

OPTIMASI METODE NAÏVE BAYES DENGAN FEATURE SELECTION INFORMATION GAIN UNTUK PREDIKSI KETERLAMBATAN PEMBAYARAN SUMBANGAN PEMBINAAN PENDIDIKAN SEKOLAH

Muqorobin¹⁾, Kusrini²⁾, Emha Taufiq Luthfi³⁾

¹⁾ Program Pascasarjana Teknik Informatika, Universitas Amikom Yogyakarta

^{2,3)} Program Pascasarjana, Universitas Amikom Yogyakarta

¹⁾robbyaullah@gmail.com, ²⁾kusrini@amikom.ac.id, ³⁾emhataufiqluthfi@amikom.ac.id

ABSTRACT

The cost of education is one component of input that is very important in implementing education. Because costs are the main requirement in an effort to achieve educational goals. SMK Al-Islam Surakarta is a private education institution that requires students to pay school fees in the form of Education Development Donations. Educational Development Donation is a routine school fee that is conducted every month. Based on last year's TU report, many students were late in paying Education Development Donations, around 60%. This is a big problem. The purpose of this study is that researchers will build a predictive system using the Naïve Bayes method. Because the method can classify the class right or late, in the payment of school fees. Data processing was taken from the dapodik data of schools in 2017/2018 with the test dataset taking 30 records. To find out the level of accuracy, this research was conducted with the Naïve Bayes Method and the Information Gain Method for feature selection. Accuracy testing is done by the Confusion matrix method. The results showed that the highest accuracy was obtained by combining the Naïve Bayes Method with the Information Gain Method obtained by 90% accuracy.

Keywords: Cost, Naïve Bayes, Information Gain, Confusion Matrix

I. PENDAHULUAN

Pendidikan merupakan salah satu kewajiban yang harus ditempuh oleh setiap anak bangsa Indonesia. Hal ini berdasarkan aturan dari Permen Dikbud No.19 tahun 2016 tentang Program Indonesia Pintar. Dalam pasal 2 menyebutkan Setiap anak layak untuk mendapatkan layanan pendidikan sampai selesai dari satuan menengah pendidikan serta dalam rangka mendukung pelaksanaan pendidikan menengah rintisan wajib belajar 12 tahun. maka dengan hal ini Program Indonesia Pintar bertujuan untuk meningkatkan akses belajar bagi anak usia 6 sampai 21 tahun.

Biaya pendidikan adalah salah satu masukan pendukung dalam kegiatan penyelenggaraan pendidikan. Biaya berperan penting untuk mencapai tujuan pendidikan baik dari sekolah atau universitas. Dalam hal ini biaya dibagi menjadi 2 yaitu biaya pendidikan secara keseluruhan dan biaya per siswa. Untuk biaya per siswa bersifat wajib yang umumnya dikenal dengan Sumbangan Pembinaan Pendidikan (SPP). Biaya SPP ini umumnya diterapkan oleh sekolah swasta yang dibebankan pada setiap siswa. Karena sekolah swasta dalam mengelola pendidikan dibebankan oleh masyarakat atau kebijakan lokal. Berbeda dengan sekolah Negeri yang biaya pengelolaan sekolah masih ada bantuan dari biaya pemerintah (Sanjiwani & Ayu, 2012).

Sekolah SMK Al-Islam Surakarta merupakan salah satu lembaga pendidikan swasta dari Yayasan Al-Islam Surakarta yang fokus mengajarkan ilmu pengetahuan Teknologi Informasi dan Ilmu pengetahuan Agama Islam. Dalam pembiayaan operasional sekolah sebagian besar dibebankan kepada siswa, terutama pada pembayaran SPP Sekolah. Karena berdasarkan Dirjen Dikdasmen Kementerian Pendidikan dan Kebudayaan, meluruskan

bahwa regulasi sekolah gratis alias terbebas dari pungutan SPP hanya diberlakukan pada jenjang Sekolah Dasar (SD) dan Sekolah Menengah Pertama (SMP) Negeri. Kebijakan itu berlaku secara nasional. Sementara kalau pembebasan SPP untuk SMA dan SMK itu kebijakan lokal (Hamid, 2017). Sehingga dalam hal ini pihak Sekolah SMK Al-Islam Surakarta dalam hal mencukupi biaya operasional Sekolah membebaskan pembiayaan sekolah kepada orang tua siswa melalui pembayaran SPP Sekolah yang harus dibayarkan setiap bulannya.

Masalah yang muncul terkait dalam pembayaran SPP sekolah yaitu apabila para siswa terlambat dalam membayar SPP sekolah. Hal tersebut menjadi masalah karena pembayaran SPP sekolah merupakan salah satu sumber dana tetap dalam meningkatkan kualitas pendidikan sekolah. Berdasarkan data dari bagian keuangan tata usaha, bahwa siswa yang terlambat membayar pada tahun ajaran 2017/2018 sekitar 60% dari total seluruh siswa yang berjumlah 236 siswa. Hal itu disebabkan oleh beberapa faktor seperti : pendapatan orang tua, pendidikan orang tua, tanggungan keluarga dan usia. Hal tersebut menjadi masalah yang besar karena banyaknya siswa yang terlambat dalam membayar SPP akan menurunkan pemasukan dana tetap sekolah. Sedangkan biaya pengeluaran untuk gaji para guru dan karyawan sebagian besar dialokasikan dari biaya SPP sekolah. Maka perlu adanya solusi berupa sistem prediksi yang dapat memperkirakan siswa yang tepat atau terlambat dalam melakukan pembayaran SPP sekolah.

Tujuan dari penelitian ini adalah membangun suatu sistem prediksi dengan metode *Naïve Bayes*. Karena metode tersebut dapat melakukan klasifikasi kelas tepat atau terlambat, dalam pembayaran biaya sekolah. *Algoritma Naïve Bayes* dipilih karena *Naïve Bayes* termasuk dalam Top 10 algorithms in data mining. Dimana paper tersebut telah dipublikasikan pada bulan desember 2006 oleh IEEE *International Conference on Data Mining* (Zierath, Rachholz, Woernle, & Müller, 2014).

Penelitian ini mengacu beberapa penelitian sebelumnya seperti : Penelitian tahun 2017 berjudul “Prediksi tingkat kelancaran pembayaran kredit bank menggunakan algoritma *Naïve Bayes* berbasis forward selection” (Hasan, 2017); Penelitian tahun 2017 yang berjudul “Prediksi tingkat kelulusan mahasiswa tepat waktu menggunakan *Naïve Bayes*” (Salmu & Solichin, 2017); Penelitian tentang komparasi *algoritma C4.5, naïve bayes, neural network* dan *logistic regression*. dengan metode pengujian *Confusion Matrix*, menghasilkan *Naïve Bayes* memiliki nilai akurasi paling tinggi (Saputra, 2014).

