KESATRIA: Jurnal Penerapan Sistem Informasi (Komputer & Manajemen)
Vol. 2, No. 2, April (2021), pp. 108-117

Teknik Data Mining Dalam Prediksi Jumlah Siswa Baru Dengan Algoritma Naive Bayes

Aston Maruli Sitompul[1], Suhada[2], Saifulah[3]
[1,3] *STIKOM Tunas Bangsa, Pematangsiantar, Sumatera Utara, Indonesia*
[2]*AMIK Tunas Bangsa, Pematangsiantar, Sumatera Utara, Indonesia*
Jln. Sudirman Blok A No. 1-3 Pematangsiantar, Sumatera Utara
[1]*astonsitompul66@gmail.com,* [2]*suhada.atb@gmail.com,*
[3]*saiful_siantar@yahoo.com*

Abstract

Sukosari Public Elementary School 095126 is an elementary school located in Gunung Malela sub district, Simalungun Regency. Prediction is an important tool in planning effective and efficient, especially in the field of education. In the modern world knowing the situation to come is not only important to see the good or bad but also aims to make forecast preparation. An important step after a prediction has been made is to verify the prediction in such a way that it reflects past data and the underlying system that underlies the request. As long as the prediction representation can be trusted, the predicted results can continue to be used. Schools are formal educational institutions that systematically carry out guidance, teaching, and training programs in order to help students to be able to develop their potential both in terms of moral, spiritual, intellectual, emotional and social aspects. The inference method used is the Naive Bayes method, the system will input student data such as the criteria needed for admission of new students at SD N 095126 Sukasari by collecting data on Indonesia Smart Cards (KIP), BPJS cards, Parent income, these data will be grouped and obtained from the past period is processed into the system and training process with the Rapid Miner application.

Keywords: *Sukosari 095126 Elementary School Students, Prediction, Data mining, Naive Bayes*

Abstrak

SD Negeri Sukosari 095126 adalah sekolah dasar yang berada di kecamatan gunung malela Kabupaten Simalungun. Prediksi merupakan alat bantu yang penting dalam perencanaan yang efektif dan efisien khususnya dalam bidang pendidikan. Dalam dunia modern mengetahui keadaan yang akan datang tidak saja penting untuk melihat yang baik atau buruk tetapi juga bertujuan untuk melakukan persiapan peramalan. Langkah penting setelah prediksi dilakukan adalah verifikasi prediksi sedemikian rupa sehingga mencerminkan data masa lalu dan sistem penyebab yang mendasari permintaan tersebut. Sepanjang representasi prediksi tersebut dapat dipercaya, hasil prediksi dapat terus digunakan. Sekolah merupakan lembaga pendidikan formal yang sistematis melaksanakan program bimbingan, pengajaran, dan latihan dalam rangka membantu siswa agar mampu mengembangkan potensinya baik yang menyangkut aspek moral, spiritual, intelektual, emosional maupun sosial. Metode inferensi yang digunakan adalah metode naive bayes sistem akan menginput data siswa seperti Oleh karena itu kriteria yang diperlukan untuk penerimaan siswa baru SD N 095126 Sukasari 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.

Keywords: *Siswa SDN Sukosari 095126, Prediksi, Data mining, Naive Bayes*

ISSN: 2720-992X

1. Pendahuluan

Dengan berkembangnya teknologi saat ini hampir semua perangkat komputer mampu menghasilkan sistem yang lebih handal. Di masa mendatang diperkirakan semua perangkat elektronik dan komputer akan jauh lebih cerdas dan mampu mengembangkan aplikasi – aplikasi yang dapat diaplikasikan dalam kehidupan nyata. Bisa digunakan dalam memprediksi hal – hal yang akan terjadi. Salah satunya Data mining yang dapat memprediksi sesuatu yang akan terjadi dengan metode *naive bayes.* Naive Bayes merupakan teknik prediksi yang bebasis probabilistik sederhana berdasarkan penerapan teorema Bayes menggukan asumsi ketidak tergantungan yang kuat. Model fitur independen adalah metode yang digunakan naive bayes.

