



Penerapan Metode *Naïve bayes* dalam Memprediksi Prestasi Siswa

Riani Maila Apsari

Teknik Informatika, Teknologi, Institut Teknologi Pagar Alam,
rianimaila641@gmail.com

Abstract

This research aims to apply the Naïve bayes algorithm method in predicting students' achievement at SMA Negeri 3 Pagar Alam. Based on the results of observations, determining student' achievement at SMA Negeri 3 Pagar Alam was still done by looking at the assessment points one by one, students' scores were used to determine who had the highest score as the most successful student, therefore there were many students who have achievements. There were various achievements of the students in SMA Negeri 3 Kota Pagar Alam. There were several students who had highest, average, and lowest achievements. Thus, there was needed the anticipation in order to improve the quality of education dealing with students who do not achieve well to produce a good and potential generation. The system development method used in this research is the CRIPS-DM method with some stages such as business understanding, data understanding, data preparation, modeling, Evaluation and dissemination. The method used was the Naïve bayes method with Testing using Cross validation. The results of this test were obtained Class recall for students with a Performance score for high achieving students is 100%, class prediction is 55.56%, for average achieving students is 95.40% with class prediction is 97.65% and those with less achievement are 89.47%, class precision is 100.00% for Performance results, the accuracy level is 94.55%. Those were the results in applying of Naive Bayes algorithm method in predicting outstanding students at SMA Negeri 3 Pagar Alam.

Keywords: Prediction, Naive Bayes, Cross-Validation

Abstrak

Penelitian ini bertujuan untuk menerapkan metode algoritma *naïve bayes* dalam memprediksi siswa berprestasi pada SMA Negeri 3 Pagar Alam. berdasarkan hasil observasi menentukan siswa berprestasi pada SMA Negeri 3 Pagar Alam masih dilakukan dengan cara melihat poin penilaian satu persatu nilai siswa untuk menentukan siswa mana yang memiliki nilai dengan rata-rata tertinggi untuk mendapatkan predikat siswa berprestasi. Oleh sebab itu, banyak siswa-siswi yang memiliki prestasi yang berbeda-beda di SMA Negeri 3 Kota Pagar Alam terdapat siswa-siswi yang memiliki prestasi tinggi atau berprestasi dan kurang berprestasi sehingga pihak sekolah harus melakukan antisipasi dalam memperbaiki mutu pendidikan dalam menangani siswa-siswi yang kurang berprestasi untuk menghasilkan generasi yang baik dan berpotensi. Metode pengembangan sistem yang digunakan dalam penelitian ini adalah metode *CRIPS-DM* dengan tahapan yaitu: Pemahaman bisnis, pemahaman data, persiapan data, pemodelan, evaluasi dan penyebaran. Metode yang digunakan yaitu Metode *Naïve bayes* dengan pengujian menggunakan *Cross validation*. Hasil dari pengujian tersebut didapatkan *Class recall* siswa dengan nilai *Performance* siswa kelas berprestasi 100 % class predcision 55.56% untun siswa cukup berprestasi 95.40% dengan class predcision 97.65% dan yang kurang berprestasi class *recall* 89.47% class precision 100.00% pada hasil *Performance* hasil tingkat akurasi 94.55% pada class Predicision. hasil dengan Penerapan Metode algoritma *Naive Bayes* dalam memprediksi siswa berprestasi pada SMA Negeri 3 Pagar Alam.

Kata kunci : Prediksi, *Naive Bayes*, *Cross-Validation*.

1. Pendahuluan

Prestasi merupakan kemampuan siswa dalam mencapai hasil belajar yang selama ini mereka pelajari menyangkut pemahanan dengan mengingat hasil yang mereka pelajari dan apa yang disampaikan guru mampu mereka cerna dan mereka terapkan pada pengembangan diri mereka masing-masing, bahkan dalam kondisi ujian mereka mampu mengingat dan menyelesaikan masalah-masalah soal yang terdapat pada lembar soal ujian selain itu untuk mengetahui siswa itu berprestasi dapat dilihat dari keaktifan dan ambisinya dalam belajar[1].

Maka dari itu prestasi belajar dapat dikatakan sebagai hasil kemampuan dan daya ingat siswa dari hasil belajar yang diperoleh selama belajar di sekolah yang bersifat kognitif dan biasanya hasil akhir yang dicapai melalui pengukuran dan penilaian. Prestasi belajar dapat dilihat melalui nilai hasil ulangan atau nilai raport sesuai dengan jumlah pelajaran yang diberikan oleh guru dengan setiap mata pelajaran. Untuk hasil belajar yang telah diperoleh siswa-siswi berupa nilai raport untuk setiap mata pelajaran nilai raportnya berbeda-beda yang di terima siswa maka dari itu untuk mengetahui banyaknya siswa berprestasi dan kurang berprestasi yang mempengaruhi jumlah kenaikan kelas nilai raport tersebut dapat digunakan untuk prediksi dengan *Algoritma Nive Bayes* [2].

Algoritma *Naive Bayes* merupakan salah satu metode yang dapat digunakan untuk mengklasifikasikan suatu kumpulan data. Algoritma ini menggunakan metode probabilitas dan statistik yang diusulkan oleh ilmuwan Inggris Thomas Bayes untuk memprediksi probabilitas masa depan berdasarkan pengalaman masa lalu [3]. Pembelajaran mesin yang menggunakan perhitungan probabilitas menggunakan konsep pendekatan Bayesian. Kata merendahkan naif berasal dari asumsi bahwa pengaruh skor atribut tidak bergantung pada kemungkinan kelas tertentu dari skor atribut lainnya. Menggunakan *teorema Bayes* dalam algoritma *naive bayes* terdiri dari menggabungkan probabilitas sebelumnya dan probabilitas bersyarat dalam sebuah formula yang bisa dipakai buat menghitung probabilitas berdasarkan setiap kemungkinan klasifikasi.

