EJECTS : E-Journal Computer, Technology and Informations System
Vol.2 No.2 Maret 2023, P-ISSN: 2988-5477 E-ISSN: 2988-4446
LPPM Universitas Darwan Ali

Data Mining Algoritma Naive Bayes Untuk Memprediksi Jumlah Siswa Baru

Ahmad Nasrullah

Program Studi Sistem Informasi, Universitas Darwan Ali

Email : ipsiarul@gmail.com.

ABSTRACT— MI Miftahussalam is an Ibtidaiyah Madrasah in the Hanau sub-district. Prediction is a tool in planning, especially in the field of education. In this world to know the circumstances that will come, to see the good or bad but also aims to make preparations. An important step after prediction is verification of the prediction so that the desired data for a problem can be obtained. Madrasas are institutions that systematically carry out teaching and learning programs in order to help students to be able to develop their potential both in terms of moral, religious, intellectual and social aspects. Therefore the criteria needed for admitting new MI Miftahussalam students are by collecting data on Smart Indonesia Cards (KIP), BPJS cards, parental income. This data will be grouped and obtained from the past period, processed into the system and training process with the Rapid Miner application.

Keywords— Data mining, naïve bayes, number of new students, KIP.

ABSTRAK— MI Miftahussalam adalah Madrasah Ibtidaiyah yang ada di kecamatan Hanau. Prediksi merupakan sebuah alat dalam perencanaan, khususnya dalam bidang pendidikan. Di dunia ini mengetahui keadaan yang akan dating, untuk melihat yang baik atau buruk tetapi juga bertujuan untuk melakukan persiapan. Langkah penting setelah prediksi adalah verifikasi prediksi sehingga dapat data yang diinginkan untuk sebuah persoalan. Madrasah merupakan lembaga yang sistematis melaksanakan program belajar mengajar dalam rangka membantu siswa agar mampu mengembangkan potensinya baik yang menyangkut aspek secara moral, agama, intelektual maupun sosial. Oleh karena itu kriteria yang diperlukan untuk penerimaan siswa baru MI Miftahussalam dengan menggumpulkan data Kartu Indonesia Pintar (KIP), kartu BPJS, Penghasilan orang tua data ini akan dikelompokan dan diperoleh dari periode lalu diolah kedalam sistem dan proses pelatihan dengan aplikasi Rapid Miner.

Kata kunci— Data mining, naïve bayes, jumlah mahasiswa baru, KIP.

I. PENDAHULUAN

MI Miftahussalam adalah Madrasah Ibtidaiyah yang ada di Kabupaten Seruyan. Mampu menerima siswa baru setiap tahunya. Pihak Madrasah mempunyai kendala dalam menerima siswa baru karena jumlah siswa yang mendaftar setiap tahunnya meningkat. Hal ini dikarenakan laju pertumbuhan penduduk yang terus meningkat. Untuk mengatasi masalah tersebut saya menggunakan Data mining dengan menggunakan metode Naive Bayes. Digunakan sebagai metode prediksi penerimaan siswa baru di MI Miftahussalam. Ada 5 peran data mining yaitu estimasi, prediksi, kalasifikasi, clustering, asosiasi. Naive Bayes merupakan sebuah pengklasifikasian probabilistik yang sederhana untuk menghitung sekumpulan probabilitas dengan cara menjumlahkan frekuensi yang ada dan kombinasi dari nilai dataset yang diberikan oleh data yang sudah ada [1]. (Alvino Dwi Rachman Prabowo) menerapkan naïve bayes dengan hasil bahwa algoritma PSO dapat digunakan sebagai algoritma feature selection yaitu dengan pembobotan atribut. Dari berawal pengolahan data dengan 16 atribut didapat nilai akurasi 82,19%, setelah dilakukan pembobotan atribut dengan algoritma PSO nilai akurasinya bertambah menjadi 89,70%.