SD Negeri Sukosari 095126 adalah sekolah dasar yang terletak di kabupaten simalungun. Mampu menerima siswa baru setiap tahunya dengan kapasitas tertentu. Pihak guru memiliki kendala dalam penerimaan siswa baru karena jumlah siswa yang mendaftar pada SD Negeri Sukosari 095126 meningkat setiap tahunnya hal ini dikarenakan laju pertumbuhan penduduk yang terus meningkat. Melihat kondisi ini pihak sekolah tidak memiliki persediaan vasilitas lebih untuk penambahan penerimaan siswa baru.

Untuk mengatasi masalah yang kompleks penulis menggunakan jaringan Data mining dengan menggunakan metode *Naive Bayes.* Digunakan sebagai metode prediksi penerimaan siswa baru di SD Negeri Sukosari 095126. Ada 5 peran data mining yaitu estimasi, prediksi, kalasifikasi, *clustering*, asosiasi. *Naive Bayes* merupakan merupakan sebuah pengklasifikasian probabilistik sederhana yang menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari dataset yang diberikan [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 Thomas Bayes, yaitu memprediksi peluang dimasa depan berdasarkan pengalaman dimasa sebelumnya sehingga dikenal sebagai *Teorema Bayes* [2]. Oleh karena itu kriteria yang diperlukan untuk penerimaan siswa baru SD N 095126 Sukasari dengan menggumpulkan data siswa berdasarkan jarak rumah, usia, Tk/Paud, Penerima Kip, Penghasilan orang tua data ini akan dikelompokan dan diperoleh dari periode lalu diolah kedalam sistem dan proses pelatihan dengan aplikasi Rapid Miner. Dengan menggunakan *Naive Bayes* ini diharapkan dapat memberikan alternative lain dalam prediksi penerimaan siswa baru di SD Negeri Sukosari 095126 setiap tahunya.

2. Metodologi Penelitian

2.1. Data mining

Data mining ialah proses menganalisa data dari persfektif dan menyimpulkannya menjadi informasi-informasi penting yang dapat dipakai untuk meningkatkan keuntungan, memeperkecil biaya pengeluaran, atau bahkan kedua nya (Saleh,2015). Selain pengertian di atas beberapa definisi juga diberikan seperti. *Data mining* atau penambangan data dapat didefinisikan sebagai proses seleksi, eksplorasi, dan pemodelan dari sejumlah besar data untuk menemukan pola atau kecenderungan yang biasanya tidak didasarin keberadaannya (Soepomo, 2014).

2.2. Klasifikasi

Klasifikasi Naive bayes ialah klasifikasi berdasarkan teorema bayes dan digunakan untuk menghitung probabilitas tiap kelas dengan asumsi bahwa antar satu kelas dengan

kelas lain tidak saling tergantung atau independent (Nugroho &Subanar 2013). Komponen-komponen utama dari proses klasifikasi antara lain (Rodiyansyah, 2014):

a) Kelas, yaitu merupakan variabel tidak bebas yang merupakan dari hasil klasifikasi.
b) Prediktor, merupakan variable bebas suatu model berdasarkan dari karakteristik atribut data yang diklasifikasi.
c) Set data pelatihan, yaitu sekumpulan data lengkap yang berisi kelas dan predictor untuk dilatih agar model dapat mengelompokan ke dalam kelas yang tepat.
d) Set data uji, berisi data-data baru yang akan dikelompokan oleh model guna mengetahui akurasi dari model yang telah dibuat.

