

## IMPLEMENTASI ALGORITMA NAÏVE BAYES TERHADAP PENERIMA KARTU INDONESIA PINTAR

Weni Agustin<sup>1</sup>, Agus Bahtiar<sup>2</sup>

<sup>1</sup> Program Studi Manajemen Informatika, STMIK Ikmi Cirebon

<sup>2</sup> Program Studi Komputerisasi Akuntansi, STMIK Ikmi Cirebon

Jalan Perjuangan No. 10 B Majasem, Kota Cirebon

weniagustin202@gmail.com

### ABSTRAK

Pentingnya peran Kartu Indonesia Pintar (KIP) sebagai program pemerintah untuk mendukung akses pendidikan bagi siswa berprestasi yang memiliki keterbatasan finansial. Analisis ini diharapkan dapat memberikan kontribusi terhadap perbaikan proses seleksi penerimaan KIP, sehingga dukungan finansial dapat disalurkan secara lebih dan tepat sasaran. KIP merupakan program bantuan pendidikan yang dirancang untuk mendukung siswa yang berprestasi dari latar belakang keterbatasan ekonomi bertujuan untuk memberikan program tunjangan pendidikan. Dengan metode Algoritma *Naïve Bayes* bisa memudahkan dalam menghadapi penerimaan program bantuan pendidikan. Dalam mempengaruhi penerimaan Kartu Indonesia Pintar (KIP) termasuk nilai ujian, prestasi, penghasilan orangtua, punya kartus sejenis seperti KIP, dan kepemilikan rumah yang ditempati. Dengan mengembangkan model yang dapat memprediksi penerimaan KIP dengan melihat tingkat akurasi yang tinggi menggunakan algoritma *Naive Bayes*. Oleh karena itu ada beberapa poin yang menjadi dasar penelitian ini untuk mengidentifikasi Penerimaan KIP melalui dengan Hasil tujuan yang pertama menganalisis dan mengevaluasi algoritma *Naïve Bayes* dalam menentukan penerimaan kartu Indonesia pintar (KIP). Dari hasil penentuan ada beberapa siswa yang lulus dan tidak lulus selain itu ada juga hasil prediksi atau bisa disebut memprediksi kemungkinan lulus dan tidak lulus. Pada gambar di 4.17 hasil dari proses penyeleksian menunjukkan prediksi tidak lulus dengan keterangan *true* tidak lulus sekitar 112 siswa, dan prediksi yang *true* lulus sekitar 15 siswa. Dan prediksi lulus dengan keterangan *true* tidak lulus sekitar 23 siswa, yang *true* lulus sekitar 50 siswa. Dan terdapat nilai akurasi sebesar 81.00%. Dengan *class recall* yang *true* tidak lulus sebesar 82.96% sedangkan yang lulus *calass recall* nya hanya mendapatkan 76.92%. Dari hasil analisis tersebut bisa disimpulkan bahwa yang *true* tidak lulus mendapat angka paling tertinggi dengan *class racall* nya sebesar 82.96%. Dan hasil tujuan yang kedua Mengukur tingkat akurasi dari model prediktif yang dikembangkan menggunakan algoritma *Naive Bayes* pada dataset 1000 siswa. Hasilnya adaalah bisa dilihat Pada gambar 4.18 akurasi yang didapat adalah 81.00% dengan split data 80/20. dengan rincian Prediksi Tidak lulus dan *True* Tidak lulus sebesar 112, Prediksi lulus dan *True* lulus sebesar 15, Prediksi Lulus dan *True* Tidak sebesar 23, Prediksi Lulus dan *True* Lulus sebesar 50. Dengan *class recall True* Tidak Lulus sebesar 82.96% dan *class recall True* Lulus sebesar 76.92%. eksplorasi data fokus studi pada analisis pendaftaran KIP dan menekankan perlunya memahami dan memperbaiki proses yang menentukan penerimaan bantuan pendidikan penekanan pada penggunaan algoritma *Naïve Bayes* mencerminkan metode klasifikasi yang dipilih untuk mengembangkan model.