Berdasarkan uraian diatas diperlukan sebuah sistem prediksi keterlambatan pembayaran SPP sekolah, menggunakan pengolahan data mining berdasarkan paramateri : pendapatan, pendidikan, tanggungan keluarga dan usia. Menggunakan metode *Naïve Bayes* dengan metode *Information Gain* untuk melakukan prediksi keterlambatan pembayaran biaya sekolah. Sistem prediksi tersebut dibuat sebagai solusi dalam menangani permasalahan keterlambatan pembayaran SPP sekolah. maka pihak sekolah dapat melihat informasi mengenai perkiraan siswa-siswi yang akan membayar tepat waktu atau terlambat. Sehingga dengan adanya hasil informasi prediksi tersebut maka pihak sekolah akan dapat melakukan tindakan berupa pembinaan terhadap siswa-siswi yang diperkirakan akan terlambat dalam melakukan pembayaran SPP Sekolah.

II. TINJAUAN PUSTAKA

Pada penelitian yang berjudul “Prediksi tingkat kelancaran pembayaran kredit bank menggunakan algoritma *Naïve Bayes* berbasis *Forward Selection*” menjelaskan bahwa tujuan dari penelitian ini adalah mengoptimalkan algoritma *Naïve Bayes* dengan seleksi fitur *forward selection* untuk dapat meningkatkan hasil akurasi atau tingkat keberhasilan

yang didapatkan dari prediksi pembayaran kredit. Data yang akan digunakan dalam penelitian ini berasal dari Bank XY yang berada di Gorontalo. Data yang diperoleh berkaitan dengan semua aspek dari nasabah kredit termasuk informasi pribadi dari nasabah. Desain eksperimen dalam penelitian ini menggunakan dataset nasabah kredit. Sedangkan analisis yang digunakan adalah Model *algoritma Naïve Bayes* dengan seleksi fitur *forward selection*. Prediksi tingkat kelancaran pembayaran kredit menggunakan algoritma *Naïve Bayes* berbasis *forward selection* mampu memprediksi kelancaran pembayaran kredit kedepannya hal ini terbukti dengan perolehan nilai akurasi *Naïve Bayes* berbasis *forward selection* mampu mencapai nilai akurasi 71,97% (Hasan, 2017).

Penelitian tahun 2015 yang berjudul prediksi nilai proyek akhir mahasiswa menggunakan algoritma klasifikasi data mining. Penelitian ini bertujuan untuk membuat aturan yang dapat memprediksi nilai proyek akhir mahasiswa program diploma manajemen informasi berdasarkan nilai-nilai matakuliah yang mendukung penyusunan proyek akhir dengan menggunakan model klasifikasi data mining. Penelitian yang dilakukan juga akan menganalisis prestasi mahasiswa pada matakuliah yang mendukung penyusunan proyek akhir dengan pencapaian nilai proyek akhir mereka. Prediksi ini diharapkan dapat membantu dalam mengidentifikasi nilai berdasarkan matakuliah yang mendukung proyek akhir mereka. Berdasarkan penelitian yang telah dilakukan, analisis prediksi menggunakan ID3 memiliki akurasi sebesar 62,66%, CHAID 63,66% dan *Naïve Bayes* 65,67% (Mayadewi & Rosely, 2015).

Penelitian yang dilakukan dengan tujuan untuk mengetahui prediksi lama masa studi dari setiap mahasiswa yang telah menempuh kuliah hingga minimal semester IV dengan menggunakan dasar data kelulusan sebelumnya. Dari hasil uji coba diperoleh tingkat kesalahan prediksi berkisar 20% sampai dengan 50% dengan data training dan testing yang diambil secara random. Namun rata-rata tingkat kesalahan berkisar 20% hingga 34%. Tinggi rendahnya tingkat kesalahan dapat disebabkan oleh jumlah *record* data dan tingkat konsistensi dari data training yang digunakan. Sedangkan hasil prediksi dari ketepatan lama studi dari mahasiswa angkatan 2008 adalah sebesar 254 mahasiswa diprediksi "Tepat Waktu" dan sisanya yaitu 4 orang diprediksi "Tidak Tepat Waktu" (Jananto, 2013).

Algoritma yang dipakai dalam seleksi fitur dibahas secara singkat. *Seleksi fitur*, kita bisa deskripsikan dengan cara formal sebagai berikut : suatu masalah dengan banyak fitur $f_i \in n$ dengan $F=\{f_1, f_2, \dots, f_k\}$, bila fitur bernilai riil (R) bisa dinyatakan sebagai satu himpunan contoh subset $V=\{v_1, v_2, \dots, v_n\}$ dengan $n < k$ merupakan subset kelas C dengan *klasifier* $K: R^k \rightarrow C$ didefinisikan seperti pada rumus (1).

$$\forall v_i \in V, j \in \{1, \dots, k\}, v_i, j \in F_j \dots \dots \dots (1)$$

Information Gain adalah ukuran simetris, yaitu jumlah informasi yang diperoleh Y setelah mengamati X adalah sama dengan jumlah informasi yang diperoleh X setelah mengamati Y. *Simetrical* adalah properti yang diinginkan untuk mengukur *feature-feature* yang saling berkorelasi. Perhitungan pada koefisiensi *symmetrical* yaitu membagi nilai *gain* atribut dengan hasil *entropy* (y) + hasil *entropy* (x), lalu hasilnya dikali 2 maka akan diketahui nilai koefisien *symmetrical*. Rumus (2) untuk menghitung *koefisien symmetrical uncertainty Symmetrical uncertainty*:

$$2x \left[\frac{\text{gain}}{H(y)+H(x)} \right] \dots \dots \dots (2)$$

Information Gain (IG) merupakan suatu pengukuran yang dilakukan untuk melakukan seleksi terhadap atribut-atribut sehingga dapat disimpulkan atribut apa saja yang akan digunakan. *Information Gain* menggunakan *entropy* untuk menentukan atribut terbaik. *Entropy* merupakan ukuran ketidakpastian dimana semakin tinggi *entropy*, maka

semakin tinggi ketidakpastian (Jiawei, Kamber, & Pei, 2006). Rumus dari *entropy* adalah seperti pada rumus (3).

$$Entropy (S) = \sum_{i=k}^n - p_i * \log_2 p_i \dots \dots \dots (3)$$

Confusion matrix adalah suatu metode yang digunakan untuk melakukan perhitungan akurasi pada konsep data mining. Evaluasi dengan *confusion matrix* menghasilkan nilai akurasi, presisi dan recall. Akurasi dalam klasifikasi adalah persentase ketepatan record data yang diklasifikasikan secara benar setelah dilakukan pengujian pada hasil klasifikasi (Jiawei, Kamber, & Pei, 2006). 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 (Kusrini & Emha, 2009).