2.3. *Naive Bayes*

Naive Bayes ialah sebuah pengklasifikasian probabilitas dengan menjumlahkan frekuensi dan kombinasi dari nilai dataset yang diberikan (Manalu, Sianturi & Manalu, 2017). Naive Bayes merupakan pengklasifikasian dengan metode probabilitas dan statistik yang dikemukankan oleh ilmuan inggris Thomas Bayes, yaitu memprediksi peluang dimasa depan berdasarkan pengalaman sebelumnya senhingga dikenal sebagai Teoroma Bayes, Teoroma tersebut dikombinasikan dengan "naive"dimana diasumsikan kondisi antara atribut saling bebas.

Kelebihan algoritma Naive Bayes Casifer lebih mudah untuk digunakan karena hanya memiliki alur perhitungan yang tidak panjang, hanya memerlukan sejumlah kecil data pelatihan untuk mengestimasi parameter (rata-rata variasi variabel) yang untuk klasifikasi, dan tokoh terhadap atribut yang tidak relavan. Untuk menyelesaikan metode Naive Bayes dapat dilakukan dengan persamaan-persamaan sebagai berikut (Fadlan, Ningsih, & Windarto, 2018) :

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

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

Penjabaran lbih lanjut rumus Bayes terebut dilakukan dengan menjabarkan (C/XI....Xn) menggunakan aturan perkalian sebagai berikut

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

Dapat diihat bahwa semakin banyak faktor-faktor yang semakin kompleks yang mempengaruhi nilai probabilitas, maka semakin mustahil untuk menghitung nilai tersebut satu persatu. Akibatnya perhitungan semakin sulit untuk dilkukan, disinilah digunakan asumsi, indepensi yang sangat tinggi, bahwa masing-masing atribut dapat saling bebas. Dengan asumsi tersebut, diperlukan persamaan (3) :

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

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

Keterangan empat (4) merupakan Teorema Bayes yang kemudian akan digunakan untuk melakukan perhitungan klasifikasi. Untuk klasifikasi dengan data continue

atau data angka mengunakan rumus distribusi Gaussian dengan dua parameter : mean μ dan varian σ :

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

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 seluruh atribut

Dalam metode Naive Bayes diperlukan data latih dan data uji yang ingin diklasifikasikan, dalam Naive Bayes sebanyak data latih yang dilibatkan, semakin baik hasil yang prediksi yang diberikan. Menghitung P(Ci) yang merupakan probabiitas prior untuk setiap sub kelas C yang akan dihasilkan menggunakan persamaan :

$$P(ci) = \frac{si}{s} \quad (6)$$

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

3. Hasil dan Pembahasan

Hasil penelitian ini disajikan sesuai penelitian yang dilakukan. Data yang digunakan dalam penelitian ini adalah data yang diperoleh langsung dari pihak sekolah untuk menentukan apakah siswa tersebut layak atau kurang layak di sekolah tersebut. Kriteria yang di tentukan yaitu Jarak Rumah, Usia, TK/PAUD, KIP dan Penghasilan Ortu. Selanjutnya data yang telah didapatkan ditransformasikan ke format data *excel* 2010 dan dilakukan perhitungan manualnya. Kemudian data tersebut diuji dengan *software Rapidminer versi* 5.3. Sehingga dapat diketahui berapa nilai klasifikasi Layak dan Tidak Layak.

Proses pengolahan data yang dilakukan dalam penelitian ini adalah dengan cara melakukan proses perhitungan manual dengan Ms. Excel 2010. Berikut uraian perhitungan manual proses klasifikasi dalam menentukan apakah siswa tersebut layak atau tidak layak untuk diterima di SD N 095126 sukosari. Dalam proses pengklasifikasian Kelayakan siswa dilakukan dengan menentukan data yang digunakan dalam penelitian ini. Berikut data yang digunakan dalam penelitian ini dapat dilihat pada tabel 1.