Naive Bayes adalah metode yang dikembangkan oleh ilmuan inggris Bernama Thomas Bayes, gunanya memprediksi peluang dimasa depan berdasarkan pengalaman dimasa sebelumnya sehingga dikenal sebagai Teorema Bayes [2]. Karena itu kriteria yang diperlukan untuk penerimaan siswa baru MI Miftahussalam dengan menggumpulkan data siswa madrasah berdasarkan jarak antar rumah, usia, pernah sekolah Tk/Paud, Penerima Kip, Penghasilan orang tua data ini akan dikelompokan dan diperoleh dari data sebelumnya kemudian dimasukkan kedalam sistem dan diproses aplikasi Rapid Miner..

II. METODOLOGI PENELITIAN

A. Data mining

Data Mining adalah proses pengambilan informasi atau pola dari data yang besar dan kompleks. Proses ini melibatkan aplikasi teknik statistika, pembelajaran mesin, dan ilmu komputer untuk menemukan informasi yang berguna dari data. Data mining membantu untuk mengidentifikasi pola dan hubungan dalam data, mengekstrak informasi berguna, dan mengambil keputusan berdasarkan data tersebut[10].

Data Mining juga dapat didefinisikan sebagai proses mengidentifikasi dan mengekstrak informasi berguna dari data yang besar, bersifat kompleks, dan tidak terstruktur. Ini membutuhkan teknik dan algoritma yang berbeda

untuk mengurai dan memahami data, dan membantu untuk menemukan pola, hubungan, dan informasi berguna dalam data tersebut.

B. Klasifikasi

Naïve Bayes Classifier adalah metode klasifikasi atau teknik machine learning yang popular/umum digunakan dalam suatu klasifikasi teks, memiliki performa yang baik pada banyak domain, lebih sederhana dan sangat efisien. Namun Naïve Bayes sangat sensitive dalam pemilihan fitur, terlalu banyak jumlah dalam fitur dapat mengakibatkan meningkatnya waktu perhitungan dan menurunkan akurasi klasifikasi [3]. Klasifikasi Naive bayes berdasarkan pada teorema bayes dan dapat digunakan untuk menghitung suatu probabilitas tiap kelas dengan asumsi bahwa antar satu kelas dengan kelas lain tidak saling bergantung atau independent [4].

C. Naive Bayes

Naive Bayes merupakan pengklasifikasian dengan metode probabilitas dan statistik yang ditemukan oleh ilmuan inggris Bernama Thomas Bayes, gunanaya yaitu memprediksi peluang dimasa depan berdasarkan pengalaman sebelumnya sehingga dikenal sebagai Teoroma Bayes, Teoroma tersebut dikombinasikan dengan "naive" dimana bisa kita asumsikan kondisi antara atribut dan atribut lainnya saling bebas. Untuk menyelesaikan metode pada Naive Bayes dapat dilakukan dengan persamaan sebagai berikut [2];

$$P(H/X) = \frac{P(H/X)}{P(X)} \quad (1)$$

Dimana:
X : Data dengan class yang belum diketahui
H : Hipotesis data merupakan suatu spesifik.
P(H/X): Probabilitas hipotesis H berdsarkan kondisi x (posteriori probabilitas)
P(H) : Probabilitas hipotesis H (Prior Probabilitas)
P(X/H): Probabilitas X Berdasarkan kondisi pada Hipotesis H
P(X) : Probabilitas X
Penjelasan lebih lanjut rumus Bayes terebut dilakukan dengan menjabarkan
(C/XI.....Xn) menggunakan aturan perkalian sebagai berikut[8]:

$$\begin{aligned} P(C/x1,\ldots,xn &= P(C)\,P(x1,\ldots,xn|C) \\ &= P(C)P(X1|C)P(X2,\ldots,Xn|C,X1) \\ &= (C)P(X1|C)P(X2|C,X1)P(X3,\ldots,Xn|C,X1,X2 \\ &= (C)P(X1|C)P(X2|C,1)P(X3|C,X1,X2) \\ &\quad P(X4,\ldots,Xn|C,X1,X2,X3) \\ &= P(C)\ P\ (X1|C)P(X2|C,X1)P(X3|C,X1,X2) \\ &\quad (Xn|C,X1,X2,X3,\ldots,Xn\text{-}1 \end{aligned} \quad (2)$$