Dari hasil wawancara, observasi dan dokumentasi yang dilakukan di SMA Negeri 3 Pagaralam didapatkan. Pada saat jam pelajaran berlangsung banyak murid yang tidak fokus pada pelajaran yang sedang diajarkan tidak paham apa yang mereka pikirkan oleh karena itu mereka sulit memahami materi yang disampaikan oleh guru yang sedang mengajar sehingga pada saat pemberian tugas-tugas atau ulangan harian yang di berikan oleh guru siswa-

siswi mengalami kesulitan menjawab soal-soal yang diberikan sehingga berdampak pada nilai kemudian banyak juga siswa-siswi yang bermalas-malasan atau tidak mengerjakan tugas yang diberikan oleh guru sehingga nilai akademik yang keluar tidak optimal atau menurun. oleh sebab itu banyak siswa-siswi yang memiliki prestasi yang berbeda-beda di SMA Negeri 3 Kota Pagar Alam terdapat siswa-siswi yang memiliki prestasi tinggi atau berprestasi dan kurang berprestasi Sehingga pihak sekolah harus melakukan antisipasi dalam memperbaiki mutu pendidikan dalam menangani siswa-siswi yang memiliki prestasi yang kurang sehingga dapat menghasilkan generasi yang baik dan memiliki potensi siswa yang baik.

Berdasarkan uraian diatas dapat disimpulkan bahwa algoritma *naive bayes* mampu dalam hal memprediksi kejadian yang akan terjadi dimasa depan. Maka dari itu untuk memprediksi siswa berprestasi dapat menggunakan algoritma *Naive bayes* untuk membantu guru SMA Negeri 3 Pagar Alam dalam memprediksikan siswa berprestasi dengan algoritma *Naive bayes*. Prediksi digunakan untuk membantu meningkatkan kualitas belajar siswa yang akan mendatang dimana keuntungannya sekolah dapat mengeluarkan siswa yang unggul dan berprestasi dan memiliki peluang yang besar untuk mendapatkan beasiswa berprestasi .

2. Metode Penelitian

2.1 Data Mining

Data mining adalah salah satu cabang ilmu kecerdasan buatan dalam penggalian pola-pola untuk mengubah data menjadi informasi. *Data mining* merupakan proses yang mempekerjakan satu atau lebih teknik pembelajaran computer untuk menganalisa dan mengekstraksi pengetahuan (*knowledge*) secara otomatis. *Data mining* mampu mengolah data dengan jumlah yang besar dan dapat melakukan pencarian data secara otomatis, oleh karena itu data mining memiliki peranan yang sangat penting dalam beberapa bidang kehidupan diantaranya yaitu bidang industri, keuangan, cuaca, ilmu dan teknologi[4].

Berdasarkan penjelasan diatas data mining merupakan proses pengolahan data untuk menemukan pola-pola baru sehingga menghasilkan suatu informasi atau pengetahuan.

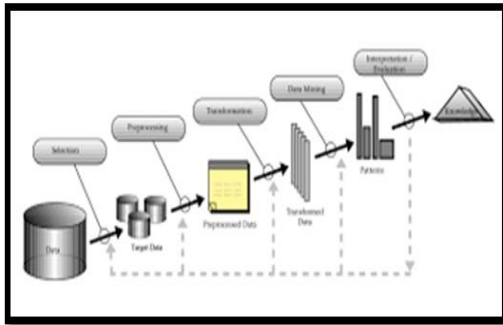
2.1.1 Proses Data Mining

Data mining, sering disebut juga sebagai *Knowledge Discovery in Database* (KDD) adalah suatu kegiatan dalam pengumpulan data, pemakaian data, riwayat untuk menemukan keteraturan, pola atau hubungan dalam suatu data yang berukuran

besar. Berikut ini adalah karakteristik data mining[5].

- Data mining berhubungan dengan penemuan sesuatu yang tersembunyi dan pola data yang sebelumnya tidak diketahui.
- Data mining biasa menggunakan data yang sangat besar.
- Data mining berguna untuk mengambil suatu keputusan yang kritis.

Tahapan-tahapan Data Mining:



Gambar 1 : Proses Data Mining

Proses data mining adalah sebagai berikut :

- Pembersihan Data (*Cleaning Data*) Penghilangan data yang tidak sesuai, tidak relevan dan tidak konsisten.
- Integrasi Data (*Data Integration*)Penggabungan dari beberapa database ke dalam suatu database yang baru.
- Seleksi Data (*Data Selection*) Data yang ada di dalam database seringkali tidak dipakai, oleh karena itu hanya data yang sesuai dan dibutuhkan yang akan di ambil dari database.
- Data Transformation Data di ubah atau di gabung dalam format yang sesuai untuk di proses dalam data mining.
- Proses Mining Suatu proses utama saat metode diterapkan untuk menemukan pengetahuan berharga yang tersembunyi dari data yang sangat besar.
- Evaluasi Pola Mengidentifikasi pola-pola menarik ke dalam *knowledge* yang ditemukan.

2.1.2 Metode Data Mining

Dalam Data Mining terdapat beberapa jenis metode sesuai dengan pemanfaatannya .Fungsi-fungsi yang umum diterapkan dalam data mining yaitu:

- Assosiation*, adalah proses untuk menemukan aturan asosiasi antara suatu kombinasi item dalam suatuwaktu.
- Sequence*, proses untuk menemukan aturan asosiasi antara suatu kombinasi item dalam suatu waktu dan diterapkan lebih dari satu periode.
- Clustering*, adalah proses pengelompokan sejumlah data/obyek ke dalam kelompok data sehingga setiap kelompok berisi data yang mirip.