Tabel 1. Data Penelitian

No	Nama	Jarak Rumah	Usia	Tk/Paud	Penerima Kip	Penghasiln Ortu	Hasil
1	Walia Anesia	Kurang Dari 1 Km	7 Thn	Ya	Dapat	1.000.000	Layak
2	Nando Saputra	Lebih Dari 1 Km	6,5 Thn	Ya	Dapat	1.000.000	Tidak Layak
3	Prasetio	Kurang Dari 1 Km	7 Thn	Ya	Tidak	1.000.000	Layak
4	Rindi Artika	Kurang Dari 1 Km	6,5 Thn	Tidak	Dapat	1.000.000	Tidak Layak
5	Dewi Pratiwi	Kurang Dari 1 Km	7 Thn	Ya	Tidak	1.500.000	Layak
6	Arya Nanda Prasetiyo	Kurang Dari 1 Km	7 Thn	Ya	Dapat	1.000.000	Layak
7	Aldi	Kurang Dari 1 Km	7 Thn	Ya	Dapat	1.500.000	Layak
8	Anggun Sri Fatmalasari	Kurang Dari 1 Km	7 Thn	Ya	Tidak	2.000.000	Layak

No	Nama	Jarak Rumah	Usia	Tk/Paud	Penerima Kip	Penghasiln Ortu	Hasil
9	Andika Fahruzi Raja Gukguk	Lebih Dari 1 Km	7 Thn	Ya	Dapat	1.500.000	Layak
10	Andi	Kurang Dari 1 Km	6,5 Thn	Tidak	Tidak	3.000.000	Layak
11	Bagas	Lebih Dari 1 Km	7 Thn	Ya	Tidak	3.000.000	Layak
12	Dwi Cancri Handayani	Kurang Dari 1 Km	7 Thn	Ya	Dapat	1.500.000	Layak
13	Elfira Sandra	Lebih Dari 1 Km	7 Thn	Ya	Tidak	1.000.000	Layak
14	Ferdi Vilanuansa	Lebih Dari 1 Km	6,8 Thn	Tidak	Dapat	1.500.000	Layak
…	…	…	…	…	…	…	…
…	…	…	…	…	…	…	…
34	Zahira	Kurang Dari 1 Km	7 Thn	Ya	Dapat	1.000.000	Layak
35	Zaenab	Kurang Dari 1 Km	7 Thn	Ya	Tidak	1.000.000	Layak
36	Rakha Pratama	Kurang Dari 1 Km	6,8 Thn	Ya	Dapat	2.000.000	??
37	Denny Firmansyah	Lebih Dari 1 Km	6,8 Thn	Tidak	Tidak	1.000.000	??
38	Sari Dewi	Lebih Dari 1 Km	6,8 Thn	Tidak	Dapat	1.000.000	??
39	Dinda Anjani	Lebih Dari 1 Km	6,8 Thn	Tidak	Tidak	2.000.000	??
40	Risky Pratama	Kurang Dari 1 Km	6,5 Thn	Ya	Dapat	3.000.000	??

Setelah data telah ditentukan, langkah selanjutnya penulis 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 :

$$P(Layak) = \frac{28}{35} = 0{,}8$$

Sedangkan perhitungan probabilitas Tidak Layak yaitu :

$$P(Tidak\ Layak) = \frac{7}{35} = 0{,}2$$

Setelah probabilitas dari masing-masing prior telah diketahui, selanjutnya penulis menghitung masing-masing probabilitas dari setiap kriteria yang digunakan. Kriteria yang digunakan penulis yaitu Jarak Rumah, Usia, TK/PAUD, KIP dan Penghasilan Ortu. Dalam menentukan probabilitas setiap kriteria, penulis menghitung bagian-bagian yang terdapat pada setiap kriteria, pada penelitian. Sehingga dalam menentukan probabilitas setiap kriteria dilakukan dengan menghitung jumlah Layak dan Tidak Layak pada bagian-bagian setiap kriteria yang digunakan. Sehingga perhitungan probabilitas masing-masing kriteria dapat dilihat pada beberapa tabel-tabel berikut. Untuk menghitung probabilitas kemungkinan dari kriteria Jarak Rumah dapat dilihat pada tabel 2 sebagai berikut.