Dapat dilihat bahwa semakin banyak faktor-faktor yang kompleks dapat mempengaruhi nilai keseluruhan probabilitas, maka akan semakin mustahil untuk menghitung seluruh nilai tersebut satu persatu. Akibatnya melakukan perhitungan akan semakin sulit untuk dilakukan, disinilah digunakannya data asumsi, indepensi yang sangat tinggi, bahwa seluruh atribut-atribut dapat saling bebas seluruhnya. Dengan asumsi maka sangat diperlukan persamaan (3) :

$$P(Xi|c, Xj) = P(Xi|C) \ldots\ldots \quad (3)$$

Keterangan keempat (4) merupakan Teorema Bayes yang kemudian akan digunakan untuk melakukan perhitungan klasifikasi. Untuk klasifikasi data continue atau data angka akan menggunakan rumus distribusi Gaussian yaitu klasifikasi paling sederhana dalam metode naïve bayes dengan cara menjumlahkan dua parameter [5]: mean μ dan varian σ :

$$P(C|X1,\ldots\ldots, Xn = P(X1|C) \quad (4)$$

$$P(Xi = Xi\ |C = cj) = 1/\sqrt{2\pi\sigma ij}\ \exp\ (xi\text{-}\mu ij)2/2\sigma\sigma 2ij \quad (5)$$

Dimana : P : Peluang
Xi : Atribut ke i
Xj : Nilai atribut ke i
C : Kelas yang dicari
Ci : Sub kelas Y yang dicari
μ : Devinisi standar, menyatakan varian dari seluruh atribut

Dalam metode Naive Bayes juga diperlukan data latih dan data uji untuk diklasifikasikan, dalam Naive Bayes sebanyak mungkin data latih yang akan dilibatkan, semakin baik pula hasil yang bisa diprediksi dan yang diberikan. Menghitung P(Ci) yang merupakan salah satu probabilitas prior untuk setiap sub kelas C yang dihasilkan menggunakan persamaan [2][6]:

$$P(ci) = si/s \quad (6)$$

Dimana Si ialah jumlah data training kategori Ci, dan S adalah jumlah total data dari training[9]. Menghitung P(Xi|Ci) yang merupakan probabilitas dari posterior Xi dengan syarat C menggunakan Persamaan 4 (empat)[7].

III. DESAIN, HASIL DAN PEMBAHASAN

Hasil penelitian ini sesuai penelitian yang dilakukan. Data yang digunakan dalam penelitian ini adalah data yang diperoleh langsung dari pihak Madrasah untuk menentukan apakah siswa tersebut layak atau kurang layak di sekolah tersebut. Kriteria yang di tentukan yaitu Jarak antar Rumah, Usia siswa, pernah bersekolah TK/PAUD, mempunyai KIP atau tidak dan Penghasilan dari Orangtua. Selanjutnya data yang telah didapatkan akan ditransformasikan ke dalam format data excel 2019 dan dilakukan perhitungan manualnya. Data tersebut kemudian diuji dengan software Rapidminer versi 5.3. Sehingga jika dijumlahkan kita akan mengetahui berapa nilai dari klasifikasi Layak dan Tidak Layaknya. Proses pengolahan data yang dilakukan dalam penelitian ini diperoleh dengan melakukan proses perhitungan secara manual dengan menggunakan Ms. Excel 2019. Berikut ini ialah uraian dari perhitungan manual proses klasifikasi untuk menentukan siswa tersebut layak atau tidak layaknya diterima di MI Miftahussalam. Dan Dalam

proses pengklasifikasian Kelayakan pada siswa dilakukan dengan menentukan data yang akan digunakan pada penelitian ini. Berikut ini data yang akan digunakan pada penelitian ini dapat dilihat pada tabel I.

