

PENGGUNAAN ALGORITME NAÏVE BAYES UNTUK PREDIKSI HASIL BELAJAR SISWA SMAN 3 RANGKASBITUNG BERDASARKAN SOSIAL EKONOMI

Ahmad Subaji¹, Dewi Kusumaningsih²

^{1,2} Program Studi Teknik Informatika, Fakultas Teknologi Informasi, Universitas Budi Luhur, Jakarta, Indonesia

Email: 1*1711501708@student.budiluhur.ac.id, 2dewi.kusumaningsih@budiluhur.ac.id,

(* : corresponding author)

Abstrak- Setiap tahun ajaran baru setiap siswa baru diharuskan mengisi biodata yang memuat data sosial-ekonomi keluarganya. Data tersebut dituliskan dalam formulir yang disediakan SMAN 3 Rangkasbitung. Setelah menempuh pembelajaran selama satu tahun ajaran siswa mendapatkan hasil belajar, namun tidak semua siswa mendapatkan hasil belajar yang memuaskan. Data yang diterima sekolah setiap tahun terus bertambah, sehingga mengakibatkan berlimpahnya data siswa. *Data mining* merupakan cara yang mampu mengolah data yang sangat besar tersebut dapat menjadi sebuah informasi. *Data mining* dengan metode Naive Bayes mampu memprediksi masa depan menggunakan probabilitas yang menerapkan aturan Bayes dengan asumsi memprediksi masa depan dengan menggunakan data yang terkait. Tujuan penelitian ini adalah mengetahui implementasi teknik data mining dengan metode Naive Bayes menyajikan informasi dalam memprediksi hasil belajar siswa. Penelitian ini bersifat studi kasus dan untuk memprediksi hasil belajar siswa dengan data mining metode Naive Bayes. Hasil penelitian pengujian dengan data *training* sebanyak 132 data diperoleh keakuratan sebesar 87,5%. Data testing sebanyak 80 siswa sebagai data uji didapatkan hasil bahwa siswa dengan hasil belajar memuaskan sebanyak 66 siswa atau sekitar 82,50% dari jumlah data *testing* dan siswa dengan hasil belajar kurang memuaskan sebanyak 14 siswa atau sekitar 17,50%.

Kata Kunci: Data Mining, Naive Bayes, Prediksi, Hasil Belajar

USING THE NAÏVE BAYES ALGORITHM FOR PREDICTION OF STUDENTS' LEARNING OUTCOMES AT SMAN 3 RANGKASBITUNG BASED ON SOCIO-ECONOMY

Abstract- Every new academic year, each new student is required to fill out a bio that contains socio-economic data for their family. The data is written in a form provided by SMAN 3 Rangkasbitung. After studying for one academic year students get learning results, but not all students get satisfactory learning results. The data that schools receive every year continues to grow, resulting in an abundance of student data. *Data mining* is a way that can process very large data into information. *Data mining* with the Naive Bayes method is able to predict the future using probabilistics that apply Bayes' rules with the assumption of predicting the future using related data. The purpose of this study was to determine the implementation of data mining techniques with the Naive Bayes method to present information in predicting student learning outcomes. This research is a case study and is used to predict student learning outcomes using the Naive Bayes data mining method. The results of testing research with training data of 132 data obtained an accuracy of 87.5%. Data testing of 80 students as test data showed that students with satisfactory learning outcomes were 66 students or around 82.50% of the total testing data and students with unsatisfactory learning outcomes were 14 students or around 17.50%.

Keywords: Data Mining, Naive Bayes, Predictions, Learning Outcomes

1. PENDAHULUAN

Kelangsungan, kemajuan dan mandirinya suatu bangsa terletak pada pendidikan yang ada pada bangsa itu sendiri. Artinya, maju dan mundurnya suatu bangsa terletak dari bagaimana bangsa itu mengelola pendidikannya. Oleh karena itu, pendidikan nasional harus dikelola dengan manajemen yang berkualitas [1].

Pengelolaan pendidikan yang berkualitas, tidak hanya pada proses pembelajaran saja. Semua faktor yang terkait dengan pendidikan sangat penting apabila dikelola dengan baik. Hal tersebut dapat menjadi modal penting dalam pengelolaan pendidikan yang berkualitas [2]. Salah satu faktor penting keberhasilan pendidikan adalah peserta didik. Namun, banyak lembaga pendidikan yang dihadapkan masalah hasil belajar yang tidak sesuai dengan harapan. Umumnya, upaya pemecahan masalahnya difokuskan pada upaya perbaikan metode pembelajaran dan pemenuhan sarana-prasarana. Upaya tersebut tentunya belum cukup, karena masih banyak faktor yang belum dimanfaatkan sebagai sumber informasi guna memecahkan permasalahan hasil belajar tersebut. Oleh karena itu, perlu adanya upaya deteksi dini guna memecahkan masalah tersebut. Hal tersebut

dianggap penting karena dengan adanya deteksi dini, lembaga pendidikan dapat mengambil langkah-langkah antisipatif guna mengatasi hambatan yang dihadapi peserta didik dalam mencapai hasil belajar yang baik.

Berdasarkan hal tersebut, dengan informasi secara komprehensif dengan memanfaatkan keseluruhan data yang ada, salah satunya adalah data peserta didik saat melakukan pendaftaran menjadi siswa pada sekolah, yaitu data sosial ekonomi. Pengelolaan data sosial ekonomi tersebut dengan tepat bisa menjadi salah satu upaya guna memprediksi sejak dini mengenai potensi siswa tersebut kedepannya. Hal tersebut dapat menjadi bahan guna mengambil langkah antisipatif oleh sekolah untuk menghindari kebijakan yang kurang bijaksana diterapkan kepada peserta didik yang hasil belajarnya tidak baik, misalnya memutuskan peserta didik tersebut tidak naik kelas atau kemungkinan yang lebih jauh adalah dikeluarkan dari sekolah.