Pengukuran akurasi dilakukan dengan metode pengujian *Confusion matrix* yang dapat dilihat pada Tabel 1.

Tabel 1. Model *confusion matrix*

<i>Correct Classification</i>	<i>Classification</i>	
	Positif	Negatif
Positif	TP	FN
Negatif	FP	TN

$$\begin{aligned}
 Akurasi &= \frac{\sum_{i=1}^l \frac{TP_i + TN_i}{TP_i + TN_i + FP_i + FN_i}}{l} * 100\% \\
 Presisi &= \frac{\sum_{i=1}^l TP_i}{\sum_{i=1}^l (FP_i + TP_i)} * 100\% \\
 Recall &= \frac{\sum_{i=1}^l TP_i}{\sum_{i=1}^l (TP_i + FN_i)} * 100\% \dots \dots \dots (4)
 \end{aligned}$$

Keterangan :

TP adalah *True Positive*, yaitu jumlah data positif yang terklasifikasi dengan benar oleh sistem.

TN adalah *True Negative*, yaitu jumlah data negatif yang terklasifikasi dengan benar oleh sistem.

FN adalah *False Negative*, yaitu jumlah data negatif namun terklasifikasi salah oleh sistem.

FP adalah *False Positive*, yaitu jumlah data positif namun terklasifikasi salah oleh sistem.

III. METODE PENELITIAN

3.1 Metode Pengumpulan Data

Adapun metode pengumpulan data yang digunakan adalah :

a. Wawancara

Wawancara berguna untuk mendapatkan *data* atau informasi secara lengkap dari narasumber yang berkompeten. Dalam hal ini peneliti melakukan wawancara secara langsung kepada Bp. Muhammad Nur Kholis Dwi Putranto selaku ketua TU sekolah. Wawancara ini dilakukan untuk mendapatkan informasi data siswa dan laporan pembayaran SPP sekolah.

b. Observasi

Observasi berguna untuk melakukan pengamatan secara langsung bagaimana proses pendataan pembayaran SPP sekolah. Pengamatan yang dilakukan yaitu mencatat data siswa seperti nomor induk, nama siswa, kelas, alamat. dan data orang tua siswa sebagai atribut parameter prediksi seperti : Pendapatan, Pendidikan, Tanggungan Keluarga dan Usia.

c. Studi Pustaka

Studi Pustaka yaitu dengan cara mencari referensi atau teori yang diperlukan melalui buku atau jurnal ilmiah yang ada kaitannya dengan masalah-masalah penelitian. Pada tahap ini dilakukan studi *literature* yaitu mengumpulkan bahan-bahan referensi baik dari buku, artikel, jurnal ilmiah, proseding, maupun situs internet mengenai sistem prediksi keterlambatan pembayaran SPP sekolah dan beberapa referensi lain yang menunjang tujuan penelitian.

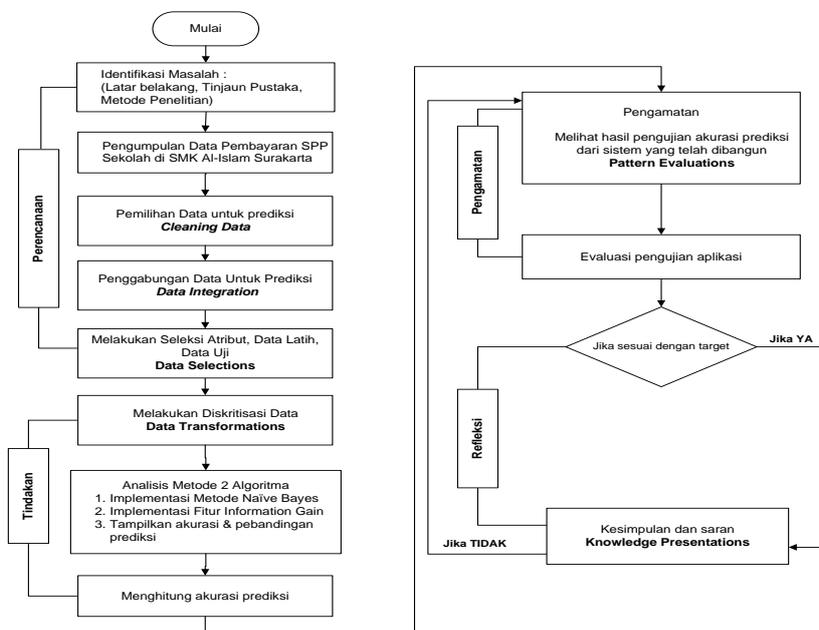
3.2 Metode Analisa Data

Data *training* dan data *testing* menggunakan data kelengkapan siswa (nomor induk, nama siswa, kelas, alamat, pendapatan orang tua, pendidikan, tanggungan keluarga dan usia) pada tahun ajaran 2017/2018. Data training digunakan untuk membuat pola pengetahuan pada prediksi keterlambatan pembayaran SPP Sekolah. Data testing digunakan untuk mengetes agar diperoleh data keterlambatan pembayaran SPP sekolah.

Metode penelitian yang digunakan pada penelitian ini adalah metode penelitian tindakan (*action research*) karena bertujuan untuk melakukan tindakan perubahan, perbaikan dan peningkatan kinerja organisasi khususnya dibidang prediksi pembayaran SPP sekolah. Menurut Lewin (1996), terdapat empat komponen pokok dalam penelitian tindakan yaitu *planning*, *action*, *observing* dan *reflecting*.

3.3 Alur Penelitian

Dalam penelitian ini menggunakan metode tindakan, sehingga langkah-langkah dalam alur penelitian ini dapat dibuat dengan model *action research* yang dapat dilihat pada Gambar 1.



Gambar 1. Alur penelitian

Keterangan Alur Penelitian :

Pada tahap perencanaan adalah melakukan identifikasi masalah dengan menentukan latar belakang masalah, mencari rumusan masalah, parameter dan solusi permasalahan. Data yang digunakan berasal dari data dapodik sekolah. Untuk mengumpulkan informasi yang berkaitan dengan studi kasus penelitian. Setelah data identitas siswa diperoleh kemudian dilakukan pembersihan data yaitu melakukan penghapusan data yang kurang penting dan hanya mengambil data siswa dan atribut dari data orang tua siswa yang berguna sebagai parameter prediksi keterlambatan pembayaran SPP sekolah.

Pada tahap pelaksanaan yaitu proses transformasi data pada atribut. Data atribut yang ditransformasi adalah data usia ayah dan usia ibu. Setelah melakukan transformasi data, dilakukan proses prediksi keterlambatan pembayaran SPP sekolah menggunakan metode *Naïve Bayes* dengan metode *information gain*.

