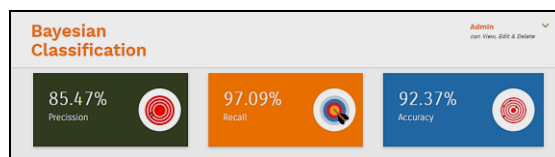
PERHITUNGAN AKHIR

P (C1)	P (C2)	Nilai Akhir C1	Nilai Akhir C2	Hasil Klasifikasi
0.39312977	0.60687023	0.0445811024	0.0015228358	C1 / <i>Kuliah</i>

Gambar 4.23 Hasil Perkalian Terhadap Probabilitas Jumlah Kasus

5) Confusion Matrix

Pengukuran menggunakan metode *Confusion Matrix* hanya dapat dilakukan pada data *training* karena pada data *training* terdapat data kelas C1 dan C2 yang didapat berdasarkan fakta di lapangan sehingga dapat dibandingkan dengan hasil klasifikasi yang telah dihitung sebelumnya sedangkan data testing belum ada data kelas tersebut. Pada aplikasi ditunjukkan pada Gambar 4.24



Gambar 4.24 Hasil Confusion Matrix

4.2 Pengujian Aplikasi

1) Pengujian Aplikasi Menggunakan Data Testing

Pengujian aplikasi menggunakan data testing bertujuan untuk memastikan bahwa aplikasi yang dibuat mampu menghitung klasifikasi yang menghasilkan status alumni yang dicari. Berbeda dengan form data training pada Gambar 4.18, form data testing tidak terdapat kolom masukan untuk status alumni, karena status alumni akan ditentukan sendiri oleh aplikasi.

Insert data testing

Nama

NIS

Jenis Kelamin

Predikat UN

Prestasi Akademik

Pendidikan Orang Tua

Pekerjaan Orang Tua

Gambar 4.29 Form Data Testing

2) Pengujian Aplikasi Menggunakan Perhitungan Manual

Pengujian menggunakan data perhitungan manual bertujuan untuk menguji ketelitian perhitungan yang dilakukan oleh aplikasi yang dibuat. Caranya adalah membandingkan hasil klasifikasi dari data yang dihitung secara manual dan telah di verifikasi

kebenarannya dengan perhitungan yang dikerjakan secara otomatis oleh aplikasi. Selain membandingkan hasil klasifikasi, juga dibandingkan hasil perhitungan dari *Confusion Matrix*.

Dari perhitungan manual dengan perhitungan yang dilakukan oleh aplikasi memiliki hasil yang sama, sehingga aplikasi telah bekerja sesuai yang diharapkan.

3) Pengujian Aplikasi Dengan Banyak Record

Pengujian aplikasi dengan menggunakan banyak record bertujuan untuk mengetahui hasil klasifikasi dan hasil *Confusion Matrix*. Record yang digunakan sebanyak 10.109 record.

Tabel 4.10 Hasil Eksekusi Masing-masing Dataset

Dataset	Presisi	Recall	Akurasi	Waktu Eksekusi
Data 262	85,47%	97,09%	92,37%	17 Detik
Data 10.109	83,69%	97,63%	91,86%	20 Menit, 59 Detik

V. PENUTUP

5.1 Kesimpulan

1. Dari data training sebanyak 262 record yang didapat dari fakta di lapangan, hasil analisa klasifikasi menggunakan metode dan aplikasi Naive Bayes menunjukkan alumni sebanyak 117 alumni memilih untuk kuliah dan sebanyak 145 alumni memilih untuk bekerja.
2. Dari data training sebanyak 262 record, aplikasi mampu melakukan proses analisa

dengan akurasi sebesar 92,37%, presisi sebesar 85,47% dan recall sebesar 97,09%.

3. Dari hasil pengujian menggunakan beberapa dataset menunjukkan bahwa waktu eksekusi berbanding lurus dengan banyaknya record, semakin banyak record semakin banyak waktu yang dibutuhkan untuk melakukan proses analisa sedangkan hasil dari presisi, recall dan akurasi tidak.