**Kata kunci :** *Kartu Indonesia Pintar (KIP), klasifikasi, Naive Bayes.*

### 1. PENDAHULUAN

Dalam penelitian ini peneliti menggunakan metode algoritma *Naïve Bayes*. *Naïve bayes* merupakan metode pengklasifikasian paling populer digunakan dengan tingkat keakuratan yang baik. Banyak penelitian tentang pengklasifikasian yang telah dilakukan dengan menggunakan algoritma ini. Berbeda dengan metode pengklasifikasian dengan logistic regression ordinal maupun nominal, pada algoritma *naive bayes* pengklasifikasian tidak membutuhkan adanya pemodelan maupun uji statistik. Pada algoritma ini pembelajaran lebih ditekankan pada pengestimasi probabilitas. Keuntungan algoritma *naive bayes* adalah tingkat nilai error yang didapat lebih rendah ketika dataset berjumlah besar, selain itu akurasi *naive bayes* dan kecepatannya lebih tinggi pada saat diaplikasikan ke dalam dataset yang

jumlahnya lebih besar. dilatarbelakangi oleh pentingnya peran Kartu Indonesia Pintar (KIP) sebagai program pemerintah untuk mendukung akses pendidikan bagi siswa berbakat yang memiliki keterbatasan finansial. Pendidikan merupakan landasan terpenting bagi pembangunan masyarakat dan negara, dan pemerintah Indonesia mengambil langkah nyata dengan memberikan dukungan finansial melalui program KIP.

Namun, mengingat jumlah siswa yang memenuhi kriteria dan keterbatasan anggaran, efisiensi proses seleksi penerimaan KIP perlu ditingkatkan. Untuk meningkatkan efisiensi dan akurasi seleksi, penggunaan algoritma klasifikasi khususnya *Naive Bayes* dapat menjadi solusi. Algoritma ini telah terbukti berhasil dalam berbagai aplikasi, termasuk analisis data yang besar dan kompleks. Penelitian ini bertujuan untuk memahami faktor-faktor yang

mempengaruhi keputusan penerimaan KIP dengan menerapkan *Naive Bayes* pada dataset sebanyak 1.000 siswa. Analisis ini diharapkan dapat memberikan kontribusi yang signifikan terhadap perbaikan proses seleksi penerimaan KIP, sehingga dukungan finansial dapat disalurkan secara lebih tepat sasaran. Hasil penelitian ini diharapkan dapat memberikan wawasan lebih dalam mengenai pola dan variabel yang mempengaruhi keputusan penerimaan KIP. Dengan cara ini, kami dapat memperbaiki dan menyesuaikan kriteria seleksi untuk memastikan bahwa bantuan keuangan menjangkau siswa yang benar-benar membutuhkannya.

Selain itu, penelitian ini juga dapat memberikan landasan yang kuat untuk mengembangkan metode analisis yang lebih canggih dan efektif dalam menentukan penerimaan KIP di masa depan. Pendidikan merupakan aspek penting dalam pembangunan suatu negara, dan pemerintah Indonesia memberikan dukungan finansial melalui program seperti Kartu Indonesia Pintar (KIP) untuk meningkatkan akses terhadap pendidikan bagi siswa berbakat namun memiliki keterbatasan finansial. Seiring kemajuan teknologi, pendekatan analisis data menjadi semakin penting untuk meningkatkan efisiensi penerapan program-program ini. Tujuan dari penelitian ini adalah untuk memberikan kontribusi yang signifikan terhadap pemahaman pola dan faktor yang mempengaruhi penerimaan KIP dengan menerapkan algoritma klasifikasi, khususnya *Naive Bayes*, pada dataset sintetik yang terdiri dari dataset 1.000 siswa.

Analisis *Naive Bayes* diharapkan dapat memberikan prediksi yang akurat mengenai siswa mana yang berhak mendapatkan bantuan pendidikan melalui KIP dengan membedah variabel seperti nilai ujian, prestasi, dan faktor ekonomi. Penggunaan metode ini bertujuan untuk meningkatkan transparansi dan objektivitas proses seleksi penerimaan, mengurangi risiko kesalahan manusia, dan memastikan dukungan keuangan disalurkan tepat sasaran. Sebagai langkah pertama, pemahaman yang lebih baik tentang potensi metode-metode ini untuk meningkatkan efektivitas program dukungan pendidikan dapat memberikan landasan yang kuat untuk perbaikan kebijakan dan prosedur di masa depan.