Tahap pengamatan merupakan suatu tindakan untuk mengetahui kinerja sistem yang telah dibangun. Dalam tahap ini dilakukan pengamatan hasil prediksi, dimana hal-hal yang perlu diamati adalah nilai *precision*, nilai *recall* dan nilai *accuracy*.

Tahap yang terakhir adalah melakukan penulisan atau menyimpulkan hasil penelitian yang telah dilakukan. Beberapa hal yang akan disampaikan yaitu hasil prediksi dan nilai kinerja dari sistem prediksi yang telah dibangun.

IV. HASIL DAN PEMBAHASAN

Penelitian ini membahas tentang optimasi algoritma untuk sistem prediksi keterlambatan pembayaran SPP Sekolah. Tujuan dari penelitian ini adalah menerapkan metode *Naïve Bayes* dengan metode *Information Gain* untuk keterlambatan pembayaran sekolah. Metode *Information Gain* merupakan salah satu metode untuk menyeleksi fitur sehingga diharapkan dapat diperoleh akurasi yang lebih baik. Implementasi dalam penelitian ini dilakukan 2 skenario uji seperti perhitungan dengan 1 metode yaitu *Naïve Bayes* dan perhitungan 2 metode yaitu gabungan antara *Naïve Bayes* dengan *information gain*. Dengan adanya 2 skenario uji akan dapat diketahui perbandingan dari tingkurasi, sehingga dapat menentukan akurasi tertinggi.

Penggunaan metode *Naïve Bayes* menggunakan 30 data sampel yang telah diacak. Data ini diambil dari hasil olahan data dapodik sekolah. Data tersebut dibagi menjadi dua yaitu 20 data training dan 10 data testing. Data ini untuk dilakukan perhitungan manual, akan tetapi data dalam perhitungan manual ini tidak bisa di jadikan acuan dalam penelitian ini hal ini dikarenakan hasil yang akan didapatkan akan berhubungan dengan jumlah total data yang akan dihitung dan perhitungan ini berguna untuk membandingkan 2 buah metode sehingga akan terlihat perbedaan tingkat akurasi dari kedua metode tersebut.

Langkah yang dilakukan dalam pembahasan penelitian ini secara umum berdasarkan alur penelitian mulai dari pengambilan data dari dapodik sekolah, kemudian proses *cleaning data* (menghapus atribut yang kurang penting), setelah dilakukan transformasi data hingga menjadi sebuah data set seperti pada Tabel 2. Perhitungan metode *Naïve Bayes* adalah menghitung Probabilitas pada tabel data hitung manual, adapun pengertian probabilitas adalah suatu nilai untuk mengukur tingkat kemungkinan terjadinya suatu kejadian yang tidak pasti. (Rozzaqi, 2015). Pengolahan dalam metode ini menerapkan 2 buah metode yaitu metode *Naïve Bayes* dan metode *Information Gain*. Hal ini bertujuan untuk melihat perbandingan tingkat akurasi prediksi diantara kedua metode tersebut sehingga akan terlihat optimasi metode prediksi dalam implementasi.

Tabel 2. Dataset (data training & data testing)

No	Nama	Penghasilan Orang Tua	Tanggunggan Keluarga	Pendidikan Ayah	Umur Ayah	Pendidikan Ibu	Umur Ibu	Ket
1	Abdullah	2 - 4 Juta	Cukup	SD	Lansia Awal	SMP	Dewasa Awal	Tepat
2	Abu Musa	< 1 Juta	Banyak	SD	Lansia Awal	SMP	Lansia Awal	Terlambat
3	Achmad	1 - 2 Juta	Sedikit	S1	Lansia Awal	D3	Lansia Awal	Tepat
4	Adi Sadewo	< 1 Juta	Cukup	SMP	Lansia Akhir	SD	Lansia Akhir	Terlambat
5	Afifah Luthfitah	1 - 2 Juta	Sedikit	SMA	Dewasa Awal	SMP	Dewasa Akhir	Terlambat
6	Afilia Nur	< 1 Juta	Banyak	SMA	Lansia Akhir	SD	Manula	Tepat
7	Agal Febri	1 - 2 Juta	Sedikit	D3	Dewasa Awal	SMA	Dewasa Akhir	Tepat
8	Ahmad Jadid	2 - 4 Juta	Cukup	SD	Lansia Akhir	D3	Lansia Awal	Terlambat
9	Ahmad Latif	1 - 2 Juta	Sedikit	SMP	Manula	SD	Manula	Terlambat
10	Ahmad Migdad	< 1 Juta	Cukup	SD	Manula	SMA	Lansia Akhir	Tepat
11	Ahmad Rifqi	< 1 Juta	Cukup	SMA	Lansia Awal	SD	Dewasa Awal	Terlambat
12	Ahmad Sofyan	> 4 Juta	Sedikit	SMP	Lansia Awal	SMP	Lansia Awal	Tepat
13	Ahmad Thoriq	2 - 4 Juta	Banyak	S1	Lansia Awal	S1	Lansia Akhir	Terlambat
14	Ahna Hafiz	1 - 2 Juta	Banyak	S1	Dewasa Akhir	SMA	Lansia Awal	Terlambat
15	Ainun Mardiyah	1 - 2 Juta	Sedikit	SMP	Dewasa Awal	SD	Dewasa Awal	Tepat
16	Ajeng Juliawati	1 - 2 Juta	Sedikit	D3	Lansia Awal	SMA	Lansia Awal	Terlambat
17	Aji Putro Guritno	< 1 Juta	Banyak	SD	Manula	SD	Dewasa Akhir	Terlambat
18	Alfaisah Ikfani Safaria	2 - 4 Juta	Banyak	SMA	Dewasa Akhir	SMP	Dewasa Akhir	Tepat
19	Alfan Rizky Affandi	> 4 Juta	Cukup	S1	Dewasa Akhir	D3	Lansia Akhir	Terlambat
20	Alfian	> 4 Juta	Sedikit	SMP	Dewasa Akhir	S1	Dewasa Akhir	Tepat
21	Alfina Putri Ariyani	> 4 Juta	Banyak	SMP	Lansia Awal	SMP	Lansia Awal	Terlambat
22	Ali Amientoni	< 1 Juta	Cukup	SMA	Dewasa Akhir	SD	Lansia Awal	Terlambat
23	Alif Nanda Altusan	> 4 Juta	Cukup	D3	Dewasa Awal	SMA	Dewasa Awal	Tepat
24	Alim Endar Marfa'i	1 - 2 Juta	Banyak	SMP	Dewasa Akhir	SD	Lansia Akhir	Terlambat
25	Althaf Rizky	2 - 4 Juta	Cukup	SMA	Dewasa Akhir	SMA	Lansia Awal	Tepat
26	Alwi	1 - 2 Juta	Banyak	SMP	Lansia Akhir	SD	Lansia Awal	Terlambat
27	Amalia Nur Sa'adah	< 1 Juta	Sedikit	SMA	Dewasa Akhir	SMP	Dewasa Awal	Tepat
28	Amanda Aulia	< 1 Juta	Banyak	SD	Manula	SD	Lansia Awal	Terlambat
29	Amanda Putri Tarisa	1 - 2 Juta	Cukup	SMA	Lansia Akhir	SD	Lansia Awal	Terlambat
30	Ameilia Duwi	> 4 Juta	Banyak	S1	Dewasa Awal	D3	Dewasa Akhir	Tepat

Berdasarkan Dataset pada Tabel 1 terdapat 30 data. Pada dataset tersebut dibagi dua yaitu 20 untuk data training dan 10 untuk data testing.