Pemanfaatan data-data tersebut yang sebelumnya hanya dianggap sebagai “sampah” atau belum diketahui manfaatnya dilakukan melalui proses yang disebut *Data Mining*. Proses tersebut dilakukan dengan mengekstraksi data yang sebelumnya tidak diketahui manfaatnya, tetapi dapat diduga bermanfaat untuk pengambilan keputusan yang penting [3], juga berguna bagi upaya pengembangan [4]. *Data Mining* juga menjadi penting dalam bidang pendidikan karena dapat menjadi salah satu solusi mengatasi permasalahan di bidang pendidikan.

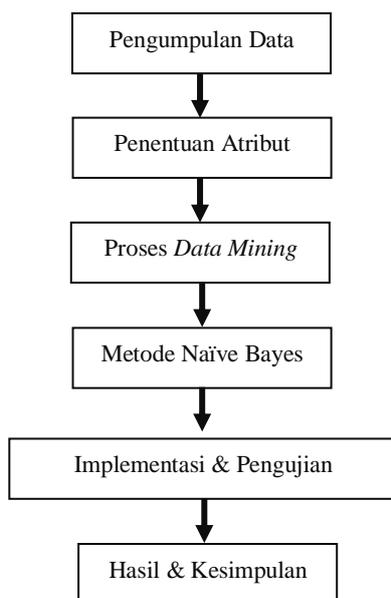
Dengan adanya *data mining*, maka tumpukan data diekstrak guna diperoleh data yang penting dan relevan untuk penyelesaian masalah. Selanjutnya, dari data-data tersebut digunakan untuk memprediksi hasil belajar siswa. Metode prediksi yang dapat digunakan adalah metode atau algoritme Naïve Bayes. Penerapan algoritme Naïve Bayes dalam berbagai kasus sering kurang berhasil, namun banyak kasus di bidang pendidikan penerapan algoritme Naïve Bayes dapat diandalkan juga prosesnya cepat dan hasilnya akurat. Data yang bisa digunakan dalam penerapan algoritme Naïve Bayes guna memprediksi hasil belajar adalah data siswa saat PPDB (Penerimaan Peserta Didik Baru) yang didalamnya memuat data sosial ekonomi calon peserta didik.

Penelitian Firmansyah & Yulianto yang berjudul *Prediksi Hasil Belajar Menggunakan Naïve Bayes Classifier pada Tingkat Sekolah Dasar* [5] menunjukkan hasil bahwa Dengan Naïve Bayes Classifier dapat memprediksi belajar dari peserta didik sekolah dasar berdasarkan variabel kehadiran dan peringkat dengan nilai akurasi mencapai 89,93%, nilai presisi mencapai 68,95 % dan nilai mencapai 65%. Hasil yang sama juga ditunjukkan pada penelitian baru yang berjudul *Implementation of Naïve Bayes Classifier-based Machine Learning to Predict and Classify New Students at Matana University* [6].

2. METODE PENELITIAN

2.1 *Flowchart* Penelitian

Proses penelitian menggunakan metode KDD (*Knowledge Data Discovery*) sehingga dalam prosesnya tidak terjadi penyimpangan atau kekeliruan dari hasil yang diharapkan. Metode ini juga menjadikan penelitian menjadi lebih baik sesuai dengan tujuan. Alur atau *flowchart* penelitian menggunakan metode *Waterfall* ditunjukkan pada Gambar 1.



Gambar 1. *Flowchart* Penelitian

2.2 Penerapan Metode yang Digunakan

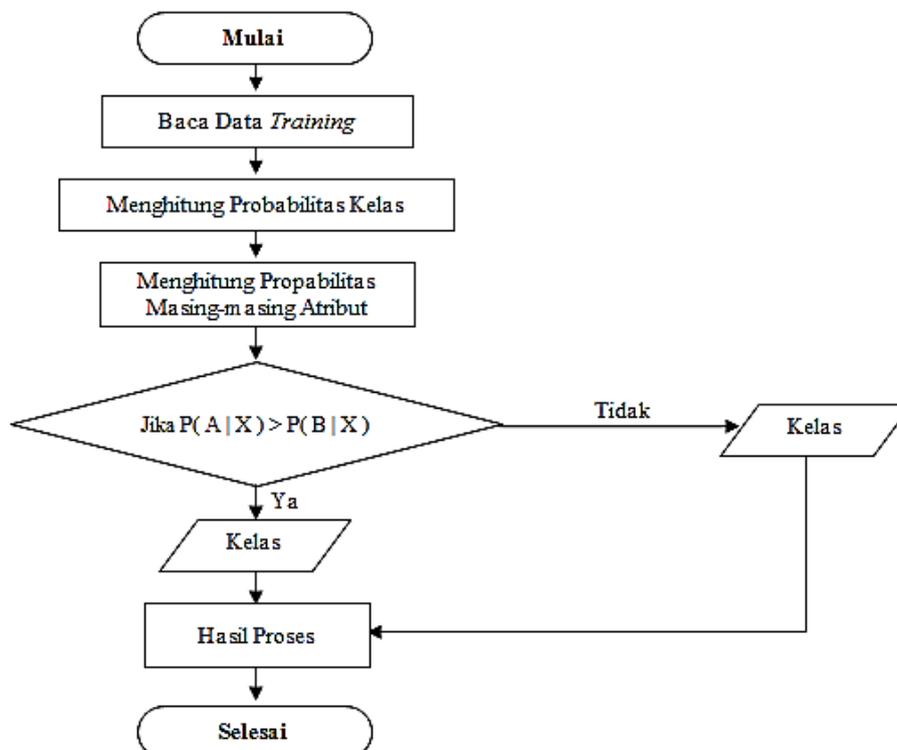
Ketepatan dalam penerapan metode penelitian menjadi kunci keberhasilan dari penelitian agar dapat menjawab atau menyelesaikan permasalahan atau penemuan kebenaran. Sesuai dengan tujuan, penelitian ini merupakan penelitian kuantitatif karena dalam prosesnya dilakukan guna mencari hubungan antar variabel yang diteliti. Data mining yang diperoleh selanjutnya diolah menggunakan metode atau algoritma Naïve Bayes untuk memprediksi hasil belajar siswa SMA Negeri 3 Rangkasbitung. Penelitian ini menggunakan data pendaftaran siswa baru tiga tahun terakhir, yaitu tahun 2020-2022. Data diperoleh dari dokumen Ledger SMAN 3 Rangkasbitung Tahun 2019-2022. Tabel 1. merupakan atribut yang digunakan dalam memprediksi menentukan hasil belajar siswa meliputi: Gender, Status Orang Tua, Pendidikan Ayah, Pendidikan Ibu, Pekerjaan Orang Tua, Pendapatan Orang Tua, Jumlah Saudara, Jarak Rumah ke Sekolah, Akses Transportasi, dan Keputusan (Hasil Belajar Siswa).