- Classification*, proses penemuan model atau fungsi yang menjelaskan atau membedakan konsep atau kelas data, dengan tujuan untuk dapat memperkirakan kelas dari suatu objek yang lebelnya tidak diketahui.
- Regression*, adalah proses pemetaan data dalam suatu nilai prediksi.
- Forecasting*, adalah proses pengestimasi nilai prediksi berdasarkan pola-pola di dalam sekumpulan data.
- Solution*, adalah proses penemuan akar masalah dan problem solving dari persoalan bisnis yang dihadapi atau paling tidak sebagai informasi dalam pengambilan keputusan

2.2 Pengertian Algoritma

Algoritma adalah suatu ilmu yang mendeskripsikan cara menyelesaikan suatu permasalahan dengan menggunakan serangkaian langkah-langkah tertentu yang disusun secara sistematis dan menggunakan bahasa yang logis dan objektif. Algoritma merupakan suatu sistem kerja komputer yang dilengkapi dengan brainware, perangkat lunak, dan perangkat keras. Sistem komputer tidak akan bekerja dengan baik tanpa perangkat ini. Algoritma yang disebutkan dalam rujukan lain adalah sebuah prosedur dari serangkaian instruksi terpadu atau cara khusus untuk menyelesaikan masalah praktis[6].

2.3 Naive bayes

Naive bayes merupakan pengklasifikasian probabilistik sederhana yang menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari dataset yang diberikan. Algoritma menggunakan teorema Bayes dan mengasumsikan semua atribut independen atau tidak saling ketergantungan yang diberikan oleh nilai pada variabel kelas . Definisi lain mengatakan *Naive Bayes* merupakan pengklasifikasian dengan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya. *Naive Bayes* didasarkan pada asumsi penyederhanaan bahwa nilai atribut secara kondisional saling bebas jika diberikan nilai output. Dengan kata lain, diberikan nilai output, probabilitas mengamatisecara bersama adalah produk dari probabilitas individu. Keuntungan penggunaan *Naive Bayes* adalah bahwa metode ini hanya membutuhkan jumlah data pelatihan (*Training Data*) yang kecil untuk menentukan estimasi parameter yang diperlukan dalam proses pengklasifikasian. *Naive Bayes* sering bekerja jauh lebih baik dalam kebanyakan situasi dunia nyata yang kompleks dari pada yang diharapkan[7]. Keuntungan penggunaan adalah bahwa metode ini hanya membutuhkan jumlah data pelatihan (*Training data*) yang kecil untuk menentukan estimasi parameter yang diperlukan dalam proses

pengklasifikasian. Karena yang diasumsikan sebagai variable independent, maka hanya varians dari suatu variable dalam sebuah kelas yang dibutuhkan untuk menentukan klasifikasi, bukan keseluruhan dari matriks kovarians [8].

Kegunaan *Naive Bayes* mengklasifikasikan dokumen teks seperti teks berita ataupun teks akademis juga sebagai metode machine learning yang menggunakan probabilitas untuk membuat diagnosis medis secara otomatis mendeteksi atau menyaring spam. Kelebihan *Naive Bayes* bisa dipakai untuk data kuantitatif maupun kualitatif Tidak memerlukan jumlah data yang banyak Tidak perlu melakukan data *Training* yang banyak Jika ada nilai yang hilang, maka bisa diabaikan dalam perhitungan. Perhitungannya cepat dan efisien mudah dipahami, mudah dibuat pengklasifikasian dokumen bisa dipersonalisasi, disesuaikan dengan kebutuhan setiap orang jika digunakan dalam bahasa pemrograman, code-nya sederhana juga bisa digunakan untuk klasifikasi masalah biner ataupun *multiclass*.

Kekurangan *Naive Bayes* apabila probabilitas kondisionalnya bernilai nol, maka probabilitas prediksi juga akan bernilai nol asumsi bahwa masing-masing variabel independen membuat berkurangnya akurasi, karena biasanya ada korelasi antara variabel yang satu dengan variabel yang lain keakuratannya tidak bisa diukur menggunakan satu probabilitas saja. Butuh bukti-bukti lain untuk membuktikannya. Untuk membuat keputusan, diperlukan pengetahuan awal atau pengetahuan mengenai masa sebelumnya. Keberhasilannya sangat bergantung pada pengetahuan awal tersebut Banyak celah yang bisa mengurangi efektivitasnya dirancang untuk mendeteksi katakata saja, tidak bisa berupa gambar.

Persamaan dari teorema *Bayes* adalah :

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

Di mana :

- X* :Data dengan *class* yang belum diketahui
- H*: Hipotesis data merupakan suatu *class* spesifik
- P(H/X)*:Probabilitas hipotesis *H* berdasar 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*

2.4 Teori Confusion Matrix

Confusion matrix adalah alat ukur berbentuk matrix yang digunakan untuk mendapatkan jumlah ketepatan klasifikasi terhadap kelas dengan algoritme yang dipakai. Berikut akan disajikan bentuk confusion matrix pada Tabel di bawah ini [9].

Tabel 1. Bentuk Confusion Matrix dari Dua Kelas

Confusion Matrix		Nilai Sebenarnya	
		True	False
Nilai Prediksi	True	TP (<i>True Positive</i>) <i>Correct result</i>	FP (<i>False Positive</i>) <i>Unexpected result</i>
	False	FN (<i>False Negative</i>) <i>Missing result</i>	TN (<i>True Negative</i>) <i>Correct absence of result</i>

Pada Tabel 1 nilai TP (*True positive*) dan TN (*True negative*) menunjukkan tingkat ketepatan klasifikasi. Umumnya semakin tinggi nilai TP dan TN semakin baik pula tingkat klasifikasi dari akurasi, presisi, dan *recall* . Jika label prediksi keluaran bernilai benar (*True*) dan nilai sebenarnya bernilai salah (*False*) disebut sebagai *False positive* (FP). Sedangkan jika prediksi label keluaran bernilai salah (*False*) dan nilai sebenarnya bernilai benar (*True*) maka hal ini disebut sebagai *False negative* (FN) [10]. Berikut formulasi untuk menghitung akurasi, presisi, dan *recall* pada pembentukan model klasifikasi ditunjukkan pada Persamaan (1), Persamaan (2), dan Persamaan (3).