1. Perhitungan Metode *Naïve Bayes*

Dalam perhitungan dengan metode *Naïve Bayes* dapat menggunakan persamaan *bayes* seperti pada rumus (5)

$$P(H|X) = \frac{P(X|H) \cdot P(H)}{P(X)} \dots\dots\dots(5)$$

Langkah Perhitungan *Naïve Bayes*

a. Menentukan Nilai Untuk Setiap Kelas

C1 (Kelas = “Tepat”) = jumlah “Tepat” pada kolom Keterangan = 9

C2 (Kelas = “Terlambat”) = jumlah “Terlambat” pada kolom Keterangan = 11

b. Menentukan Nilai Setiap Parameter

1. Penghasilan

Atribut Kelas “Tepat”		Atribut Kelas “Terlambat”	
➤ < 1 Juta	= 2/9 = 0,22	➤ < 1 Juta	= 4/11 = 0,36
➤ 1 – 2 Juta	= 3/9 = 0,33	➤ 1 – 2 Juta	= 4/11 = 0,36
➤ 2- 4 Juta	= 2/9 = 0,22	➤ 2 – 4 Juta	= 2/11 = 0,18
➤ > 4 Juta	= 2/9 = 0,22	➤ 4 Juta	= 1/11 = 0,09

2. Tanggungan Keluarga

Atribut Kelas “Tepat”		Atribut Kelas “Terlambat”	
➤ Sedikit	= 5/9 = 0,56	➤ Sedikit	= 3/11 = 0,27
➤ Cukup	= 2/9 = 0,22	➤ Cukup	= 4/11 = 0,36
➤ Banyak	= 2/9 = 0,22	➤ Banyak	= 4/11 = 0,26

3. Pendidikan Ayah

Atribut Kelas “Tepat”		Atribut Kelas “Terlambat”	
➤ SD	= 2/9 = 0,22	➤ SD	= 3/11 = 0,27
➤ SMP	= 3/9 = 0,33	➤ SMP	= 4/11 = 0,18
➤ SMA	= 2/9 = 0,22	➤ SMA	= 4/11 = 0,18
➤ D3	= 1/9 = 0,11	➤ D3	= 1/11 = 0,09
➤ S1	= 1/9 = 0,11	➤ S1	= 3/11 = 0,27

4. Umur Ayah

Atribut Kelas “Tepat”		Atribut Kelas “Terlambat”	
➤ Dewasa Awal	= 2/9 = 0,22	➤ Dewasa Awal	= 1/11 = 0,09
➤ Dewasa Akhir	= 2/9 = 0,22	➤ Dewasa Akhir	= 2/11 = 0,18
➤ Lansia Awal	= 3/9 = 0,33	➤ Lansia Awal	= 4/11 = 0,36
➤ Lansia Akhir	= 1/9 = 0,11	➤ Lansia Akhir	= 2/11 = 0,18
➤ Manula	= 1/9 = 0,11	➤ Manula	= 2/11 = 0,18

5. Pendidikan Ibu

Atribut Kelas “Tepat”		Atribut Kelas “Terlambat”	
➤ SD	= 2/9 = 0,22	➤ SD	= 3/11 = 0,27
➤ SMP	= 3/9 = 0,33	➤ SMP	= 3/11 = 0,27
➤ SMA	= 2/9 = 0,22	➤ SMA	= 2/11 = 0,18
➤ D3	= 1/9 = 0,11	➤ D3	= 2/11 = 0,18
➤ S1	= 1/9 = 0,11	➤ S1	= 1/11 = 0,09

6. Umur Ibu

Atribut Kelas “Tepat”

- Dewasa Awal = $2/9 = 0,22$
- Dewasa Akhir = $3/9 = 0,33$
- Lansia Awal = $2/9 = 0,22$
- Lansia Akhir = $1/9 = 0,11$
- Manula = $1/9 = 0,11$

Atribut Kelas “Terlambat”

- Dewasa Awal = $1/11 = 0,09$
- Dewasa Akhir = $2/11 = 0,18$
- Lansia Awal = $5/11 = 0,45$
- Lansia Akhir = $2/11 = 0,18$
- Manula = $1/11 = 0,09$

c. Perhitungan Data Testing

Dalam perhitungan data testing diambil 10 data testing dari 30 dataset yang dapat dilihat pada Tabel 3.

Tabel 3. Data testing

No	Nama	Penghasilan Orang Tua	Tanggung jawab Keluarga	Pendidikan Ayah	Umur Ayah	Pendidikan Ibu	Umur Ibu	Ket
21	Alfina Putri Ariyani	> 4 Juta	Banyak	SMP	Lansia Awal	SMP	Lansia Awal	Terlambat
22	Ali Amientoni	< 1 Juta	Cukup	SMA	Dewasa Akhir	SD	Lansia Awal	Terlambat
23	Alif Nanda Altusan	> 4 Juta	Cukup	D3	Dewasa Awal	SMA	Dewasa Awal	Tepat
24	Alim Endar Marfa'i	1 - 2 Juta	Banyak	SMP	Dewasa Akhir	SD	Lansia Akhir	Terlambat
25	Althaf Rizky	2 - 4 Juta	Cukup	SMA	Dewasa Akhir	SMA	Lansia Awal	Tepat
26	Alwi	1 - 2 Juta	Banyak	SMP	Lansia Akhir	SD	Lansia Awal	Terlambat
27	Amalia Nur Sa'adah	< 1 Juta	Sedikit	SMA	Dewasa Akhir	SMP	Dewasa Awal	Tepat
28	Amanda Aulia	< 1 Juta	Banyak	SD	Manula	SD	Lansia Awal	Terlambat
29	Amanda Putri Tarisa	1 - 2 Juta	Cukup	SMA	Lansia Akhir	SD	Lansia Awal	Terlambat
30	Ameilia Duwi	> 4 Juta	Banyak	S1	Dewasa Awal	D3	Dewasa Akhir	Tepat

Data testing diatas merupakan data asli dari dataset, maka untuk menguji akurasi algoritma *Naïve Bayes* dari data diatas dilakukan perhitungan dengan metode *naïve bayes*, kemudian hasilnya dibandingkan dengan data asli sebenarnya. Sehingga akan dapat dihitung tingkat akurasi dari algoritma *Naïve Bayes*.