Tabel 1. Atribut-atribut dan Kategorisasinya

No.	Atribut	Kategori
1.	Gender	Laki-laki; Perempuan
2.	Status Tempat Tinggal	Bersama Orang Tua; Pisah
3.	Pendidikan Ayah	Tidak Sekolah, SD; SMP; SMA; Minimal S1
4.	Pendidikan Ibu	Tidak Sekolah, SD; SMP; SMA; Minimal S1
5.	Pekerjaan Orang Tua	Tidak Bekerja; ASN; Non-ASN
6.	Pendapatan Orang Tua	< 5 Juta; >= 5 Juta
7.	Jumlah Saudara	<= 3 Orang; > 3 Orang
8.	Jarak Rumah ke Sekolah	< 2 KM; >= 2 KM
9.	Akses Transportasi	Pribadi; Umum
10.	Hasil Belajar Siswa	Memuaskan; Kurang Memuaskan

2.3 Algoritme Naïve Bayes

Model klasifikasi untuk prediksi hasil belajar menggunakan metode Naïve Bayes *Classifier* ini akan melalui proses data selection, data transformation dan klasifikasi dengan menggunakan metode Naive Bayes *Classifier* dengan *flowchart*-nya disajikan pada Gambar 2.



Gambar 2. Flowchart Algoritme Naïve Bayes

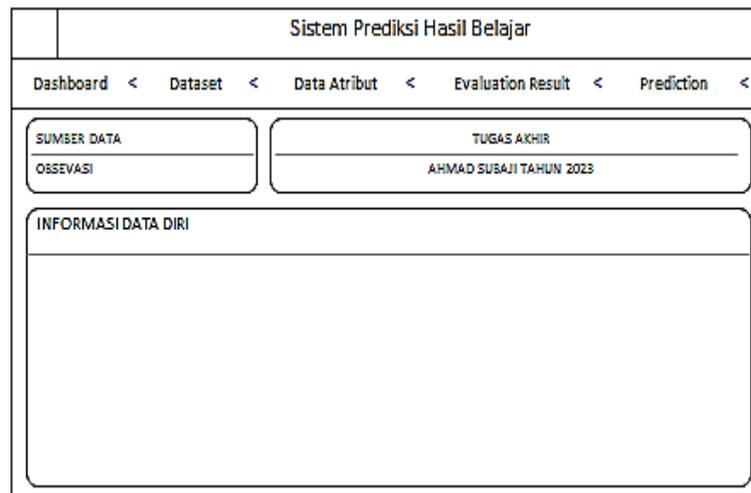
Berdasarkan Gambar 2 proses pengolahan data prediksi dengan menggunakan *Naïve Bayes Classifier* diawali dengan penggunaan *data training*.

2.4 Rancangan Menu

- Dalam pembuatan aplikasi Prediksi Hasil Belajar Siswa dibuat dalam beberapa menu, yaitu:
- Dashboard*. Menu ini merupakan tampilan awal dari aplikasi. Saat program dijalankan, maka akan muncul menu ini.
 - Dataset*. Menu ini merupakan menu yang digunakan untuk menginput data. Ada dua sub-menu pada menu ini, yaitu: *data training* dan *data test*. Proses input dapat dilakukan dengan dua cara, yaitu melalui proses upload (import data) dan input manual. Untuk proses impor file yang support adalah berupa file .EXE atau .CSV.
 - Data Atribut*. Menu ini berisi atribut-atribut yang akan digunakan dalam proses pengujian.
 - Evaluation Result*. Menu ini berisi mengenai hasil dari proses pengujian, juga hasil perhitungan probabilitas.
 - Prediction*. Menu ini berisi hasil yang diperoleh dari prediksi kasus.