## 2. TINJAUAN PUSTAKA

### 2.1. Penerimaan Kartu Indonesia Pintar (KIP)

Kartu Indonesia Pintar (KIP) harus tepat sasaran agar dapat dimanfaatkan sebaik-baiknya. Daerah diharapkan memilih pemerintah yang bertanggung jawab menyalurkan KIP dengan sebaik-baiknya agar tidak terjadi kesalahan sasaran. Sebab di setiap daerah di Indonesia banyak sekali masyarakat yang sangat membutuhkan bantuan dari pemerintah, sangat membutuhkan uang dan biaya untuk memperlancar keuangan rumah tangganya [1] Dalam pemberian beasiswa perlu diperhatikan kemampuan penerima,

memenuhi persyaratan pengajuan, kemampuan finansial dan kemampuan akademik sesuai dengan ketentuan sistem beasiswa masing-masing. *INSTIKI* merupakan salah satu universitas yang menyalurkan beasiswa kepada mahasiswa dan saat ini belum ada metode yang sistematis dan objektif untuk menyeleksi penerima beasiswa. Beasiswa adalah sebuah format Kompensasi bantuan keuangan diberikan kepada seseorang organisasi untuk prestasi kerja dihasilkan. Hadiyahnya adalah akses ke institusi dalam bentuk dana khusus atau bantuan keuangan kegiatan penelitian [2]

### 2.2. Naïve Bayes

*Naïve Bayes* adalah pengklasifikasian statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu *class*. Metode Bayes merupakan pendekatan *statistic* untuk melakukan inferensi induksi pada persoalan klasifikasi. *Naive Bayes classifier* merupakan salah satu metode pembelajaran mesin yang memanfaatkan perhitungan probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi probabilitas pada masa depan berdasarkan pengalaman pada masa sebelumnya. (Widianto, 2019) [3]. Menurut (Naomi Chatrina Siregar, 2020) *Naive Bayes* adalah salah satu metode pengklasifikasian sederhana yang sering digunakan karena mudah diterapkan dan memiliki hasil yang baik saat diterapkan pada banyak kasus. Sedangkan kelemahan dari metode ini sendiri yaitu adanya asumsi atau dengan kata lain kondisi kelas saling bebas, sehingga kurang akurat. Adapun pada prakteknya, kebergantungan ada diantara variabel. Misalnya rumah sakit: pasien, umur, keluarga ,dsb. Kebergantungan diantara variabel ini tidak dapat dimodelkan dengan menggunakan *Naive Bayesian Classifier*. [5]

Rumus dari *teorema bayes* akan disajikan pada persamaan berikut :

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

### 2.3. Flowchart Naïve Bayes

Sistem *Naive Bayes* pada Flowchart tersebut merupakan cara penulisan algoritma dengan menggunakan notasi garis. Flowchart merupakan gambar atau bagian yang memperlihatkan langkah-langkah dari suatu program dan hubungan antar proses beserta pernyataannya. Gambaran ini dinyatakan dengan symbol. Dengan demikian setiap simbol menggambarkan proses tertentu. Sedangkan antar proses digambarkan dengan garis penghubung. Dengan menggunakan Flowchart akan memudahkan kita untuk melakukan pengecekan bagian-bagian yang terlupakan dalam analisis masalah. Pada Flowchart *Naive Bayes* suatu proses klasifikasi data. Meskipun detail spesifik dari Flowchart mungkin berbeda, ada yang mencakup langkah-langkah seperti menghitung probabilitas, membagi kumpulan data, mengklasifikasikan data, dan

menampilkan hasil klasifikasi. Flowchart memberikan representasi visual dari tahapan berurutan dari algoritma Naïve Bayes, mulai dari input data hingga hasil klasifikasi akhir Algoritma Naïve Bayes merupakan algoritma klasifikasi berdasarkan teorema Bayesian dalam statistik.

#### 2.4. Data Mining

Data mining menganalisis data menggunakan tool untuk menemukan pola dan aturan dalam himpunan data. Perangkat lunak bertugas untuk menemukan pola dengan mengidentifikasi aturan dan fitur pada data. Budi Santosa menjelaskan bahwa data mining adalah kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menemukan keteraturan pola dan hubungan dalam set data berukuran besar. Suatu organisasi dapat dibanjiri dengan berbagai macam data, sangatlah tidak berguna data yang terkumpul dan tersimpan jika tidak dimanfaatkan. Berdasarkan tugasnya data mining dikelompokkan menjadi deskripsi, estimasi, prediksi klasifikasi [6]