TABEL I
DATA PENELITIAN

No	Nama	Jarak Rumah Ke Sekolah	Umur	Tk/Paud	Punya KIP	Penghasilan Ortu	Hasil
1	RADIT	Kurang Dari 1 Km	7 Th	Ya	Punya	1.000.000	Layak
2	ARINIL MADINAH	Lebih Dari 1 Km	6,5 Th	Ya	Punya	1.000.000	Tidak Layak
3	YUNIY	Kurang Dari 1 Km	7 Th	Ya	Tidak	1.000.000	Layak
4	TITI AGUSTI	Kurang Dari 1 Km	6,5 Th	Tidak	Punya	1.000.000	Tidak Layak
5	AMILIYA	Kurang Dari 1 Km	7 Th	Ya	Tidak	1.500.000	Layak
6	SISKANAH	Kurang Dari 1 Km	7 Th	Ya	Punya	1.000.000	Layak
7	NELI ASTERINA	Kurang Dari 1 Km	7 Th	Ya	Punya	1.500.000	Layak
8	SHEILA SURYANI PUTRI AULIA	Kurang Dari 1 Km	7 Th	Ya	Tidak	2.000.000	Layak
9	NORIAH NABILA SALSABILA	Lebih Dari 1 Km	7 Th	Ya	Punya	1.500.000	Layak
10	LOLA ALMUHRIPAH	Kurang Dari 1 Km	6,5 Th	Tidak	Tidak	3.000.000	Layak
11	MIHA RAHMA DIANA	Lebih Dari 1 Km	7 Th	Ya	Tidak	3.000.000	Layak
12	AMITASYA	Kurang Dari 1 Km	7 Th	Ya	Punya	1.500.000	Layak
13	AYU RINTIANINGSIH	Lebih Dari 1 Km	7 Th	Ya	Tidak	1.000.000	Layak
14	IKA TUS TIANINGSIH	Lebih Dari 1 Km	6,8 Th	Tidak	Punya	1.500.000	Layak
...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...
34	NUR JANAH	Kurang Dari 1 Km	7 Th	Ya	Punya	1.000.000	Layak
35	REVA RAHAYU	Kurang Dari 1 Km	7 Th	Ya	Tidak	1.000.000	Layak
36	WILANDA FITRIANI	Kurang Dari 1 Km	6,8 Th	Ya	Punya	2.000.000	??
37	AMALIA	Lebih Dari 1 Km	6,8 Th	Tidak	Tidak	1.000.000	??
38	NADIA AYU PAMUNGKAS	Lebih Dari 1 Km	6,8 Th	Tidak	Punya	1.000.000	??
39	CHINTYA MUTIARA	Lebih Dari 1 Km	6,8 Th	Tidak	Tidak	2.000.000	??
40	IDYNDA PRASTIKA WESNIANTY	Kurang Dari 1 Km	6,5	Ya	Punya	3.000.000	??

Setelah data ditentukan, langkah selanjutnya saya menghitung jumlah klasifikasi Layak dan Tidak Layak berdasarkan tabel 1. Dari 35 data latih yang digunakan, diketahui kelas Layak sebanyak 28 data, dan kelas Tidak Layak sebanyak 7 data. Perhitungan probabilitas prior kemungkinan Layak dalam menentukan Kelayakan siswa baru dapat dilihat pada persamaan (6), yaitu : P(Layak) = 28/35 = 0,8, Sedangkan perhitungan probabilitas Tidak Layak yaitu : P(Tidak Layak) = 07/35 = 0,2.

Setelah probabilitas dari masing-masing prioriti telah kita ketahui, selanjutnya kita akan menghitung probabilitas dari setiap masing-masing kriteria yang akan digunakan. Kriteria yang digunakan penulis yaitu Jarak Rumah ke sekolah, Usia anak, pernah bersekolah di TK/PAUD, mempunyai KIP dan Penghasilan Orang Tua. Dalam menentukan setiap probabilitas dari setiap kriteria, saya menghitung bagian yang terdapat pada setiap kriteria yang ada pada penelitian. Dalam menentukan setiap probabilitas dari setiap kriteria dilakukan dengan menghitung jumlah Layak dan Tidak Layak pada bagian setiap kriteria-kriteria yang digunakan. Sehingga perhitungan dari probabilitas dari masing-masing kriteria dapat dilihat pada beberapa table-tabel berikut ini. Untuk menghitung probabilitas dari kemungkinan dari kriteria Jarak antar Rumah ke sekolah dapat dilihat pada tabel 2 sebagai berikut.