Tabel 2. 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 dapat dilihat pada tabel 3 sebagai berikut.

Tabel 3. Probabilitas Usia

No	Usia	Kejadian yang dipilih		Probabilitas	
		Layak	Tidak Layak	Layak	Tidak Layak
1	6,5 thn	2	4	0,0714	0,5714
2	6,8 thn	4	3	0,1429	0,4286
3	7 thn	22	0	0,7857	0
	Jumlah	28	7	1	1

Probabilitas pada kriteria Usia yaitu pada kategori Layak pada usia 6,5 tahun memiliki probabilitas 0,0714, 6,8 thn memiliki probabilitas 0,1429 dan 7 thn memiliki probabilitas 0,7857. Sehingga jumlah probabilitas Layak yaitu 1. Sedangkan pada kategori Tidak Layak pada pada usia 6,5 tahun memiliki probabilitas 0,5714, 6,8 thn memiliki probabilitas 0,4286 dan 7 thn memiliki probabilitas 0. Sehingga jumlah probabilitas Layak yaitu 1. Untuk menghitung probabilitas kemungkinan dari kriteria TK/PAUD dapat dilihat pada tabel 4.

Tabel 4. Probabilitas TK/PAUD

No	TK/PAUD	Kejadian 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
	Jumlah	28	7	1	1

Probabilitas pada kriteria TK/PAUD yaitu pada kategori Layak pada data YA memiliki probabilitas 0,9286, TIDAK memiliki probabilitas 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 5.

Tabel 5. Probabilitas Penerima KIP

No	KIP	Kejadian yang dipilih		Probabilitas	
		Layak	Tidak Layak	Layak	Tidak Layak
1	Dapat	16	5	0,5714	0,7143
2	Tidak	12	2	0,4286	0,2857
	Jumlah	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 pada kategori Tidak Layak pada data Dapat memiliki probabilitas 0,7143, Tidak memiliki probabilitas 0,2857. Sehingga jumlah

probabilitas Layak yaitu 1. Untuk menghitung probabilitas kemungkinan dari kriteria Penghasilan Ortu dapat dilihat pada tabel 6.

Tabel 6. Probabilitas Penghasilan Ortu

No	Penghasilan ortu	Kejadian 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	3.000.000	4	0	0,1429	0
	Jumlah	28	7	1	1

Probabilitas pada kriteria Penghasilan Ortu yaitu pada kategori Layak pada Penghasilan 1.000.000 memiliki probabilitas 0,3214, 1.500.000 memiliki probabilitas 0,2500, 2.000.000 memiliki probabilitas 0,2857 dan 3.000.000 memiliki probabilitas 0,1429. Sehingga jumlah probabilitas Layak yaitu 1.. Sedangkan pada kategori Tidak Layak pada Penghasilan 1.000.000 memiliki probabilitas 0,7143, 1.500.000 memiliki probabilitas 0, 2.000.000 memiliki probabilitas 0,2857 dan 3.000.000 memiliki probabilitas 0. Sehingga jumlah probabilitas Layak yaitu 1.
Setelah masing-masing 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 = 0,0155

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 nilai Layak dan Tidak Layak pada data 36 sampai dengan 40 telah diketahui. Selanjutnya penulis melakukan perhitungan maksimal masing-masing klasifikasi. Perhitungan data 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

Sedangkan perhitungan maksimal nilai Tidak Layak pada data responden 36 sampai dengan 40 yaitu :

(Tidak Layak|C) = P(Rn|C) * P(Tidak Layak)
= P(36|C) * P(Tidak Layak)
= 0,0214 x 0.8 = 0,0043

Setelah menghitung pemaksimalan dari nilai Layak dan Tidak Layak, selanjutnya penulis 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).