5.2 Saran

terdapat beberapa saran untuk dilakukan perbaikan maupun penambahan analisis maupun fitur dari aplikasi itu sendiri, adapun saran yang bisa diteliti lebih lanjut adalah:

1. Membandingkan metode Naive Bayes pada penelitian ini dengan metode klasifikasi lain untuk mengukur metode klasifikasi mana yang paling cocok dengan dataset alumni.
2. Pada aplikasi, seluruh atribut maupun kelas dibuat dinamis, sehingga dapat dilakukan penambahan atribut maupun kelas, dengan demikian klasifikasi tidak hanya terpaku pada dataset alumni namun bisa berbagai macam dataset yang lain.

VI. REFERENSI

- [1] Sulastri, G. (2017, September 26). *Penerapan Data Mining Dalam Pengelompokan Penderita Thalassaemia*. Diambil kembali dari Jurnal Teknologi dan Sistem Informasi:
<http://teknosi.fti.unand.ac.id/index.php/teknosi/article/view/335>

- [2] Saleh. (2015, Juli). *Implementasi Metode Klasifikasi Naive Bayes Dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga*. Diambil kembali dari Open Journal System "Amikom OJS Journal": <http://ojs.amikom.ac.id/index.php/cite/article/view/375>
- [3] Kumara, S. (2015, November 26). *Klasifikasi Data Mining Untuk Penerimaan Seleksi Calon Pegawai Negeri Sipil 2014 Menggunakan Algoritma Decision Tree C4.5*. Diambil kembali dari UDiNus Repository: <http://eprints.dinus.ac.id/16516/>
- [4] Rosandy. (2016, Mei 1). *Perbandingan Metode Naive Bayes Classifier Dengan Metode Decision Tree (C4.5) Untuk Menganalisa Kelancaran Pembiayaan (Studi Kasus: KSPPS / BMT Al-Fadhila)*. Diambil kembali dari Portal JURNAL IBI DARMAJAYA: <https://jurnal.darmajaya.ac.id/index.php/jtim/article/view/648>
- [5] Marcos, & Hidayah. (2014). **IMPLEMENTASI DATA MINING UNTUK KLASIFIKASI NASABAH KREDIT BANK "X" MENGGUNAKAN CLASSIFICATION RULE**. researchgate.
- [6] Marisa. (2013). *Educational Data Mining (Konsep dan Penerapan)*. Diambil kembali dari JURNAL TEKNOLOGI INFORMASI: Teori, Konsep, dan Implementasi VOL. 4 NO. 2 TAHUN 2013: <http://ejournal.stimata.ac.id/index.php/TI/article/view/108/148>
- [7] Saputra, & Primadasa. (2018). *Penerapan Teknik Klasifikasi Untuk Prediksi Kelulusan Mahasiswa Menggunakan Algoritma K-Nearest Neighbour*. Techno.COM, 395-403.
- [8] Sari. (2016). *Implementasi Metode Naive Bayes Untuk Memprediksi Kelulusan Mahasiswa Di Program Studi Teknik Informatika Ft Un PGRI Kediri*. Diambil Dari Sistem Informasi Manajemen Karya Ilmiah Universitas Nusantara PGRI Kediri: http://simki.unpkediri.ac.id/mahasiswa/file_artikel/2016/12.1.03.02.0081.pdf
- [9] Shadiq. (2009, Juni). *Keoptimalan Naive Bayes Dalam Klasifikasi*. Diambil kembali dari Keoptimalan Naive Bayes - Scribd: <https://www.scribd.com/document/73487637/Keoptimalan-Naive-Bayes>
- [10] Supianto. (2014). *Pengenalan Pola. Klasifikasi - Naive Bayes*. Malang, Jawa Timur, Indonesia: Universitas Brawijaya.
- [11] Wardoyo. (2016, Desember 10). *Klasifikasi. Bayesian Classification*. Jember, Jawa Timur, Indonesia: Universitas Muhammadiyah Jember.