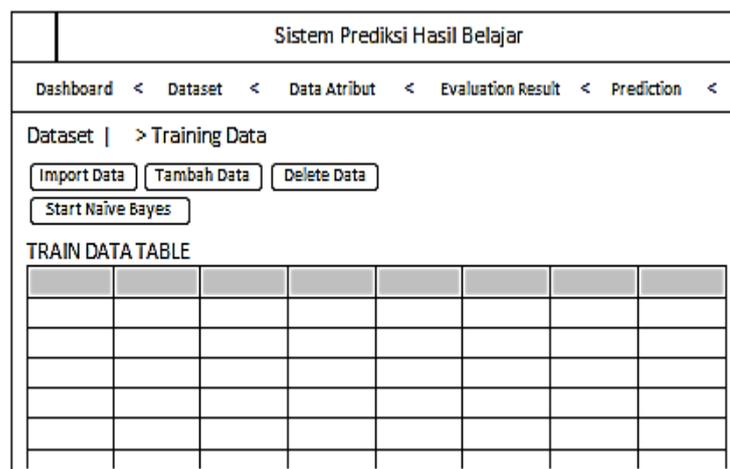
2.5 Rancangan Layar

- Dalam pembuatan aplikasi Prediksi Hasil Belajar Siswa, diperlukan rancangan layar, sebagai berikut:
- Rancangan Layar *Dashboard*



Gambar 3. Rancangan Layar Dashboard

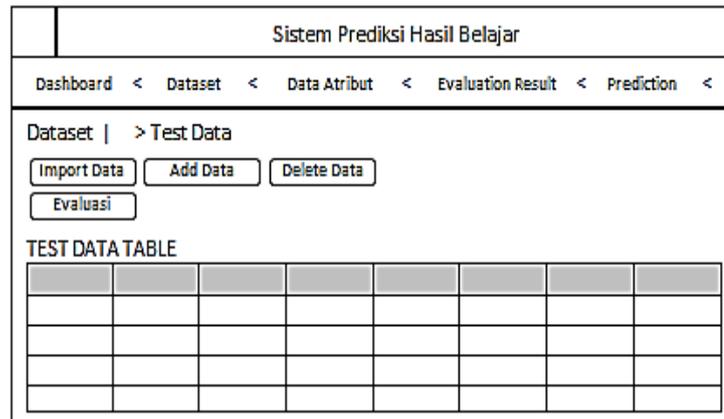
- Rancangan Layar Dataset
 - Data Training*



Gambar 4. Rancangan Layar Data Training

Tampilan layar *Data Training* pada Gambar 4. berisi beberapa fitur, yaitu: import data, tambah data, delete data dan start Naïve Bayes. Import Data merupakan fitur yang digunakan untuk mengimport data yang akan diolah. Juga fitur-fitur tambah dan delet data guna melakukan perubahan dari data yang diinput. Hasil dari input data ditampilkan dalam layar.

2. Data Test



Gambar 5. Rancangan Layar Data Test

Sama halnya dengan *Data Training* pada layar *Data Test* ini juga dilengkapi fitur Import Data, Add Data (tambah data), Delete Data (hapus data), dan Evaluasi. Pada layar ini juga ditampilkan hasil dari data *test* pada Gambar 5.

2.6 Confusion Matrix

Perhitungan *accurate*, *recall*, *precision*, dan *F-score* diperoleh dengan menerapkan metode *Confusion Matrix* [7]. *Confusion Matrix* atau disebut juga dengan *error matrix* adalah salah satu jenis tabel yang digunakan untuk menampilkan performansi dari suatu model klasifikasi. Ada empat elemen dari *Confusion Matrix*. *confusion matrix*, kelas positif diprediksi sebagai positif (positif benar), kelas negatif diprediksi sebagai negatif (negatif benar), kelas positif diprediksi sebagai negatif (positif palsu), dan kelas negatif diprediksi sebagai positif (negatif palsu) [8].

3. HASIL DAN PEMBAHASAN

Sesuai dengan tujuan penelitian, hasil dan pembahasan difokuskan pada tahapan penelitian, dilanjutkan dengan proses evaluasi dan pembahasan.

3.1 Lingkungan Percobaan

Aplikasi Prediksi Hasil Belajar ini dibuat menggunakan bahasa pemrograman PHP dan database MySQL.

3.1.1 Spesifikasi Hardware dan Software

Berikut adalah spesifikasi hardware dan software yang dibutuhkan agar aplikasi dapat berjalan dengan baik.

a. Hardware (Perangkat Keras)

Spesifikasi perangkat keras yang digunakan pada penelitian ini adalah:

1. Processor : Intel Core I3 2.26 Ghz
2. RAM : 4 GB
3. Hardisk : 320 GB

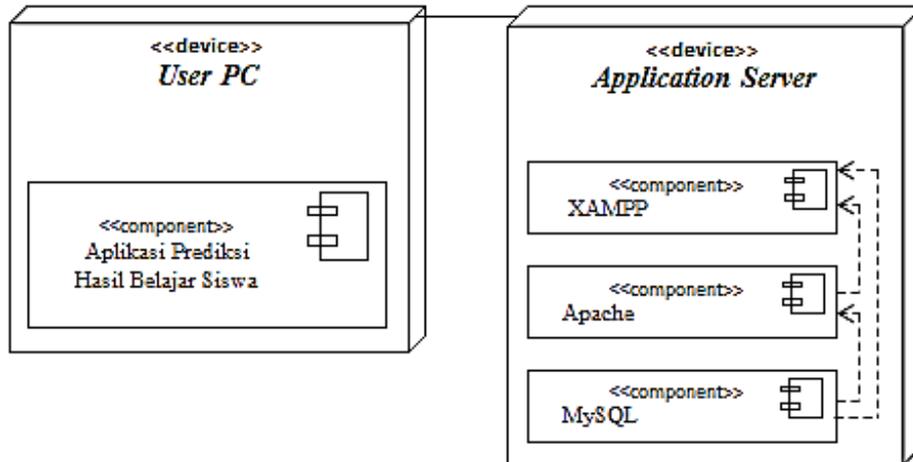
b. Software (Perangkat Lunak)

Perangkat lunak aplikasi (Application Software) yang digunakan untuk merancang aplikasi Prediksi Hasil Belajar adalah :