Dalam pengujian sistem hanya mengambil data sampling sebanyak 30 data. Untuk data training dari No 1 s/d 20, sedangkan untuk data testing dari No. 21 s/d 30. Proses perhitungan metode *Naïve Bayes* ini, menggunakan tools Ms. Excel yaitu menguji 10 data testing. Sebagai gambaran proses perhitungan maka ditampilkan sebuah perhitungan data testing dari data.

No. 21, sedangkan untuk data testing pada No. 22 s/d 30 dilakukan dengan tools Ms. Excel sehingga hasilnya langsung ditampilkan. Proses perhitungan data testing No.

21 dapat dilihat Tabel 4 sedangkan untuk hasil keseluruhan dari perhitungan 10 data testing maka dapat dilihat pada Tabel 5.

Tabel 4. Data testing ke 1

No	Nama	Penghasilan Orang Tua	Tanggung jawab Keluarga	Pendidikan Ayah	Umur Ayah	Pendidikan Ibu	Umur Ibu	Ket
21	Alfina Putri Ariyani	> 4 Juta	Banyak	SMP	Lansia Awal	SMP	Lansia Awal	?

Likehood of "TEPAT" = $(0,22 \times 0,22 \times 0,33 \times 0,33 \times 0,33 \times 0,22) = 0,000406442$

Likehood of "TERLAMBAT" = $(0,09 \times 0,36 \times 0,18 \times 0,36 \times 0,27 \times 0,45) = 0,000270947$

Hob "TEPAT" = $0,000406442 / (0,000406442 + 0,000270947) = 0,6000$

Hob "TERLAMBAT" = $0,000270947 / (0,000406442 + 0,000270947) = 0,4000$

Hob Tepat > Hob Terlambat, sehingga diperoleh prediksi = "TEPAT"

Untuk Data ke 22 s/d 30 dilakukan di excel dengan menggunakan rumus yang sama sehingga diperoleh hasil keseluruhan yang dapat dilihat pada Tabel 5.

Tabel 5. Hasil Perhitungan Data Testing

No	Nama	Penghasilan Orang Tua	Tanggung jawab Keluarga	Pendidikan Ayah	Umur Ayah	Pendidikan Ibu	Umur Ibu	Ket Hasil Asli	Ket Hasil Prediksi
21	Alfina Putri Ariyani	> 4 Juta	Banyak	SMP	Lansia Awal	SMP	Lansia Awal	Terlambat	Tepat
22	Ali Amientoni	< 1 Juta	Cukup	SMA	Dewasa Akhir	SD	Lansia Awal	Terlambat	Terlambat
23	Alif Nanda Altusan	> 4 Juta	Cukup	D3	Dewasa Awal	SMA	Dewasa Awal	Tepat	Tepat
24	Alim Endar Marfa'i	1 - 2 Juta	Banyak	SMP	Dewasa Akhir	SD	Lansia Akhir	Terlambat	Terlambat
25	Althaf Rizky	2 - 4 Juta	Cukup	SMA	Dewasa Akhir	SMA	Lansia Awal	Tepat	Terlambat
26	Alwi	1 - 2 Juta	Banyak	SMP	Lansia Akhir	SD	Lansia Awal	Terlambat	Terlambat
27	Amalia Nur Sa'adah	< 1 Juta	Sedikit	SMA	Dewasa Akhir	SMP	Dewasa Awal	Tepat	Tepat
28	Amanda Aulia	< 1 Juta	Banyak	SD	Manula	SD	Lansia Awal	Terlambat	Terlambat
29	Amanda Putri Tarisa	1 - 2 Juta	Cukup	SMA	Lansia Akhir	SD	Lansia Awal	Terlambat	Terlambat
30	Ameilia Duwi	> 4 Juta	Banyak	S1	Dewasa Awal	D3	Dewasa Akhir	Tepat	Tepat

d. Pengujian Confusion Matrik

Tabel matrik *confusion* yang digunakan dalam pengujian seperti pada Tabel 6.

Tabel 6. Matrik *confusion*

<i>Correct Classification</i>	<i>Classification</i>	
	Positif	Negatif
Positif	3	1
Negatif	1	5

Keterangan

1. *Classification* Positif – Positif = 3 karena jumlah data positif yang terklasifikasi dengan benar oleh sistem.
2. *Classification* Negatif – Positif = 1 karena jumlah data negatif namun terklasifikasi salah oleh sistem.
3. *Classification* Positif – Negatif = 1 karena jumlah data positif namun terklasifikasi salah oleh sistem
4. *Classification* Negatif – Negatif = 5 karena jumlah data negatif yang terklasifikasi dengan benar oleh system

Berdasarkan hasil perhitungan metode *Naïve Bayes* maka untuk pengujian algoritma dapat dilakukan dengan *confusion* matrik sesuai rumus ke (4).

Hasil Pengujian

$$\text{Akurasi} = \frac{3+5}{(3+1+1+5)} * 100\% = 80\%$$

$$\text{Presisi} = \frac{3}{(3+1)} * 100\% = 75\%$$

$$\text{Recall} = \frac{3}{(3+1)} * 100\% = 75\%$$

Hasil akurasi dengan metode *Naïve Bayes* diperoleh = 80% Untuk meningkatkan akurasi metode *Naïve Bayes*, maka dilakukan eksperimen yaitu dengan menggabungkan metode *Information Gain* dengan metode *Naïve Bayes*.