$$Akurasi = \frac{TP+TN}{TP+TN+F+P+FN} \times 100\% \tag{1}$$

$$Presisi = \frac{TP}{TP+FP} \times 100\% \tag{2}$$

$$Recall = \frac{TP}{TP+FN} \times 100\% \tag{3}$$

2.5 Cross validation

Cross validation adalah teknik validasi model untuk menilai keakuratan hasil analisis. Data yang sudah di praproses dilakukan *Cross validation* dengan membagi data menjadi data latih dan data uji untuk proses klasifikasi. Pembagian data dilakukan menggunakan *k-fold Cross validation* dengan nilai *k* sama dengan 5. Salah satu teknik *Cross validation* adalah *k-fold Cross validation* dimana *k* adalah bilangan bulat yang digunakan untuk membagi data. Jika nilai *k* = 5 maka data akan dibagi 5, dan proses pelatihan dan pengujian dilakukan sebanyak lima kali. Pada pelatihan dan pengujian pertama, data

bagian pertama akan digunakan sebagai data uji, dan sisanya menjadi data latih, sedangkan pada pelatihan dan pengujian kedua, data bagian kedua akan digunakan sebagai data uji, dan 4 bagian data lainnya sebagai data latih dan seterusnya[11].

2.6 Prediksi

Prediksi adalah suatu proses memperkirakan secara sistematis mengenai sesuatu yang mungkin akan terjadi di masa yang akan datang berdasarkan informasi masa lalu dan informasi saat ini yang dimiliki, agar error dan kesalahan dapat diperkecil. Dalam prediksi tidak harus memberikan jawaban yang pasti terkait kejadian yang akan datang, akan tetapi berusaha untuk mencari jawaban seakurat mungkin yang nantinya akan terjadi[12].

Prediksi atau peramalan (Forecasting) adalah kegiatan memperkirakan atau memprediksi apa yang akan terjadi pada masa yang akan datang dengan waktu yang relatif lama. Sedangkan ramalan adalah suatu situasi atau kondisi yang akan diperkirakan kan terjadi pada masa yang akan datang, peramalan atau prediksi adalah “seni dan ilmu untuk meperkirakan kejadian dimasa depan melalui pengujian di masa lalu”. Pengujian tersebut atas dasar pola-pola di waktu yang lalu dengan melibatkan pengambilan data masa lalu dan menempatkannya ke masa yang akan datang dengan model matematis. Prediksi atau peramalan berproses memperkirakan berapa kebutuhan dimasa datang yang meliputi kebutuhan kuantitas, kualitas, waktu dan lokasi yang dibutuhkan dalam rangka memenuhi permintaan barang atau jasa[13].

Berdasarkan uraian diatas Prediksi adalah suatu proses untuk memperkirakan kemungkinan yang akan terjadi dimasa akan datang dengan menggunakan kejadian yang telah terjadi sebelumnya.

3. Hasil dan Pembahasan

3.1 Hasil

Hasil dari penelitian ini adalah prediksi siswa berprestasi dengan nilai rata-rata raport tertinggi menggunakan metode algoritma *Naïve bayes* model yang dapat digunakan sebagai rekomendasi prediksi yang dinilai untuk melihat pola nilai siswa SMA Negeri 3 Pagar Alam dengan Metode Tahapan menggunakan *CRIPS-DM* yaitu Pemahaman Bisnis (*Business Understanding*), Pemahaman Data (*Data Understanding*), Pengolahan Data (*Data Preparation*), Pemodelan (*Modelling*), *Evaluation*, *Deployment*. Berdasarkan data dapat dilihat ada 6 siswa yang berprestasi dengan nilang yang tertinggi dan 105 siswa cukup berprestasi karena rata-rata nilai yang didapatkan lumayan tinggi sedangkan siswa kurang berprestasi terdapat 13 orang dengan nilai rata-rata sedikit rendah.

Kemudian dari tahapan tersebut dilakukan pengujian dengan menggunakan *Cross validation* untuk menghitung hasil tingkat keakuratan prediksi

prestasi siswa hasil dari pengujian tersebut didapatkan tingkat akurasi 94.55% dengan keterangan siswa berprestasi Berprestasi,cukup beprestasi dan kurang berprestasi untuk menghitung tingkat ke akuratan jumlah siswa berprestasi dilakukan tahapan pemodelan menggunakan RapidMiner sesuai dengan hasil prediksi *Naïve bayes* dengan pengujian *Cross validation*..

3.2 Tahapan CRIPS-DM

3.2.1 Pemahaman Bisnis (*Businnes Understanding*)

Pada pemahaman bisnis dilakukan langkah mentukan tujuan pencarian dan lingkungan pencarian dilakukan maksud pencarian ini adalah untuk Pengelompokan nilai rata-rata siswa data yang diambil dari excel data tersebut hanya digunakan untuk melihat nilai tertinggi dan terendah pada siswa SMA Negeri 3 Pagar Alam. Dari proses pengelompokan data tersebut belum pernah dilakukan pengolahan data lebih lanjut atau belum pernah dilakukan prediksi prestasi siswa dari nilai tersebut maka dari itu peneliti akan melakukan prediksi prestasi siswa dari hasil belajar siswa berupa nilai raport dengan menggunakan metode algoritma *Naïve bayes* untuk menegetahui tingkat prestasi dari hasil belajar siswa SMA Negeri 3 Pagar Alam.