TABEL II
PROBABILITAS JARAK RUMAH

No	Jarak Rumah	Kejadian yang dipilih		Probabilitas	
		Layak	Tidak Layak	Layak	Tidak Layak
1	Kurang dari 1 km	20	6	0,7143	0,8571
2	Lebih dari 1 km	8	1	0,2857	0,1429
	Jumlah	28	7	1	1

Probabilitas pada kriteria Jarak Rumah yaitu pada kategori Layak pada jarak rumah kurang dari 1 km memiliki probabilitas 0,7143, dan Lebih dari 1 km memiliki probabilitas 0,2857. Sehingga jumlah probabilitas Layak yaitu 1. Sedangkan pada kategori Tidak Layak pada data kurang dari 1 km memiliki probabilitas 0,8571, dan Lebih dari 1 km memiliki probabilitas 0,1429. Sehingga jumlah probabilitas Layak yaitu 1. Untuk menghitung probabilitas kemungkinan dari kriteria Usia bisa dilihat pada table III.

TABEL III
PROBABILITAS TK/PAUD

No	TK/PAUD	Yang Dipilih		Probabilitas	
		Layak	Tidak Layak	Layak	Tidak Layak
1	YA	26	2	0,9286	0,2857
2	TIDAK	2	5	0,0714	0,7143
		28	7	1	1

Probabilitas keseluruhan pada kriteria pernah bersekolah di TK/PAUD yaitu pada kategori Layak pada data YA memiliki probabilitas 0,9286, Sedangkan TIDAK memiliki probabilitas sebesar 0,0714 Sehingga jumlah probabilitas Layak yaitu 1. Sedangkan pada kategori Tidak Layak pada data YA memiliki probabilitas 0,2857, TIDAK memiliki probabilitas 0,7143 Sehingga jumlah probabilitas Layak yaitu 1. Untuk menghitung probabilitas kemungkinan dari kriteria Penerima KIP dapat dilihat pada tabel IV.

TABEL IV
PROBABILITAS PENERIMA KIP

No	TK/PAUD	Yang Dipilih		Probabilitas	
		Layak	Tidak Layak	Layak	Tidak Layak
1	PUNYA	16	5	0,5714	0,7143
2	TIDAK	12	2	0,4286	0,2857
		28	7	1	1

Probabilitas pada kriteria Penerima KIP yaitu pada kategori Layak pada data Dapat memiliki probabilitas 0,5714, Tidak memiliki probabilitas 0,4286 Sehingga jumlah probabilitas Layak yaitu 1. Sedangkan untuk kategori Tidak Layak pada data memiliki probabilitas sebanyak 0,7143, dan Tidak memiliki probabilitas sebanyak 0,2857. Sehingga jumlah probabilitas Layak yaitu 1. Untuk menghitung probabilitas

kemungkinan dari kriteria Penghasilan Orang tua dapat dilihat pada tabel V dibawah ini.

TABEL V
PROBABILITAS PENGHASILAN ORANG TUA

No	Penghasilan ortu	yang dipilih		Probabilitas	
		Layak	Tidak Layak	Layak	Tidak Layak
1.	1.000.000	9	5	0,3214	0,7143
2.	1.500.000	7	0	0,2500	0
3.	2.000.000	8	2	0,2857	0,2857
4.	Rp.3.000.000	4	0	0,1429	0
	Jumlah	28	7	1	1

Probabilitas pada kriteria Penghasilan Orang tua pada kategori Layak pada Penghasilan Rp.1.000.000 memiliki Jumlah probabilitas sebanyak 0,3214, Rp.1.500.000 memiliki Jumlah probabilitas sebesar 0,2500, Rp.2.000.000 memiliki Jumlah Probabilitas sebesar 0,2857 dan Rp.3.000.000 memiliki Jumlah probabilitas sebanyak 0,1429. Sehingga jumlah seluruh probabilitas Layak yaitu 1.. Sedangkan pada kategori Tidak Layak pada Penghasilan Rp.1.000.000 memiliki Jumlah probabilitas sebanyak 0,7143, Rp.1.500.000 memiliki Jumlah probabilitas Sebesar 0, Rp.2.000.000 memiliki Jumlah probabilitas Sebanyak 0,2857 dan 3.000.000 memiliki jumlah probabilitas 0. Sehingga jumlah probabilitas Layak yaitu 1.