2. Perhitungan Metode *Information Gain* (IG) + *Naïve Bayes*

Informasi *gain* adalah suatu algoritma fitur seleksi dimana algoritma ini nantinya yang akan menentukan jumlah atribut yang akan digunakan. Adapun perhitungan formula dari algoritma *Information Gain* seperti pada rumus (6).

$$\text{entropy}(s) = \sum_{i=1}^n -p_i \cdot \log_2 p_i \dots\dots\dots(6)$$

a. Menghitung Entropy Total

Diketahui

Jumlah data kelas Tepat = 9

Jumlah data kelas Terlambat = 11

Jumlah data keseluruhan = 20

$$\text{Entropy Total} = \left(-\frac{9}{20} * \log_2 \left(\frac{9}{20}\right)\right) + \left(-\frac{11}{20} * \log_2 \left(\frac{11}{20}\right)\right) = 0,993$$

b. Menghitung *Entropy* Atribut

➤ Atribut Penghasilan

< 1 Juta

Diketahui

Jumlah data kelas Tepat = 2

Jumlah data kelas Terlambat = 4

Jumlah data keseluruhan = 6

$$\text{Entropy} = \left(-\frac{2}{6} * \log_2 \left(\frac{2}{6}\right)\right) + \left(-\frac{4}{6} * \log_2 \left(\frac{4}{6}\right)\right) = 0,918$$

2 - 4 Juta

Diketahui

Jumlah data kelas Tepat = 2

Jumlah data kelas Terlambat = 2

Jumlah data keseluruhan = 4

1 - 2 Juta

Diketahui

Jumlah data kelas Tepat = 3

Jumlah data kelas Terlambat = 4

Jumlah data keseluruhan = 7

$$\text{Entropy} = \left(-\frac{3}{7} * \log_2 \left(\frac{3}{7}\right)\right) + \left(-\frac{4}{7} * \log_2 \left(\frac{4}{7}\right)\right) = 0,985$$

> 4 Juta

Diketahui

Jumlah data kelas Tepat = 2

Jumlah data kelas Terlambat = 1

Jumlah data keseluruhan = 3

$$\text{Entropy} = \left(-\frac{2}{4} \cdot \log_2\left(\frac{2}{4}\right)\right) + \left(-\frac{2}{4} \cdot \log_2\left(\frac{2}{4}\right)\right) = 1$$

$$\text{Entropy} = \left(-\frac{2}{3} \cdot \log_2\left(\frac{2}{3}\right)\right) + \left(-\frac{1}{3} \cdot \log_2\left(\frac{1}{3}\right)\right) = 0,918$$

$$\text{Gain Penghasilan} = 0,993 - \left(\left(\frac{6}{20} \cdot 0,918\right) + \left(\frac{7}{20} \cdot 0,985\right) + \left(\frac{4}{20} \cdot 1\right) + \left(\frac{3}{20} \cdot 0,918\right)\right) = \mathbf{0,03471}$$

➤ **Atribut Lain**

Untuk meringkas penulisan, pada proses perhitungan atribut lain telah dihitung dengan excel, maka menghasilkan nilai *Gain* masing-masing berikut.

$$\text{Gain Tanggungan} = 0,993 - \left(\left(\frac{8}{20} \cdot 0,954\right) + \left(\frac{6}{20} \cdot 0,918\right) + \left(\frac{6}{20} \cdot 0,18\right)\right) = \mathbf{0,06002}$$

$$\text{Gain Pend. Ayah} = 0,993 - \left(\left(\frac{5}{20} \cdot 0,971\right) + \left(\frac{5}{20} \cdot 0,971\right) + \left(\frac{4}{20} \cdot 1\right) + \left(\frac{2}{20} \cdot 1\right) + \left(\frac{4}{20} \cdot 0,811\right)\right) = \mathbf{0,0453}$$

$$\text{Gain Umur Ayah} = 0,993 - \left(\left(\frac{3}{20} \cdot 0,918\right) + \left(\frac{4}{20} \cdot 1\right) + \left(\frac{7}{20} \cdot 0,985\right) + \left(\frac{3}{20} \cdot 0,918\right) + \left(\frac{3}{20} \cdot 0,918\right)\right) = \mathbf{0,0349}$$

$$\text{Gain Pend. Ibu} = 0,993 - \left(\left(\frac{5}{20} \cdot 0,971\right) + \left(\frac{6}{20} \cdot 1\right) + \left(\frac{4}{20} \cdot 1\right) + \left(\frac{3}{20} \cdot 0,918\right) + \left(\frac{2}{20} \cdot 1\right)\right) = \mathbf{0,0125}$$

$$\text{Gain Umur Ibu} = 0,993 - \left(\left(\frac{3}{20} \cdot 0,918\right) + \left(\frac{5}{20} \cdot 0,971\right) + \left(\frac{7}{20} \cdot 0,863\right) + \left(\frac{3}{20} \cdot 0,918\right) + \left(\frac{2}{20} \cdot 1\right)\right) = \mathbf{0,0727}$$

c. **Perangkingan Nilai *Gain* Atribut**

Berdasarkan proses perhitungan *gain* atribut diatas maka dapat dibuat sebuah perangkingan sesuai Tabel 7.

Tabel 7. Perangkingan nilai *gain* atribut

No	Atribut Parameter	Nilai <i>Gain</i>	Rangking
1	Data Atribut Umur Ibu	0,0727	1
2	Data Atribut Tanggungan Keluarga	0,0600	2
3	Data Atribut Pendidikan Ayah	0,0453	3
4	Data Atribut Umur Ayah	0,0349	4
5	Data Atribut Penghasilan	0,0347	5
6	Data Atribut Pendidikan Ibu	0,0125	6

Berdasarkan perangkingan seperti pada Tabel 7, maka diambil 4 Parameter terbaik (nilai *gain* dari tertinggi ke rendah) untuk digunakan dalam implementasi metode *Information Gain* dengan *Naïve Bayes*.

d. **Implementasi Metode *Information Gain***

Dalam implementasi *information gain*, hanya menggunakan 4 parameter yang paling berpengaruh, maka berikut perhitungan dalam Metode *Naïve Bayes*. Untuk mengetahui tingkat akurasi, maka implementasi menggunakan data testing yang sama yaitu data *record* No.21 dapat dilihat pada Tabel 8.

Tabel 8. Data testing 4 parameter

No	Nama	Tanggung Keluarga	Pendidikan Ayah	Umur Ayah	Umur Ibu	Ket
21	Alfina Putri Ariyani	Banyak	SMP	Lansia Awal	Lansia Awal	?

Dari Tabel 8 terlihat bahwa Parameter yang digunakan hanyalah 4 parameter yaitu Tanggungan Keluarga, Pendidikan Ayah, Umur Ayah dan Umur Ibu. Sehingga diperoleh perhitungan sebagai berikut.

$$\text{Likelihood of "TEPAT"} = (0,22 \times 0,33 \times 0,33 \times 0,22) = 0,005486968$$

$$\text{Likelihood of "TERLAMBAT"} = (0,36 \times 0,18 \times 0,36 \times 0,45) = 0,010928215$$

$$\text{Hob "TEPAT"} = 0,005486968 / (0,005486968 + 0,010928215) = 0,3343$$

$$\text{Hob "TERLAMBAT"} = 0,010928215 / (0,005486968 + 0,010928215) = 0,6657$$

Hob Tepat < Hob Terlambat, sehingga diperoleh prediksi = **"TERLAMBAT"**

Hasil prediksi berupa “**TERLAMBAT**” yang nilainya sama dengan data asli sehingga menunjukan IG+Naïve Bayes memberikan hasil yang lebih baik.