3.2.2 Pemahaman Data (*Data Understanding*)

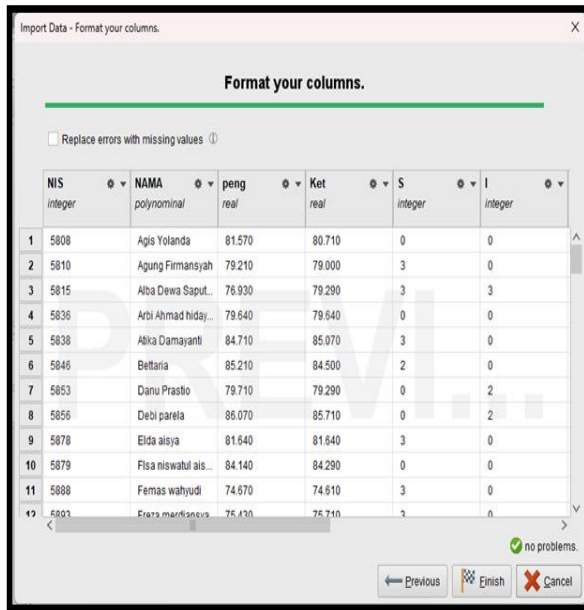
Pada fase pemahaman data ini data didapat dari wakil kesiswaan SMA Negeri 3 Pagar Alam yaitu data nilai raport siswa kelas 2 SMA Negeri 3 Pagar Alam tahun 2022-2023 dengan jumlah sampel 124 data siswa yang diambil 62 data siswa kelas ipa dan 62 lagi data dari kelas ips dengan jumlah atribut sebanyak 24 atribut kategori yang di terima dalam bentuk sub copy dan data perlu diketik ulang dalam bentuk file excel dan data yang didapat perlu di cleaning dan dan dilakukan pemilihan data untuk atribut yang digunakan.

3.2.3 Pengolahan Data (*Data Preparation*)

Pengolahan data langsung di implementasikan yang di RapidMiner dengan tiga tahapan yang ditetapkan pada pengolahan data Kemudian masukan data melalui read excel lalu pilih data yang akan digunakan untuk proses selanjutnya .

a. Data Selection

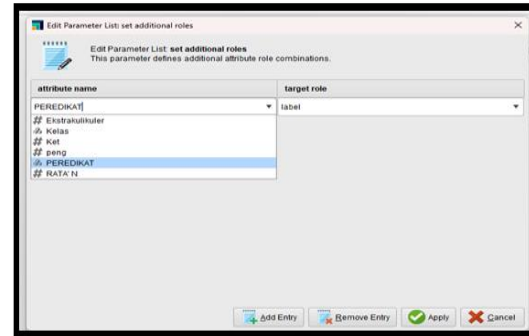
Pemilihan (seleksi) pengumpulan data informasi tentang data pekerjaan harus dilakukan sebelum dimulainya penambangan data KDD. Dilakukan pemlihan atribut dan sebagian data dalam atribut akan di konversikan agar proses data mining menjadi lebih mudah. Data yang diperoleh perlu di seleksi dengan cara di filter. Data yang difilter adalah data atribut yang akan dipakai . peneliti hanya memakai atribut yang dipilih seperti Nis, Nama, Peng, nilai keterangan dan pengetahuan, absensi, nilai rata-rata keseluruhan dan predikat.



	NIS	NAMA	peng	Ket	S	I
	integer	polynomial	real	real	integer	integer
1	5808	Agis Yolanda	81.570	80.710	0	0
2	5810	Agung Firmansyah	79.210	79.000	3	0
3	5815	Alba Dewa Saput...	76.930	79.290	3	3
4	5836	Arbi Ahmad hiday...	79.640	79.640	0	0
5	5838	Alka Damayanti	84.710	85.070	3	0
6	5846	Bettafia	85.210	84.500	2	0
7	5853	Danu Prasno	79.710	79.290	0	2
8	5856	Debi parela	86.070	85.710	0	2
9	5878	Elda aisyia	81.640	81.640	3	0
10	5879	Fisa niswahul ais...	84.140	84.290	0	0
11	5888	Femas wahyudi	74.670	74.610	3	0
12	5903	Frata marfanica	75.130	75.710	3	0

Gambar1.: Data Selection

Menambahkan set role untuk menentukan label dari data set yang digunakan untuk melakukan klasifikasi.



Gambar 2. Label Dan Atribut

b. Data Processing/Cleaning

Setelah tahap pengumpulan data dan filter data maka tahap selanjutnya yaitu cleaning data agar tidak ada duplikasi data, memeriksa data yang inkonsisten dan memperbaiki kesalahan pada data seperti kesalahan cetak, sehingga data tersebut dapat diolah dan dilakukan proses data mining. Setelah semua data yang dibutuhkan telah melalui tahap cleaning data maka penulis mendapatkan 124 data yang akan di olah menjadi 2 bagian yaitu data *Training* dan data *Testing* yang siap dimasukan pada software yang digunakan pada penelitian ini. Proses procesing merupakan proses yang mencakup cleaning dan formasi data. Pada tahap data processing peneliti sudah memastikan tidak ada lagi data yang kosong dan dapat digunakan.

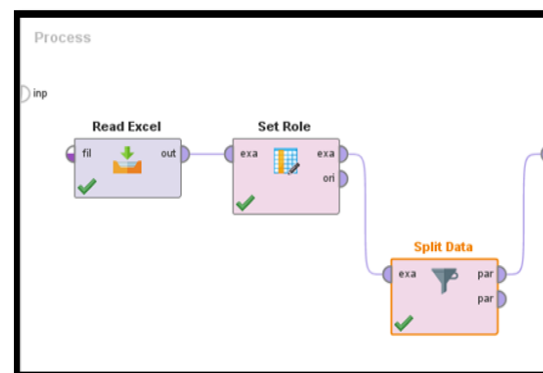
Pada tahap ini memilih atribut yang akan digunakan pada nilai raport siswa SMA Negeri 3 Pagar Alam. Penyeleksian menggunakan data dari excel dengan cara di filter dengan memilih atribut.

4. Pemodelan (Modelling)

Pada tahap ini akan dilakukan proses pemodelan awal pada prediksi prestasi siswa SMA Negeri 3 Pagar Alam dengan algoritma *Naive bayes* Menggunakan RapidMiner untuk tahap pertama yaitu mengimport data nilai siswa dari faile excel ke aplikasi RapidMiner menggunakan Read Excel.

Kemudian ke set role.

Gambar label dan atibut diatas untuk menampilkan atribut yang akan dijadikan lebel dalam klasifikasi. Setelah muncul form seperti gambar di atas lalu pilih atribut target atau label yang digunakan yaitu atribut redikat. Setelah ditentukan label yang dipilih maka klik Finish. Kemudian setelah kedua data sudah di import maka step selanjutnya yaitu drag and drop operators *Naive Bayes*, *Apply Model*, dan performace lalu hubungan dengan data *Training* dan *Testing*.