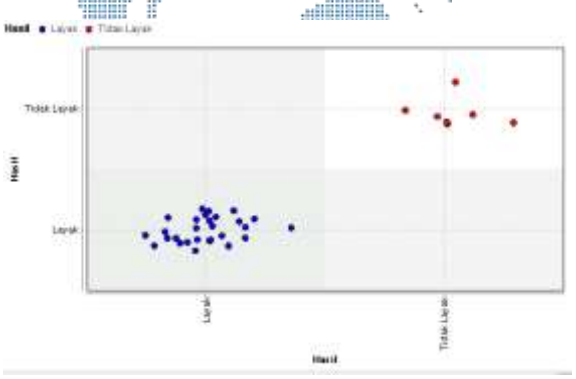
R102 = Layak >= Tidak Layak
= 0,0003 >= 0,0018
= 0,0018 (Tidak Layak).

R103 = Layak >= Tidak Layak
= 0,0003 >= 0,0018
= 0,0018 (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.

Gambar 1. Hasil Akhir

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 klasifikasi pada gambar 2 didapatkan grafik hasil dari *RapidMiner* 5.3 berikut ini :



Gambar 2. Grafik klasifikasi

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

Multiclass Classification Performance | Annotations

Table View | Plot View

accuracy: 85.00% +/- 20.00% (mikro: 82.86%)

	true Layak	true Tidak Layak	class precision
pred. Layak	24	2	92.31%
pred. Tidak Layak	4	5	55.56%
class recall	85.71%	71.43%	

Gambar 3. Nilai *Accuracy Performance*

Keterangan :

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

4. Kesimpulan

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

a) Penerapan data mining dalam menentuka klasifikasi penerimaan siswa baru dapat digunakan untuk memprediksi layak atau tidak layak seorang siswa dapat diterima dengan menggunakan algoritma naïve bayes.

b) Algoritma naïve bayes sangat cocok diterapkan dalam memprediksi peluang dimasa depan berdasarkan pengalaman dimasa sebelumnya memudahkan dalam penerimaan siswa baru.
c) 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.
d) Penelitian ini hanya menggunakan data umum untuk mendapatkan hasil yang lebih spesifik disarankan untuk penelitian selanjutnya untuk memaparkan variable yang lebih jelas.
e) Perlu dikembangkan penelitian yang lebih mendalam dan variable algoritma pelatihan supaya mendapatkan hasil yang lebih optimal dengan waktu pelatihan yang lebih singkat lagi.

Daftar Pustaka

[1] V. Rizqiyani, A. Mulwinda, And R. Defi Mahadji Putri, "Klasifikasi Judul Buku Dengan 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 Air," *Semin. Nas. Ris. Inf. Sci.*, 2019, Doi: 10.30645/Senaris.V1i0.81.
[3] J. Eska, "Penerapan Data Mining Untuk Prediksi Penjualan Wallpaper Menggunakan Algoritma C4.5," *Jurteksi (Jurnal Teknol. Dan Sist. Informasi)*, Vol. 2, 2018, Doi: 10.31227/Osf.Io/X6svc.
[4] Dan G. F. Ade Bastian, Harun Sujadi, "Penerapan Algoritma K-Means Clustering Analysis Pada Penyakit Menular Manusia (Studi Kasus Kabupaten Majalengka)," No. 1, Pp. 26–32, 2018.
[5] R. W. Sari And D. Hartama, "Data Mining : Algoritma K-Means Pada Pengelompokkan Wisata Asing Ke Indonesia Menurut Provinsi," *Semin. Nas. Sains Teknol. Inf.*, Pp. 322–326, 2018.
[6] H. Sulastri And A. I. Gufroni, "Jurnal Teknologi Dan Sistem Informasi Penerapan Data Mining Dalam Pengelompokan Penderita," Vol. 02, Pp. 299–305, 2017.
[7] T. Khotimah, "Pengelompokan Surat Dalam Al Qur'an Menggunakan Algoritma K-Means," *Simetris*, Vol. 5, No. 1, Pp. 83–88, 2014.