Data mining adalah proses pengumpulan dan pengolahan data yang bertujuan untuk mengekstrak informasi penting pada data. Data mining memiliki tiga tujuan yaitu sebagai sarana untuk menjelaskan atau *explanatory*, untuk konfirmasi atau *confirmatory*, dan untuk eksplorasi atau *exploratory*. Ia juga memiliki beberapa metode seperti *Association*, *Classification*, *Regression*, dan *Clustering*. *Association* adalah Teknik yang pertama adalah *association*. *Association* adalah metode berbasis aturan yang digunakan untuk menemukan asosiasi dan hubungan variabel dalam satu set data. Biasanya analisis ini terdiri dari pernyataan “if atau then” sederhana. *Association* banyak digunakan dalam mengidentifikasi korelasi produk dalam keranjang belanja untuk memahami kebiasaan konsumsi pelanggan. Sehingga, perusahaan dapat mengembangkan strategi penjualan dan membuat sistem rekomendasi yang lebih baik. *Classification* adalah metode yang paling umum digunakan dalam data mining. *Classification* adalah tindakan untuk memprediksi kelas suatu objek. *Regression* adalah teknik yang menjelaskan variabel dependen melalui proses analisis variabel independen. Sebagai contoh, prediksi penjualan suatu produk berdasarkan korelasi antara harga produk dengan tingkat pendapatan rata-rata pelanggan. *Clustering*

Terakhir, metode clustering. *Clustering* digunakan dalam membagi kumpulan data menjadi beberapa kelompok berdasarkan kemiripan atribut yang dimiliki. Contoh kasusnya adalah *Customer Segmentation*. Ia membagi pelanggan ke dalam beberapa grup berdasarkan tingkat kemiripannya [4]

#### 2.5. Klasifikasi

Klasifikasi data merupakan pengkategorian data yang diperoleh berdasarkan bagian-bagian penelitian yang telah ditetapkan. Klasifikasi data dilakukan agar

terdapat suatu batasan mengenai bahasan yang akan diteliti. Klasifikasi adalah metode pembelajaran data yang digunakan untuk prediksi suatu nilai yang ada pada sekumpulan atribut. Klasifikasi biasanya memperoleh seperangkat rules biasa disebut aturan. Aturan ini dipakai untuk metrik guna mendapatkan prediksi suatu objek dari data yang akan dilakukan prediksi [7] Klasifikasi merupakan suatu pekerjaan menilai objek data untuk memasukkannya ke dalam kelas tertentu dari sejumlah kelas yang tersedia. Dalam klasifikasi ada dua pekerjaan utama yang dilakukan, yaitu Pembangunan model sebagai *prototype* untuk disimpan sebagai memori dan Penggunaan model tersebut untuk melakukan pengenalan/klasifikasi/prediksi pada suatu objek data lain agar diketahui dikelas mana objek data tersebut dalam model yang sudah disimpannya [8]

#### 2.6. Rapid Miner

Pada proses ini setelah peneliti melakukan proses perhitungan secara manual, proses selanjutnya adalah melakukan pengujian dari hasil perhitungannya yang dilakukan secara manual, dengan menggunakan bantuan tools *rapidminer*. Yang fungsinya adalah untuk melakukan perbandingan tingkat akurasi pada diagnose [9] *RapidMiner* adalah platform perangkat lunak data ilmu pengetahuan yang dikembangkan oleh perusahaan dengan nama yang sama, yang menyediakan lingkungan terpadu untuk pembelajaran mesin (*machine learning*), pembelajaran mendalam (*deep learning*), penambangan teks (*text mining*), dan analisis prediktif (*predictive analytics*). Aplikasi ini digunakan untuk aplikasi bisnis dan komersial serta untuk penelitian, pendidikan, pelatihan, pembuatan *prototype* dengan cepat, dan pengembangan aplikasi serta mendukung semua langkah proses pembelajaran mesin termasuk persiapan data, visualisasi hasil, validasi dan pengoptimalan. *RapidMiner* dikembangkan dengan model open core [10]

### 3. METODE PENELITIAN

#### 3.1. Sumber Data

Penelitian ini menggunakan dataset sintetik berdasarkan karakteristik siswa dan parameter terkait perolehan Kartu Indonesia Pintar (KIP). Sumber data ini diperoleh dari *Kaggle*, sebuah platform online yang menyediakan berbagai dataset untuk keperluan penelitian dan pengembangan. Data yang digunakan dalam penelitian ini berjenis data sekunder. Link pengambilan datanya saya lampirkan dibawah ini: (<https://www.kaggle.com/datasets/gedewahyupurnama/student-data-kip>).