1. Apache
2. MySQL
3. Google Chrome
4. Sublime

3.1.2 Deployment Diagram

Deployment Diagram dari sistem yang dibuat digambarkan sebagai berikut:

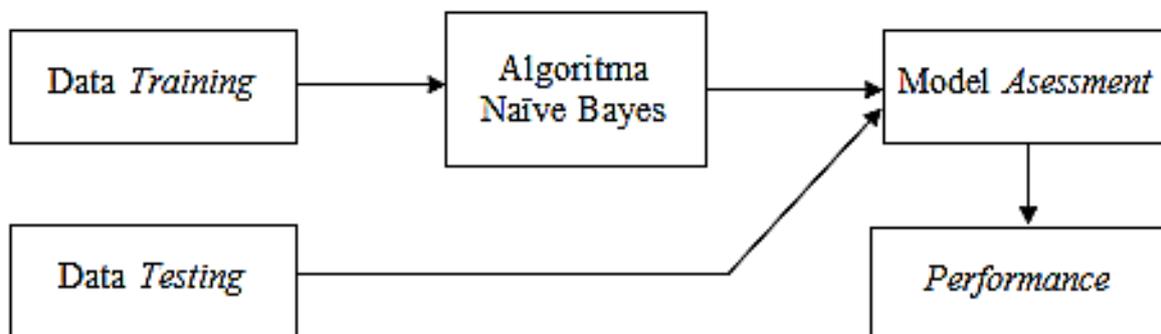


Gambar 6. *Deployment Diagram*

Gambar 6, menunjukkan bahwa aplikasi Prediksi Hasil Belajar dapat diakses dengan menggunakan PC (laptop atau komputer). Aplikasi diperuntukan oleh user.

3.1.3 Arsitektur Aplikasi

Arsitektur aplikasi perlu dibuat guna memahami konsep aplikasi yang akan dibangun. Gambar 7. merupakan gambaran dari arsitektur aplikasi yang menggambarkan garis besar proses dari keseluruhan sistem yang dibangun pada penelitian ini.



Gambar 7. Arsitektur Aplikasi

3.2 Implementasi Metode

Aplikasi Prediksi Hasil Belajar ini dibuat menggunakan algoritme Naïve Bayes. Untuk menjalankan aplikasi tersebut dilakukan dengan beberapa tahapan.

Berikut adalah tahapan-tahapan tersebut:

3.2.1 Persiapan Data

Tahapan ini diawali dengan melakukan pengambilan data sampel atau contoh data dari data siswa Kelas X tahun 2019-2022 dan nantinya akan dijadikan sebagai data training. Data yang digunakan sudah dilakukan proses pembersihan dan transformasi dalam bentuk kategori. Pengujian juga menggunakan data siswa Kelas X tahun 2019-2022. Data pengujian sebanyak 212 *record* dan diambil 132 *record* untuk digunakan sebagai data training.

Berdasarkan hasil pengolahan data dan jumlah data yang digunakan tersebut dapat dibagi menjadi dua kelas kategori “Memuaskan” sebanyak 108 siswa dan kategori “Kurang Memuaskan” sebanyak 24 siswa.

Proses pengujian, data dibagi menjadi 2 bagian yaitu data training dan data testing dengan menggunakan Naive Bayes. Data training digunakan untuk membentuk tabel probabilitas dan data testing digunakan untuk menguji probabilitas yang telah terbentuk.

Tabel 2. merupakan data 212 record yang dibagi menjadi 2, yaitu data *training* dan data *test*:

Tabel 2. Data Mentah (Gabungan Data *Training* dan Data *Test*)

Gender	Status Tempat Tinggal	Pend. Ayah	Pend. Ibu	Pekerjaan Orang Tua	Pendapatan Orang Tua	Jumlah Saudara	Jarak Rumah-Sekolah	Akses Transp	Hasil Belajar
Laki-laki	Bersama_Org_Tua	SMA	SMA	Tidak_Bekerja	<5_Jt	≤3	<2_KM	Umum	Memuaskan
Perempuan	Bersama_Org_Tua	≥S1	SMA	ASN	<5_Jt	≤3	<2_KM	Pribadi	Memuaskan
Perempuan	Bersama_Org_Tua	Tidak_Sekolah	Tidak_Sekolah	Non_ASN	<5_Jt	≤3	≥2_KM	Umum	Kurang_Memuaskan
Perempuan	Bersama_Org_Tua	SMA	≥S1	Non_ASN	<5_Jt	>3	<2_KM	Umum	Memuaskan
Perempuan	Bersama_Org_Tua	SMA	SMA	Tidak_Bekerja	<5_Jt	≤3	<2_KM	Pribadi	Memuaskan
Laki-laki	Bersama_Org_Tua	SMA	SD	Non_ASN	<5_Jt	≤3	<2_KM	Umum	Memuaskan
Laki-laki	Bersama_Org_Tua	SMP	SMA	Tidak_Bekerja	<5_Jt	>3	<2_KM	Umum	Kurang_Memuaskan
Laki-laki	Bersama_Org_Tua	≥S1	Tidak_Sekolah	ASN	<5_Jt	>3	<2_KM	Umum	Memuaskan
Perempuan	Bersama_Org_Tua	SD	SMA	Non_ASN	<5_Jt	≤3	<2_KM	Umum	Memuaskan
Perempuan	Bersama_Org_Tua	SMA	≥S1	Non_ASN	<5_Jt	≤3	<2_KM	Pribadi	Memuaskan
Laki-laki	Bersama_Org_Tua	SMA	SMP	Non_ASN	<5_Jt	≤3	≥2_KM	Umum	Memuaskan
Laki-laki	Bersama_Org_Tua	Tidak_Sekolah	Tidak_Sekolah	Non_ASN	<5_Jt	>3	<2_KM	Umum	Kurang_Memuaskan
Perempuan	Bersama_Org_Tua	SMA	Tidak_Sekolah	Non_ASN	<5_Jt	≤3	<2_KM	Umum	Memuaskan
...
Laki-laki	Bersama_Org_Tua	≥S1	SMA	ASN	<5_Jt	>3	≥2_KM	Pribadi	Memuaskan