Setelah masing-masing dari probabilitas kriteria telah diketahui, langkah selanjutnya adalah menghitung nilai dari salah satu nilai yang diberikan responden untuk menentukan nilai klasifikasi. Berdasarkan data training pada tabel 1 pada data Siswa 36 sampai dengan 40 dilakukan klasifikasi ke dalam kelas Layak. Rumus yang digunakan dalam menentukan kelas Layak dapat dilihat pada persamaan (4). Sehingga untuk menghitung nilai Layak pada data responden 36 sampai dengan 40 adalah sebagai berikut:

P(36| Layak) = P(Jarak Rumah=Kurang dari 1 km|Layak) x P(Usia =6,8 thn|Layak) x P(TK/PAUD=YA|Layak) x P(KIP = Dapat |Layak) x P(Penghasilan Ortu = 2.000.000|Layak)= 0,7143 x 0,1429 x 0,9286 x 0,5714 x 0,2857 = 0,015

P(37| Layak) = P(Jarak Rumah=Lebih dari 1 km|Layak) x P(Usia =6,8 thn|Layak) x P(TK/PAUD=Tidak|Layak) x P(KIP =Tidak|Layak) x P(Penghasilan Ortu = 1.000.000|Layak)= 0,2857 x 0,1429 x 0,0714 x 0,4286 x 0,3214 = 0,0004

P(38| Layak) = P(Jarak Rumah=Lebih dari 1 km|Layak) x P(Usia =6,8 thn|Layak) x P(TK/PAUD=Tidak|Layak) x P(KIP = Dapat |Layak) x P(Penghasilan Ortu = 1.000.000|Layak)= 0,2857 x 0,1429 x 0,0714 x 0,5714 x 0,3214 = 0,0004

P(39| Layak) = P(Jarak Rumah=Lebih dari 1 km|Layak) x P(Usia =6,8 thn|Layak) x P(TK/PAUD=Tidak|Layak) x P(KIP =Tidak|Layak) x P(Penghasilan Ortu = 2.000.000|Layak) = 0,2857 x 0,1429 x 0,0714 x 0,4286 x 0,2857 = 0,0004

Sedangkan untuk menghitung nilai Tidak Layak pada data ke-36 sampai dengan 40 rumus yang digunakan sama dengan rumus untuk menentukan nilai Layak. Sehingga untuk mendapatkan nilai dilakukan sebagai berikut :

P(36| Tidak Layak)= P(Jarak Rumah=Kurang dari 1 km|Tidak Layak) x P(Usia =6,8 thn|Tidak Layak) x P(TK/PAUD=YA|Tidak Layak) x P(KIP = Dapat |Tidak Layak) x P(Penghasilan Ortu = 2.000.000|Tidak Layak)= 0,8571 x 0,4286 x 0,2857 x 0,7143 x 0,2857 = 0,0214

P(37| Tidak Layak) = P(Jarak Rumah=Lebih dari 1 km|Tidak Layak) x P(Usia =6,8 thn|Tidak Layak) x P(TK/PAUD=Tidak|Tidak Layak) x P(KIP =Tidak|Tidak Layak) x P(Penghasilan Ortu = 1.000.000|Tidak Layak)= 0,1429 x 0,4286 x 0,7143 x 0,2857 x 0,7143 = 0,0089

Setelah menilai pada klasifikasi Layak dan Tidak Layak pada total seluruh data 36 siswa sampai dengan 40 siswa telah diketahui. Selanjutnya penulis mulai menghitung nilai maksimal dari masing-masing seluruh klasifikasi. Perhitungan data yang didapat dari responden 36 sampai dengan 40 untuk menghitung pemaksimalan nilai Layak yaitu:

P(Layak|C) = P(Rn|C) * P(Layak)
= P(36|C) * P(Layak)
= 0,0155 x 0.8 = 0,0124

Setelah menghitung maksimal dari nilai Layak dan Tidak Layak, selanjutnya penulis akan membandingkan nilai Layak dan Tidak Layak. Sehingga dapat diketahui siswa tersebut termasuk kedalam kategori Layak atau Tidak Layak.