Gambar 3. Split Data

Pada tahap ini memasukan memasukan oprator dimana mengubah data manjadi data *Training* dan data *Testing*. Data yang digunakan 10% data *Training* dan 90% data *Testing* untuk data itu sendiri menggunakan rasio 0.1 banding 0.9 untuk data *Training* dan data *Testing* menggunakan rasio 0.9 bading 0.1 data untuk data *Training* jumlah data digunakan 13 data dan data *Testing* 111 data dapat dilihat pada gambar dibawah ini.

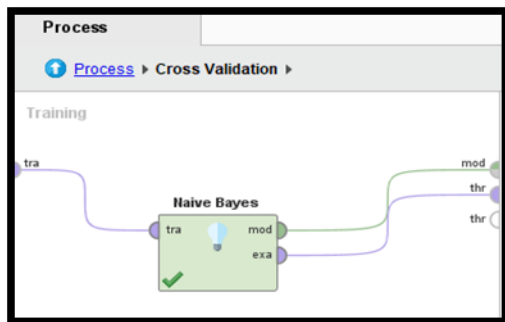
Row No.	PEREDIKAT	prediktaber	confidence	confidence	confidence	peng	Kat	Eksklusivitas	Kelas	BAKIR
1	CUKUP BER	KURANG BER	0.940	0.932	0	78.470	79.470	1	04	78.970
2	BERPRESTASI	CUKUP BER	0	1	0	89.890	89.500	1	04	89.880
3	CUKUP BER	KURANG BER	0.789	0.231	0	78.210	79	1	04	79.105
4	CUKUP BER	CUKUP BER	0.000	1.000	0	85.570	85.570	1	04	85.570
5	KURANG BER	CUKUP BER	0	1	0	77.430	78.290	1	04	77.880
6	CUKUP BER	CUKUP BER	0.000	1.000	0	82.070	82.140	2	04	82.105
7	KURANG BER	CUKUP BER	0	1	0	74.570	74.510	2	04	74.540
8	CUKUP BER	CUKUP BER	0	1	0	84.430	84.380	1	04	84.395
9	CUKUP BER	CUKUP BER	0.000	1.000	0	82.550	83.380	1	04	82.930
10	CUKUP BER	CUKUP BER	0.000	1.000	0	84.930	84.930	2	04	85.930
11	CUKUP BER	CUKUP BER	0.000	1.000	0	84.500	83.790	1	04	84.145
12	CUKUP BER	CUKUP BER	0.000	1.000	0	81.540	81.540	2	04	81.540
13	CUKUP BER	CUKUP BER	0.000	1.000	0	84.430	84.380	1	04	84.395

Gambar 4. Data Training

Row No.	PEREDIKAT	prediktaber	confidence	confidence	confidence	peng	Kat	Eksklusivitas	Kelas	BAKIR
1	BER	BER	0.907	0.909	0	81.710	82.710	04	04	82.160
2	BER	BER	0.907	0.909	0	78.760	78.280	04	04	79.030
3	CUKUP BER	CUKUP BER	0.864	0.816	0	75.760	75.870	04	04	75.830
4	CUKUP BER	CUKUP BER	1.000	0.999	0	73.930	73.740	04	04	73.830
5	BER	BER	0.900	1.000	0	82.930	85.110	04	04	84.020
6	BER	BER	0.900	1.000	0	83.140	81.710	04	04	82.420
7	BER	BER	0.900	1.000	0	82	82.630	04	04	82.910
8	BER	BER	0.900	1.000	0.000	84.140	85.290	04	04	85.270
9	BER	BER	0.900	1.000	0	82.070	82.140	04	04	82.105
10	BER	BER	0.900	1.000	0.000	84.430	84.380	04	04	84.395
11	SANGAT BER	SANGAT BER	0.900	0.900	1.000	89.890	89.500	04	04	89.880
12	CUKUP BER	CUKUP BER	1.000	0.999	0	74.570	74.510	04	04	74.540
13	BER	BER	0.900	0.900	0	81.540	82.070	04	04	82.030
14	BER	BER	0.900	1.000	0	81.760	81.500	04	04	81.645
15	BER	BER	0.940	0.944	0	78.190	79.190	04	04	78.920

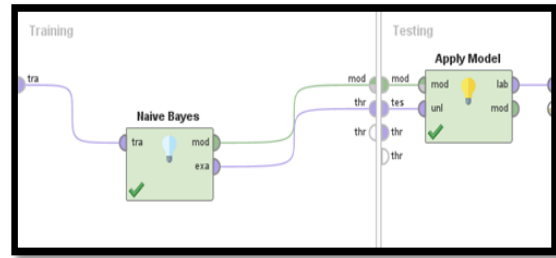
Gambar 5. Data Testing

Kemudian ke proses *Naive bayes* untuk mengklasifikasikan data set karena hasil prediksi diambil dari hasil klasifikasi. Dari hasil proses perhitungan menggunakan rapid miner dengan metode prediksi menampilkan hasil dari data *Testing* dan *Training* yang telah diuji. Kolom ini memberikan informasi tentang data siswa yang mendapatkan nilai maksimal dari hasil belajar tahun 2022 dan tahun 2023. Selanjutnya untuk mengetahui tingkat *accuracy* algoritma *Naive Bayes*,



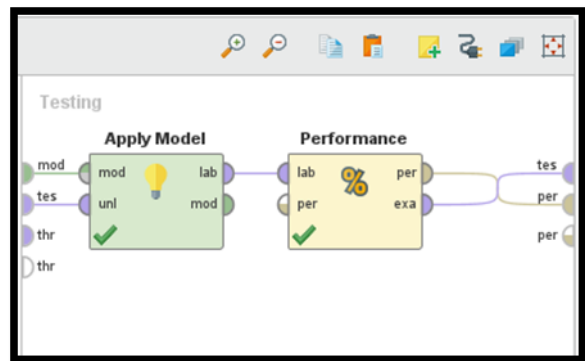
Gambar 6. Proses Naive Bayes

Untuk menjalankan model yang dibuat kita membutuhkan *Apply Model* yang menghubungkan operator *naive bayes* ke *Apply Model* seperti gambar dibawah ini.