Hasil seleksi data, dari sepuluh atribut yang digunakan hanya tujuh saja yang dipakai. Atribut Gender, Status Tempat Tinggal, dan Pendapatan Orang Tua tidak digunakan. Atribut Gender tidak digunakan karena hal tersebut bukan pilihan dan tidak relevan sebagai salah satu atribut yang menentukan keberhasilan hasil belajar siswa. Selanjutnya, untuk Status Tempat Tinggal tidak digunakan karena hampir 100% siswa tinggal bersama orang tua. Begitu juga dengan Pendapatan Orang Tua dikarenakan hampir 100% pendapatan orang tua di bawah 5 juta. Data tersebut di-drop karena dapat mengganggu proses penghitungan probabilitas Naïve Bayes. Dengan demikian, data yang digunakan hanya data Pendidikan Ayah, Pendidikan Ibu, Pekerjaan Orang Tua, jumlah Saudara, Jarak Rumah-Sekolah, Akses Transportasi, dan Hasil Belajar.

3.2.2 Pengolahan Data

Tahapan pengolahan data yang sudah diseleksi, selanjutnya diolah kedalam metode/algorithm Naïve Bayes dari teknik data mining. Berdasarkan data dan atribut yang telah didapatkan pada proses sebelumnya, maka pada bagian ini akan dijelaskan bagaimana penggunaan metode *Naïve Bayes Classifier* dalam klasifikasi data tersebut. Untuk melakukan klasifikasi, diperlukan data training sebagai data pembelajaran untuk menentukan label kelas pada data testing. Data training yang digunakan adalah data training yang telah melewati proses selection. Pengolahan data hanya menggunakan data training hasil seleksi dari data mentah, sebanyak 132 *record*. Hasil pengolahan data diperoleh nilai probabilitas akhir yang akan digunakan dalam proses selanjutnya yaitu perhitungan dengan menggunakan metode *Confusion Matrix*.

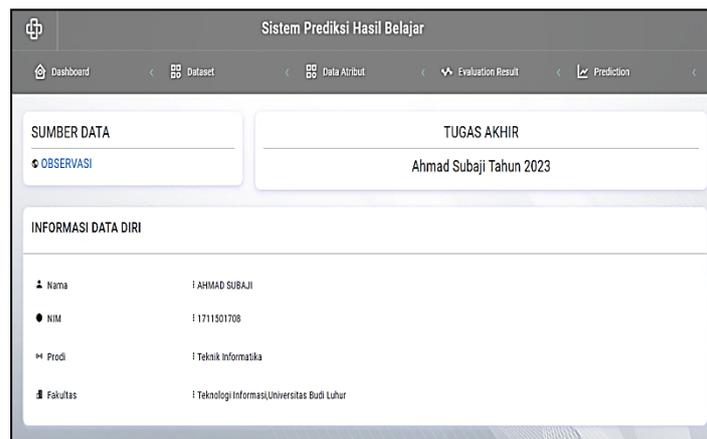
Tabel 3. Hasil Perhitungan Probabilitas Akhir

Atribut	Data Uji	Probabilitas "Memuaskan"	Probabilitas "Kurang Memuaskan"
Pend_Ayah	≥S1	0.750	0.950
Pend_Ibu	Tidak_Sekolah	0.148	0.208
Pekj_Ortu	ASN	0.194	0.041
Jml_Sdr	">3	0.417	0.417
Jarak_Rmh	"<2_KM	0.833	0.750
Akses_Transp.	Umum	0.657	0.833
Hasil_Belajar	Memuaskan	0.818	0.182
Nilai Probabilitas * $\sum_{i=1}^n P(X H)P(X)$		0.004019985	0.000384136

Tabel 3. Menunjukkan bahwa nilai probabilitas hasil belajar “*Memuaskan*” pada kasus data uji sebesar 0.004019985 dan nilai probabilitas hasil belajar “*Kurang Memuaskan*” sebesar 0.000384136. Dari kedua nilai tersebut, nilai probabilitas hasil belajar “*Memuaskan*” lebih besar dari pada nilai probabilitas “*Kurang Memuaskan*”, sehingga prediksi yang dihasilkan sesuai dengan data aktualnya siswa tersebut. Kedua hasil yang sudah ditentukan pada setiap kelas, bandingkan nilai yang paling tinggi atau mendekati 1. Apabila kelas “*Memuaskan*” paling mendekati 1, maka hasilnya “*Memuaskan*” dan begitu pula sebaliknya.