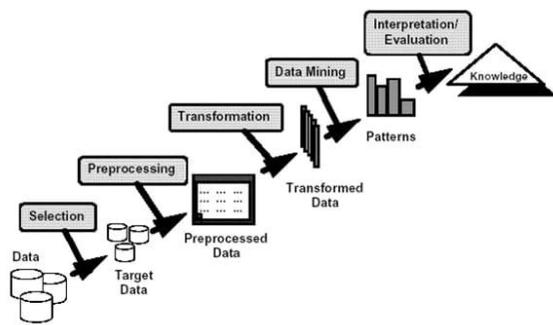
#### 3.2. Teknik Pengumpulan

Penelitian ini menggunakan dataset berdasarkan karakter siswa dan parameter terkait perolehan Kartu Indonesia Pintar (KIP) dengan 1.000 data. Sumber data ini diperoleh dari *Kaggle*, sebuah platform online yang menyediakan berbagai dataset untuk keperluan

penelitian dan pengembangan. Data yang digunakan dalam penelitian ini berjenis data sekunder.

**3.3. Tahap Perancangan**

Fase desain menggunakan *Knowledge Discovery in Databases* (KDD) adalah proses sistematis dalam mengekstraksi pengetahuan dan pola berharga dari kumpulan data besar. KDD menyertakan beberapa langkah penting yang diambil untuk lebih memahami data anda pada gambar dibawah ini.



Gambar 1. Tahapan perancangan

Berikut adalah penjelasan mengenai tahapan perancangan :

**3.3.1. Data Selection**

Seleksi data adalah proses memilih subkumpulan data tertentu yang relevan untuk dianalisis atau diproses lebih lanjut. Tujuan pemilihan data adalah untuk fokus pada bagian penting dari kumpulan data dan mengabaikan data yang tidak perlu atau tidak relevan. Proses pemilihan data mungkin melibatkan pemilihan baris (pengamatan) atau kolom (atribut) dari kumpulan data. Penggunaan metode pemilihan data bergantung pada konteks dan tujuan analisis yang diinginkan. Metode pemilihan data yang umum mencakup pemfilteran, pemilihan fitur, atau teknik lainnya bergantung pada karakteristik kumpulan data dan jenis analisis yang dilakukan.

**3.3.2. Data Pre-processing (Preprocessing)**

Pemrosesan awal data (*preprocessing*) adalah serangkaian langkah atau proses yang dilakukan pada data sebelum digunakan dalam suatu analisis atau model pembelajaran mesin. Tujuan utama pemrosesan awal data adalah untuk membersihkan, mengatur, dan mengubah data ke dalam format yang sesuai untuk

analisis dan pelatihan model.

**3.3.3. Data Transformation**

*Data Transformation* adalah proses mengubah atau memodifikasi nilai data dalam suatu dataset. Tujuan utama transformasi data adalah untuk memperbaiki atau mengubah distribusi data, mengurangi anomali, atau menyiapkan data untuk memenuhi persyaratan analisis atau pemodelan yang Anda lakukan. Transformasi data sangat penting ketika mempersiapkan data sebelum dianalisis atau dimodelkan. Memilih teknik transformasi yang tepat bergantung pada karakteristik data Anda, distribusi variabel, dan tujuan analisis atau pemodelan yang Anda inginkan.

**3.3.4. Data Mining**

*Data Mining* adalah proses ekstraksi pola-pola yang bermanfaat atau pengetahuan yang mendalam dari sekumpulan data besar atau kompleks. Tujuannya adalah untuk mengungkap informasi yang tersembunyi, relevan, dan berharga yang dapat digunakan untuk pengambilan keputusan. *Data Mining* menggunakan teknik analisis statistik, matematika, kecerdasan buatan, dan metode lainnya untuk mengeksplorasi dan menganalisis data, mengidentifikasi pola atau tren, serta memahami hubungan yang dapat memberikan wawasan yang berharga.

**3.3.5. Data Evaluation**

*Evaluation* dalam konteks analisis data atau pemodelan merujuk pada langkah-langkah untuk memahami dan mengevaluasi hasil dari suatu analisis atau pemodelan. Ini melibatkan penafsiran terhadap temuan atau output yang dihasilkan dari suatu proses analisis data atau pembuatan model. *Evaluasi* merupakan tahapan kritis dalam siklus analisis data atau pembuatan model. Langkah-langkah ini membantu memastikan bahwa hasil yang dihasilkan tidak hanya akurat secara statistik tetapi juga memiliki arti dan relevansi yang signifikan dalam konteks yang lebih luas.