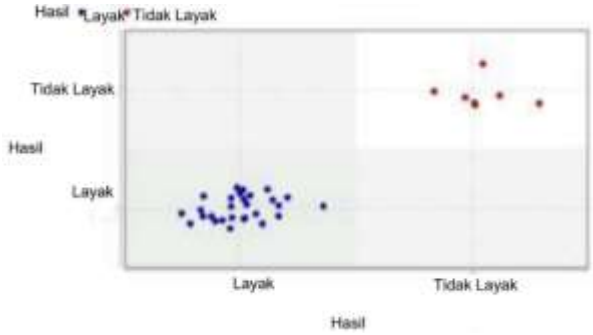
R101 = Layak >= Tidak Layak
= 0,0124 >= 0,0043
= 0,0124 (Layak).
R102 = Layak >= Tidak Layak
= 0,0003 >= 0,0018
= 0,0018 (Tidak Layak).
R102 = Layak >= Tidak Layak
= 0,0003 >= 0,0018
= 0,0018 (Tidak Layak).

Sehingga dari data perbandingan tersebut dapat diketahui bahwa data testing dari data siswa 36 dan 40 memiliki klasifikasi Layak Sedangkan dari data siswa 37,38 dan 39 memiliki Tidak Layak Pada tahap selanjutnya akan dilakukan pengujian data menggunakan tools rapidminer. Hasil akhir yang akan ditampilkan adalah berupa SimpleDistrubition yaitu menentukan banyaknya nilai dari data Layak dan Tidak Layak. Dapat

dilihat pada gambar 1. Berdasarkan gambar 1 menjelaskan bahwa kelas Layak memiliki nilai klasifikasi/probabilitas 0,800 sedangkan kelas Tidak Layak mendapatkan nilai klasifikasi/probabilitas 0,200. Sehingga berdasarkan data hasil dari klasifikasi pada gambar 2 didapatkan grafik pada hasil RapidMiner 1 berikut ini.

Gambar 1. Hasil Akhir

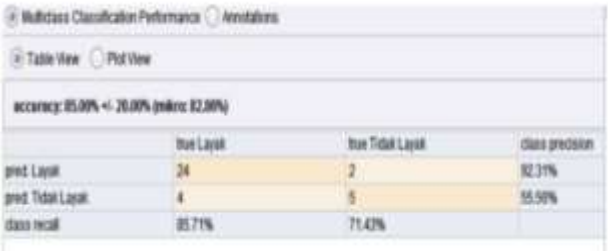
Berdasarkan pada gambar 2. dapat diketahui bahwa pada gambar titik berwarna merah (Tidak Layak) memiliki node total sebanyak 7, sedangkan gambar pada titik berwarna biru (Layak) memiliki node total sebanyak 28. Akurasi Hasil pengujian Model Algoritma Naive Bayes Classfier ditunjukan pada gambar berikut:



Gambar 2. Grafik klasifikasi

Keterangan :

a) Jumlah prediksi Layak dan kenyataannya benar Layak adalah 24 record (FN).
b) Jumlah prediksi untuk yang Layak dan kenyataannya benar Tidak Layak adalah 2
c) record (TN).
d) Jumlah prediksi kurang Tidak Layak dan kenyataannya benar Layak adalah 4
e) record (FP).
f) Jumlah prediksi Tidak Layak dan kenyataannya benar Tidak Layak adalah 5 record
g) (TP).
h) Pada gambar diatas dapat dilihat Nilai Accuracy sebesar 85.00 %. class precision
i) pada prediksi Layak memiliki nilai 92.31%, sedangkan pada prediksi Tidak Layak
j) memiliki nilai 55,56%. Class recall pada true Layak
k) memiliki nilai 85,71%, sedangkan pada true Tidak Layak memiliki nilai 71,43 %.