Gambar 7. Apply Model

Kemudian menambahkan operator *Performance* untuk menampilkan tingkat mengukur akurasi model yang dibuat.



Gambar 8. Performance

Kemudian hasil akhir dari keluaran (*output*) dari sistem pada pohon keputusan yang telah dijalankan pada tahap akhir yang menampilkan keputusan pada RapidMiner. Hasil akhir yang ditampilkan adalah berupa *sample distribution* yang menentukan banyaknya nilai dari data berprestasi, cukup berprestasi, berprestasi dan menjelaskan bahwa siswa yang berprestasi memiliki nilai 0.045 sedangkan yang cukup berprestasi 0.7874 . dan yang kurang berprestasi 0.171 berdasarkan data hasil klasifikasi yang didapatkan.

```

SimpleDistribution

Distribution model for label attribute PEREDIKAT

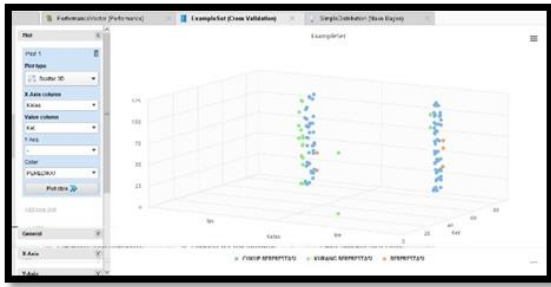
Class CUKUP BERPRESTASI (0.784)
5 distributions

Class KURANG BERPRESTASI (0.171)
5 distributions

Class BERPRESTASI (0.045)
5 distributions
    
```

Gambar 9. Hasil Akhir

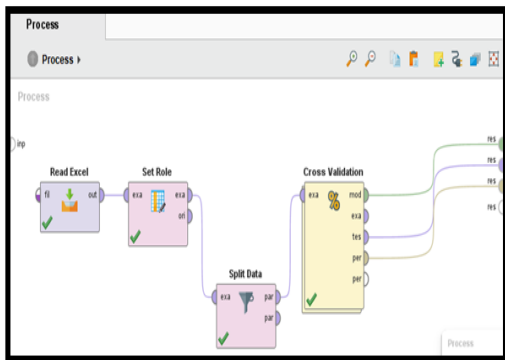
Tahap selanjutnya yaitu menampilkan hasil klasifikasi dengan grafik dibawah ini menjelaskan bahwa pada titik orange menampilkan siswa berprestasi sedangkan warna biru cukup berprestasi dan yang hijau itu siswa yang kurang berprestasi.



Gambar 10. Grafik Klasifikasi

5. Evaluation

Melakukan evaluasi dengan pengujian cross validation untuk menampilkan akurasi dari kalsifikasi untuk madapatkan hasil prediksi. Selanjutnya adalah mengatur operator *Cross validation* dengan cara double klik operator *Cross validation* Kemudian drag dan drop operator *Naive Bayes* dan hubungkan port tra dan mod pada kolom *Training*. Selanjutnya drag and drop operator *Apply Model* dan *Performance*, kemudian hubungkan port mod, tes pada operator *Apply Model* dan lab ke operator *Performance* pada kolom *Testing*. Pada operator *Performance* hubungkan port per ke port per dan port exa dengan port tes, Setelah semuanya telah terhubung dan parameter operator *Cross validation* terisi, klik Run Untuk menampilkan perhitungan.



Gambar 11 : Cross validation

Pada tahap ini adalah hasil akhir dari kalsifikasi yang menentukan hasil Prediksi siswa berprestasi dan ini adalah hasil dari Pengujian model yang dibuat dapat dilihat pada gambar dibawah ini untuk class recall True baik 95.40% dan class predicsion 97.65% sedang class recal sangat baik 100.00% dan class 55.56% untuk class recall cukup baik ada 89.47 % dan hasil predicsion 100.00% Berdasarkan hasil tabel diatas didapatkan hasil *Performance* dari metode yang telah digunakan peneliti pada prediksi siswa berprestasi menggunakan data nilai raport siswa SMA Negeri 3 Pagar Alam tahun 2022-2023 menpakatkan nilai akurasi 94.55% sehingga dapat digunakan untuk prediksi siswa yang berprestasi.

Tabel 1: *Performance*

	True Cukup Berprestasi	True Kurang Berprestasi	True Berprestasi	Class Precison
Pred. Cukup Berprestasi	83	2	0	97.55%
Pred. kurang Berprestasi	0	17	0	100.00 %
Pred. Berprestasi	4	0	5	55.56%
Class recall	95.40%	89.47%	100.00%	

Accuracy.94.55% +/- 9.77%(micro average:94.59%)

6. Deployment

Pada tahap ini aladah tahap akhir dari proses pengolahan dimana hasil dari proses pengolahan data dapat dilihat dari tahap ini yaitu prediksi siswa berprestasi dan kurang berprestasi. Dibawah ini dapat dilihat siswa berprestasi, cukup berprestasi dan kurang berprestasi berdasarkan Nis, Nama, absensi dan nilai rata-rata siswa kelas 11 SMA Negeri 3 Pagar Alam.

7. Kesimpulan

Berdasarkan hasil dan pembahasan dapat disimpulkan bahwa prediksi siswa berprestasi menggunakan algoritma *naive bayes* bisa digunakan berdasarkan hasil pengolahan data siswa ini dengan menggunakan Metode *Naive bayes* dan pengujian *Cross validation* didapatkan Class recall siswa dengan nilai *Performance* siswa kelas berprestasi 100 % class predcsion 55.56% untun siswa cukup berprestasi 95.40% dengan class predcsion 97.65% dan yang kurang berprestasi class recall 89.47% class precision 100.00% pada hasil *Performance* hasil tingkat akurasi 94.55% pada class Predicsion Prestasi siswa.