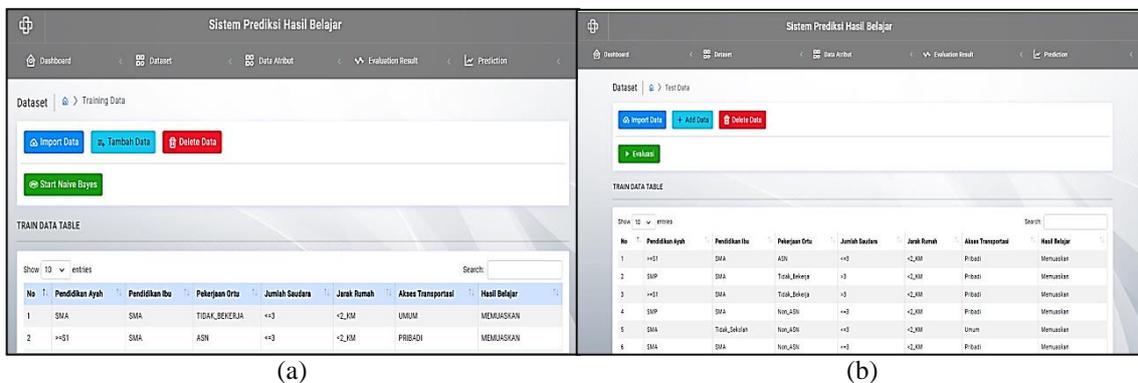
3.3 Tampilan Layar Aplikasi

Aplikasi Prediksi Hasil Belajar Siswa dirancang dengan 5 (lima) menu utama, yaitu: Dashboard (sama dengan Home), Dataset, Data Atribut, Evaluation Result, dan Prediction. Dalam layar Dashboard (Gambar 8.) berisi informasi data admin dan informasi lainnya.



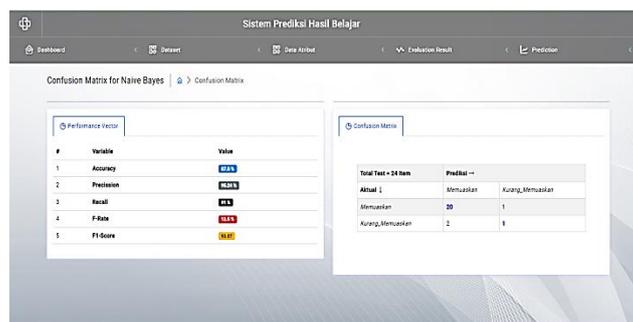
Gambar 8. Tampilan Layar *Dashboard*

Selanjutnya adalah tampilan layar menu *Dataset*. Pada menu ini dapat ditampilkan sub-menu *Data Training* dan *Data Test*, Gambar 9.



(a) (b)
Gambar 9. Tampilan Data *Training* (a) dan Data *Test* (b)

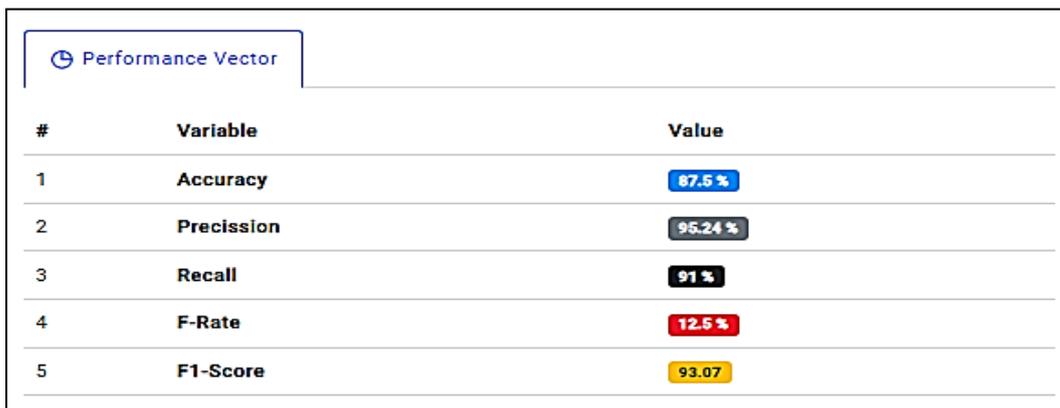
Hasil akhir dari Aplikasi Prediksi Hasil Belajar adalah hasil perhitungan dari Confusion Matrix yang menampilkan hasil dari *Accurate*, *Recall*, *Precision*, *F-Rate*, dan *F-Score*, disajikan pada Gambar 10.



Gambar 10. Layar Hasil Evaluasi Program Prediksi Hasil Belajar

3.4 Hasil Prediksi Aplikasi

Hasil pengujian diperoleh nilai *Accuracy* (menggambarkan seberapa akurat model dapat mengklasifikasikan dengan benar) untuk memprediksi hasil belajar siswa sebesar 87,5%, *Precision* (menggambarkan tingkat keakuratan antara data yang diminta dengan hasil prediksi yang diberikan oleh model) sebesar 95,24%, *Recall* (menggambarkan keberhasilan model dalam menemukan kembali sebuah informasi) sebesar 91%, *F-Rate* (merupakan gambaran tingkat kegagalan model dalam memprediksi dengan benar) sebesar 12,5%, dan *F1-Score* (merupakan rata-rata harmonik dari *Precision* dan *Recall*) sebesar 93,07, disajikan pada Gambar 11..



#	Variable	Value
1	Accuracy	87.5 %
2	Precision	95.24 %
3	Recall	91 %
4	F-Rate	12.5 %
5	F1-Score	93.07

Gambar 11. Hasil Perhitungan *Confusion Matrix*

Nilai 87,5% membuktikan bahwa metode Naïve Bayes baik digunakan untuk melakukan klasifikasi prediksi hasil belajar siswa. Hasil tersebut sesuai dengan pendapat yang dikemukakan oleh Anagora *et al.* yang menyatakan bahwa *Naïve Bayes Classifier* sangat baik dan akurat untuk memprediksi model [9].

Sebanyak 80 siswa Kelas X tahun 2019-2022 sebagai data testing dengan menggunakan metode Naïve Bayes ternyata didapatkan hasil penelitian ini bahwa siswa yang hasil belajarnya memuaskan sebanyak 66 siswa atau sekitar 82,50% dari jumlah data testing dan siswa dengan hasil belajar kurang memuaskan sebanyak 14 siswa atau sekitar 17,50%.