Gambar 3. Nilai Accuracy Performance

IV. KESIMPULAN

Berdasarka hasil dari penelitian yang dilakukan maka diperoleh kesimpulan sebagai berikut:

1) Penerapan data mining dalam menentuka klasifikasi penerimaan siswa baru dapat
2) digunakan untuk memprediksi layak atau tidak layak seorang siswa dapat diterima dengan menggunakan algoritma naïve bayes.
3) Algoritma naïve bayes sangat cocok dterapkan dalam memprediksi peluang dimasa depan berdasarkan pengalaman dimasa sebelumnya memudahkan dalam penerimaan siswa baru.
4) Metode naïve bayes memanfaatkan data training untuk menghasilkan probabilitas setiap kriteria untuk class yang berbeda, sehingga nilai-nilai probabilitas dari kriteria tersebut dapat dioptimalkan untuk memprediksi kelayakan penerimaan siswa baru berdasarkan proses klasifikasi yang dilakukan oleh metode naïve bayes itu sendiri
5) Penelitian ini hanya menggunakan data umum untuk mendapatkan hasil yang lebih spesifik disarankan untuk penelitian selanjutnya untuk memaparkan variable yang lebih jelas.
6) Perlu dikembangkan penelitian yang lebih mendalam dan variable algoritma pelatihan supaya mendapatkan hasil yang lebih optimal dengan waktu pelatihan yang lebih singkat lagi.

V. REFERENSI

[1] V. Rizqiyani, A. Mulwinda, And R. Defi Mahadji Putri, “Klasifikasi Judul pada Buku
Algoritma Naive Bayes Dan Pencarian Buku Pada Perpustakaan Jurusan Teknik Elektro,” J. Tek. Elektro, Vol. 9, No. 2, Pp. 60–65, 2017.

[2] T. Imandasari, E. Irawan, A. P. Windarto, And A. Wanto, “Algoritma Naive Bayes
Dalam Klasifikasi Lokasi Pembangunan Sumber Mata Air,” Semin. Nas. Ris. Inf. Sci.,
2019, Doi: 10.30645/Senaris.V1i0.81.

[3] D. A. Muthia, "Opinion Mining pada Review Buku Menggunakan Algoritma Naive

Bayes,"JurnalTeknik Komputer Amik BSI, Vol.II, No.1,pp. 1-8, 2016.

[4] J. Eska, “Penerapan Data Mining Untuk Prediksi Penjualan Wallpaper Menggunakan Algoritma C4.5,” Jurnal Teknologi. Dan Sistem Informasi), Vol. ke 2, 2018, Doi: 10.31227/Osf.Io/X6svc.

[5] Dan G. F. Ade Bastian, Harun Sujadi, “Penerapan Algoritma K-Means Clustering Analysis Penyakit Menular Pada Manusia (Studi Kasus di Kabupaten Majalengka),” No. 1, Pp.26–32, 2018.

[6] R. W. Sari And D. Hartama, “Data Mining : Algoritma K-Means Pada Pengelompokkan Wisata Asing Ke Indonesia Berdasarkan Provinsi,” Semin. Nas. SainsTeknol. Inf., Pp. 322–326, 2018.

[7] H. Sulastri And A. I. Gufroni, “Jurnal Teknologi Dan Sistem Informasi Penerapan Data Mining Dalam Pengelompokan Penderita,” Vol. 02, Pp. 299–305, 2017.

[8] T. Khotimah, “Pengelompokan Surat Didalam Al Qur’an cara Menggunakan Algoritma K-Means,” Simetris, Vol. 5, No. 1, Pp. 83–88, 2014

[9] Ahmad, Abu. (2017). Mengenal Artificial Intelegence, Machine Learning, Neural Network, dan Deep Learning.

[10] Darmawan, Rudi. (2008). Modul Teknologi Informasi dan Komunikasi. Solo: Hayati.