3. Perbandingan Tingkat Akurasi

Berdasarkan perhitungan dari kedua metode tersebut maka dapat dikomparasikan dari kedua metode tersebut dengan data asli maka diperoleh sesuai Tabel 9.

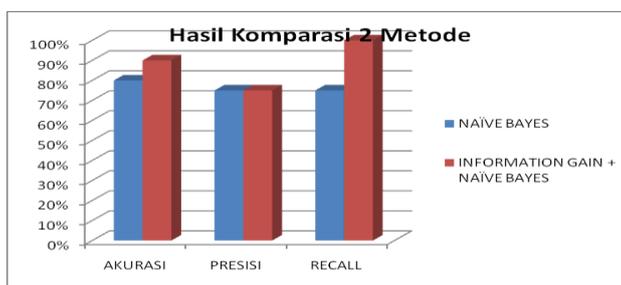
Tabel 9. Hasil Prediksi Komparasi Metode

No	Nama	Hasil Ket Asli	Hasil Naïve Bayes	Hasil Information Gain + Naïve Bayes
21	Alfina Putri Ariyani	Terlambat	Tepat	Terlambat
22	Ali Amientoni	Terlambat	Terlambat	Terlambat
23	Alif Nanda Altusan	Tepat	Tepat	Tepat
24	Alim Endar Marfa'i	Terlambat	Terlambat	Terlambat
25	Althaf Rizky	Tepat	Terlambat	Terlambat
26	Alwi	Terlambat	Terlambat	Terlambat
27	Amalia Nur Sa'adah	Tepat	Tepat	Tepat
28	Amanda Aulia	Terlambat	Terlambat	Terlambat
29	Amanda Putri Tarisa	Terlambat	Terlambat	Terlambat
30	Ameilia Duwi	Tepat	Tepat	Tepat

Dari hasil perbandingan Tabel 9 maka dapat diperoleh Nilai Komparasi Pengujian sistem pada Tabel 10 dan Gambar 2.

Tabel 10. Hasil Komparasi

No	Pengujian	Naïve Bayes	Information Gain + Naïve Bayes
1	Akurasi	80%	90%
2	Presisi	75%	75%
3	Recall	75%	100%



Gambar 2. Hasil komparasi 2 metode

Hasil komparasi menunjukan bahwa gabungan metode *Information Gain + Naïve Bayes* memberikan hasil akurasi tertinggi yaitu 90%.

V. KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil pembahasan dalam penelitian ini maka dapat diambil kesimpulan bahwa penerapan gabungan antara metode algoritma *Information Gain* dengan *Naïve Bayes* untuk prediksi keterlambatan pembayaran SPP Sekolah diperoleh hasil akurasi lebih

baik yaitu 90% sesuai dengan Tabel 10 dan Gambar 2. Sedangkan jika tanpa penerapan algoritma *information gain* hanya diperoleh akurasi 80%. Maka dengan ini peneliti menyimpulkan bahwa optimasi metode *Naïve Bayes* dengan Metode *information gain* dapat membantu dalam melakukan prediksi keterlambatan pembayaran SPP sekolah dengan hasil akurasi yang lebih baik.

5.2 Saran

Berdasarkan hasil penelitian yang telah dilakukan maka muncul gagasan-gagasan yang dirangkum dalam usulan dan saran untuk penelitian yang berhubungan dengan prediksi keterlambatan pembayaran SPP (sumbangan pembinaan pendidikan) sekolah antara lain :

1. Dalam Penelitian prediksi hendaknya pemilihan data dilihat nilai homogenya terlebih dahulu, karena dalam penelitian ini pengambilan data traning terlalu kompleks, hal ini nantinya akan sangat mempengaruhi akurasi.
2. Dalam melakukan penelitian yang berkaitan dengan prediksi haruslah memilah algoritma yang sesuai dengan jenis data (algoritma yang menyesuaikan data).

DAFTAR PUSTAKA

- A. Jananto. (2013). "Algoritma *Naïve Bayes* untuk Mencari Perkiraan Waktu Studi Mahasiswa," *Tekno. Inf.*, vol. 18, no. 1, pp. 9–16.
- Rozzaqi, A. R. (2015). "*Naïve Bayes* dan Filtering *Feature Selection Information Gain* untuk Prediksi Ketepatan Kelulusan Mahasiswa," *Jurnal Informatika UPGRIS*, vol. 1, 30–41.
- Sanjiwani, E. & I. Ayu, A. (2012). Biaya, and P. Pembelajaran, "Terhadap Kualitas Proses Pembelajaran dan Aspirasi Pendidikan Siswa (Studi Tentang Persepsi Para Siswa SMA Dwijendra Denpasar Program Pascasarjana)"
- Hamid. (2017). Kompas.com, "Sejumlah Sekolah Diperkenankan Pungut SPP", <https://nasional.kompas.com/read/2017/01/19/23323911/sejumlah.sekolah.diperkenankan.pungut.spp>. Sheila Respati.
- Jiawei, H., Kamber, M., & Pei, J. (2006), *Data Mining : Concept and Techniques Second Edition*, Morgan Kaufmann Publishers.
- J. Zierath, R. Rachholz, C. Woernle, & A. Müller. (2014). *Load Calculation on Wind Turbines: Validation of Flex5, Alaska/Wind, MSC.Adams and SIMPACK by Means of Field Tests*.
- Kusrini & Emha, T.L. (2009). *Algoritma Data Mining*, Penerbit Andi, Yogyakarta.
- Lewin, K. (1996). *Action Research and Minority Problems*, Publis in *Journal of Social Issues*, 2: 34- 46.
- M. Hasan. (2017). "Menggunakan Algoritma *Naïve Bayes* Berbasis," vol. 9, pp. 317–324.
- Mayadewi, P. & Rosely, E. (2015). "Prediksi Nilai Proyek Akhir Mahasiswa Menggunakan Algoritma Klasifikasi Data Mining," , 329-334
- Saputra, R.A. (2014). "Komparasi Algoritma Klasifikasi Data Mining Untuk Memprediksi Penyakit Tuberculosis (TB): Studi Kasus Puskesmas Karawang Sukabumi," *Semin. Nas. Inov. dan Tren*, 1-8.
- S. Salmu & A. Solichin. (2017). "Prediksi Tingkat Kelulusan Mahasiswa Tepat Waktu Menggunakan *Naïve Bayes*: Studi Kasus UIN Syarif Hidayatullah Jakarta Prediction of Timeliness Graduation of Students Using *Naïve Bayes* : A Case Study at Islamic State University Syarif Hidayatullah Jakarta," no. April, 701–709.