Berdasarkan hasil implementasi dan evaluasi program terdapat kelebihan dan kekurangan dari program tersebut. Kelebihan Program 1) Program ini mampu menghadirkan hasil pengujian yang lebih cepat dan tingkat keakuratannya tinggi. 2) Dapat membantu pengambilan keputusan yang lebih tepat guna memprediksi hasil belajar siswa, apakah hasilnya memuaskan atau kurangmemuaskan, sehingga dapat segera diambil langkah tindak lanjutnya . b. Kekurangan Program 1) Aplikasi Prediksi Hasil Belajar ini masih bersifat *localhost* dan belum *online*. 2) Aplikasi belum dibuat secara full interaktif, misalnya menghapus atribut atau menambah atribut. 3) Belum adanya grafik yang menggambarkan kondisi dari masing-masing atribut.

4. KESIMPULAN

Implementasi data *training* sebanyak 132 data dengan algoritme Naïve Bayes berhasil memprediksi besarnya hasil belajar siswa dengan persentase keakuratan (*Accuracy*) sebesar 87,5%, *Precision* sebesar 95,24%, *Recall* sebesar 91%, *F-Rate* sebesar 12,5%, dan *F1-Score* sebesar 93,07. Data mining dan Naïve Bayes juga mampu menampilkan informasi prediksi hasil belajar siswa dengan menggunakan data sosial-ekonomi siswa Kelas X Tahun 2019-2022 sebagai data *training* dan data *testing*. Sebanyak 80 siswa sebagai data *test* yang dihasilkan penelitian ini bahwa siswa yang hasil belajarnya memuaskan sebanyak 66 siswa atau sekitar 82,50% dari jumlah data *testing*.

Aplikasi Hasil Belajar Siswa dimasa mendatang diharapkan menggunakan tambahan data dengan jumlah yang lebih besar dan variabel atau atribut dalam penentuan prediksi hasil belajar siswa. Input data banyaknya atribut hendaknya dapat dibuat sesuai dengan kebutuhan, juga untuk pengembangan metode klasifikasi dimasa mendatang dapat menggunakan dua atau lebih metode sehingga bisa untuk membandingkan keakuratannya.

DAFTAR PUSTAKA

- [1] Preddy Marpaung dan R. Fanry Siahaan, “Penerapan Algoritma K-Means Clustering Untuk Pemetaan Kepadatan Penduduk Berdasarkan Jumlah Penduduk Kota Medan”, *J-SAKTI: Jurnal Sains Komputer dan Informatika*, Vol. 5, No. 1, pp. 503-521, 2021, <https://tunasbangsa.ac.id/ejurnal/index.php/jsakti>
- [2] Devi Heryana, “Data Mining untuk Memprediksi Kelulusan Mahasiswa Pendidikan Matematika UIN Raden Intan Lampung Menggunakan Naïve Bayes”, *Skripsi Tidak Dipublikasikan*. Lampung: Universitas Islam Negeri Raden Intan, 2019, <http://repository.radenintan.ac.id/id/eprint/6430>
- [3] I Kadek Juni Arta, Gede Indrawan, dan Gede Rasben Dantes, “Data Mining Rekomendasi Calon Mahasiswa Berprestasi di STMIK Denpasar Menggunakan Metode *Technique for Others Reference by Similarity to Ideal Solution*”, *Jurnal Ilmu Komputer Indonesia (JIKI)*, Vol. 4, No. 1, pp. 11-21, 2019, <https://ejournal-pasca.undiksha.ac.id/index.php/jik/article/view/2765/1330>
- [4] Megawati Noperia, Ishak, dan Vina Winda Sari, “Implementasi Data Mining Pengelompokan Data Nilai Untuk Menentukan Minat Belajar Seni Budaya”, *JURNAL SISTEM INFORMASI TGD*, Vol. 2, No. 1, pp. 65-72, 2023, <https://ojs.trigunadharma.ac.id/index.php/jsi/article/viewFile/5656/1978>
- [5] Firmansyah dan Agus Yulianto, “Prediksi Hasil Belajar Menggunakan Naïve Bayes Classifier pada Tingkat Sekolah Dasar”, *Remik: Riset dan E-Jurnal Manajemen Informatika Komputer*, Vol. 7, No. 2, pp. 1174-1182, 2023, <http://doi.org/10.33395/remik.v7i2.12375>
- [6] Simon Pranata Barus, “Implementation of Naïve Bayes Classifier-based Machine Learning to Predict and Classify New Students at Matana University”, *Journal of Physics: Conference Series*, Vol. 1842, pp. 1-7, 2021, <http://doi:10.1088/1742-6596/1842/1/012008>
- [7] M. A. Aditya, A. Helen, dan I. Suryana, “Naïve Bayes and Maximum Entropy Comparison for Translated Novel’s Genre Classification”, *Journal of Physics: Conference Series*, Vol. 1722(012007), 2021, <https://iopscience.iop.org/article/10.1088/1742-6596/1722/1/012007/pdf>
- [8] Ade Clinton Sitepu, Wanayumini, dan Zakarias Situmorang, “Determining Bullying Text Classification Using Naive Bayes Classification on Social Media”, *Jurnal Varian*, Vol. 4, No. 2, pp. 133-140, 2021, <https://journal.universitasbumigora.ac.id/index.php/Varian>
- [9] Roni Anagora, R. A., Rudini, R., Rohmat Taufiq, R. T., Ahmad Dedi Jubaedi, A. D. J., Rio Wirawan, R. W., & Arman Syah Putra, “The Classification of Phishing Websites using Naïve Bayes Classifier Algorithm”, *International Journal Of Science, Technology & Management*, Vol. 3, No. 2, pp. 553-562, 2022, <https://doi.org/10.46729/ijstm.v3i2.498>