8. Saran

Pada penelitian ini masih banyak kekurangan juga kelemahan yang perlu dikembangkan untuk penelitian selanjutnya semoga jauh lebih baik dan lebih sempurna dari penelitian ini. Adapun saran dari peneliti yaitu sebagai berikut :

- Pada penelitian ini terdapat banyak kekurangan dan kelemahan bagi peneliti selnjutnya bisa membuat sistem agar lebih mudah melakukan prediksi.
- Perlu menggunakan banyak data dan atribut yang lebih dari peneliti ini.
- Bisa mencoba menggunakan algoritma lain sebagai perbandingan untuk menemukan algoritma yang sangat cocok untuk melakukan prediksi siswa berprestasi..

Daftar Pustaka

- [1] Almufqi, F. M., Voutama, A., Informasi, S., Komputer, I., Karawang, U. S., & Perbandingan Metode *Data Mining* Untuk Memprediksi Presatasi. *15*(1), 61–66. <https://doi.org/10.30736/jt.v15i1.929>
- [2] Daulay, P. I., Islam, U., Sumatera, N., Islam, U., & Sumatera, N. (2023). *Penerapan Algoritma Pemrograman dalam Pembelajaran Ilmu Komputer Pinkan Indriani Daulay Yahfizham*. *1*(6).
- [3] Fauzia, N. S., & Dana, R. D. (2023). *Implementasi Algoritma Naive Bayes dalam Klasifikasi Status Kesejahteraan Masyarakat Desa Gunungsari*.
- [5] Hasanah, M. A., Soim, S., & Handayani, A. S. (2021). *Implementasi CRISP-DM Model Menggunakan Metode Decision Tree dengan Algoritma CART untuk Prediksi Curah Hujan Berpotensi Banjir*. *5*(2).
- [6] Hasudungan, R., & Pranoto, W. J. (2021). *Implementasi Teorema Naive bayes Pada Prediksi Prestasi Mahasiswa*. *5*(1), 10–16.
- [8] Informatika, J. T., Dan, P. A. K., Persero, P. T. P., Mega, J., Barat, K., Kawasan, L., & Kuningan, M. (2021). *Berdasarkan Tingkat Kepentingan Pada Prodi / Jurusan D4 Teknik*.
- [9] Iyah, S., Metode, M., & Series, T. (n.d.). *Prediksi Penerimaan Siswa Baru Pada MADRASAH ALIYAH AS-*.
- [10] Kadek, N., Patrianingsih, W., & Sugianta, I. K. A. (2023). *Penerapan Naive bayes pada Potensi Akademik Siswa SD Negeri 5 Singakerta*. *8*(2), 154–163.
- [11] Lestari, S., & Badrul, M. (2020). *Implementasi Klasifikasi Naive Byes Untuk Prediksi*. *7*(1), 8–16.
- [12] Mardiana, L., Kusnandar, D., & Satyahadewi, N. (2022). *Analisis Diskriminan Dengan K Fold Cross validation Unruk Klasifikasi Kualitas Air DI KOTA PONTIANAK*. *11*(1), 97–102.
- [13] Mesin, J. T., Dan, E., Jtmei, I., & Afyudin, A. (2024). *Analisis Peramalan Kebutuhan Komponen Minifix Bolt Dia di PT XYZ Menggunakan Metode Time Series*. *3*(1).
- [14] Navisa, S., Hakim, L., & Nabilah, A. (2021). *Komparasi Algoritma Klasifikasi Genre Musik pada Spotify Menggunakan CRISP-DM. Jurnal Sistem Cerdas*, *4*(2), 114–125.
- Perbandingan Data Untuk Klasifikasi Wine Menggunakan Agoritma K-NN*. (2019). *4*(1), 78–82.
- [15] Prima, J., Sistem, J., Komputer, I., & No, V. (2021). *Penerapan Metode Naive Bayes Dalam Memprediksi Prestasi Siswa Di SMA Negeri 1 Panombeian Panei*. *4*(2).
- [16] Rafi Nahjan, M., Nono Heryana, & Apriade Voutama. (2023). *Implementasi Rapidminer Dengan Metode Clustering K-Means Untuk Analisa Penjualan Pada Toko Oj Cell. JATI (Jurnal Mahasiswa Teknik Informatika)*, *7*(1), 101–104. <https://doi.org/10.36040/jati.v7i1.6094>
- [17] Rifai, M. F., Jatnika, H., & Valentino, B. (2019). *Penerapan Algoritma Naive bayes Pada Sistem Prediksi Tingkat Kelulusan Peserta Sertifikasi Microsoft Office Specialist (MOS)*. *12*(2), 131–144.
- [18] Rosaly, R. (n.d.). *Pengertian Flowchart Beserta Fungsi dan Simbol-simbol Flowchart yang Paling Umum Digunakan*.
- [19] Teknorama, J., El, T., & Rahma, S. El. (2023). *Prediksi Lanjut Studi Mahasiswa Menggunakan Algoritma Naive bayes*. *1*(1), 30–34.
- [20] Utama, T. P., & Haibuan, M. S. (2023). *Penerapan Algoritma Naive Bayes Dan Forward Selection Prediksi Penyakit Stroke*. *17*, 3051–357.
- [21] Watratan, A. F., B, A. P., Moeis, D., Informasi, S., & Makassar, S. P. (2020). *JOURNAL OF APPLIED COMPUTER SCIENCE AND TECHNOLOGY (JACOST) Implementasi Algoritma Naive Bayes Untuk Memprediksi Tingkat Penyebaran Covid-19 Di Indonesia*. *1*(1), 7–14.
- [22] Wiguna, S., Islam, P. A., Artikel, I., App, M., & Ability, W. (2022). *Implementasi Penggunaan Aplikasi Mendeley dalam Membangun Kemampuan Menulis Karya Tulis Ilmiah pada Mahasiswa STAI Jam ' iyah Mahmudiyah Tanjung Pura Langkat*. *1*(3). <https://doi.org/10.54259/diajar.v1i3.965>