**4. HASIL DAN PEMBAHASAN**

**4.1. Hasil Tujuan 1 Menganalisis dan mengevaluasi algoritma Naive Bayes dalam menentukan Penerimaan Kartu Indonesia Pintar (KIP)**

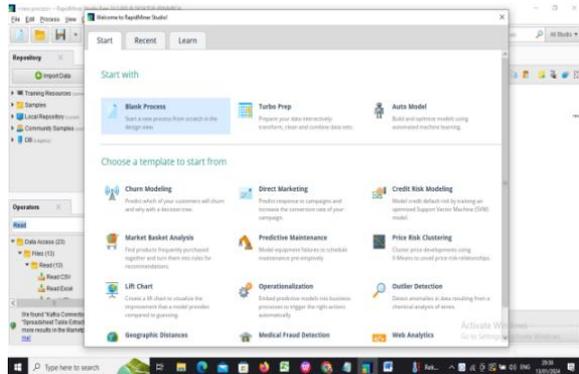
**4.1.1. Pengumpulan data**

No.	Alias	Prestasi	Nilai Ujian	Penghasilan Orangtua	.....	Target
1	_Rare_	0	65	2.1	.....	2
2	_Rare_	0	70	2.7	.....	2
.....	....	....	.....	.....	.....	.....
1000	_Rare_	1	78	2.2	.....	2

4.1.2. Data Selection

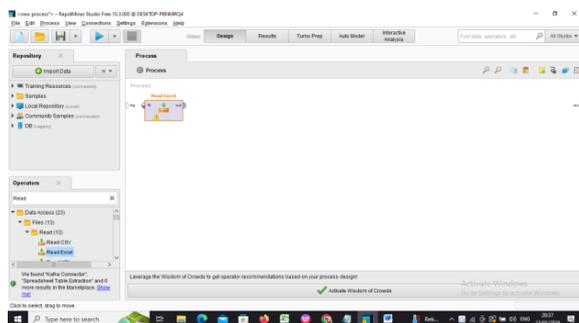
Analisis *Naïve bayes* menggunakan *RapidMiner*

- a. Langkah pertama yaitu klik aplikasi *RapidMiner*
- b. Membuat dataset dengan format *.xlsx* pada *Ms.Excel* dan klik *Blank Process*.



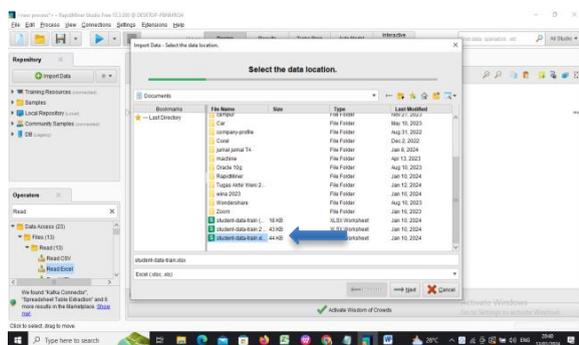
Gambar 1. Blank Process

- c. Menambahkan data, pilih *read excel* pada menu *operators*. Untuk menambahkan data yang akan diolah, maka perlu ditambahkan *read excel* yang ada pada menu *operators*. Seperti terlihat pada gambar berikut.



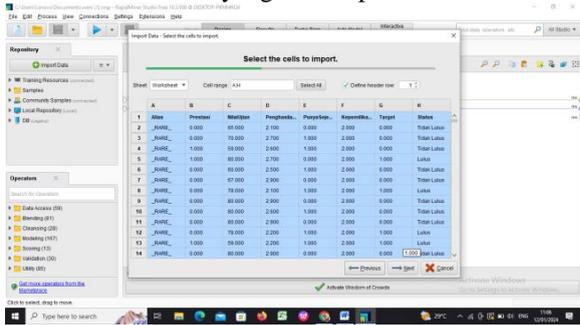
Gambar 1. Pemilihan Operators Read Excel

- d. Pilih data yang ingin ditambahkan, lalu klik *Next*.



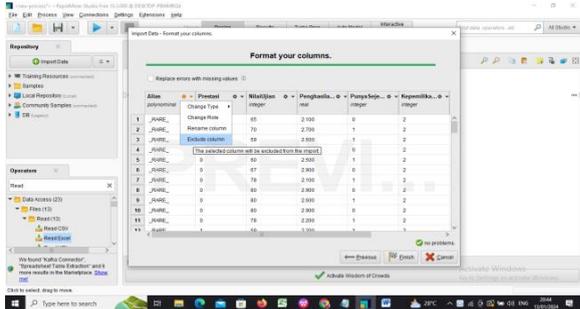
Gambar 2. Klik operator *read excel* kemudian klik pilih file *.xlsx*

- e. Pilihlah atribut yang akan diproses.

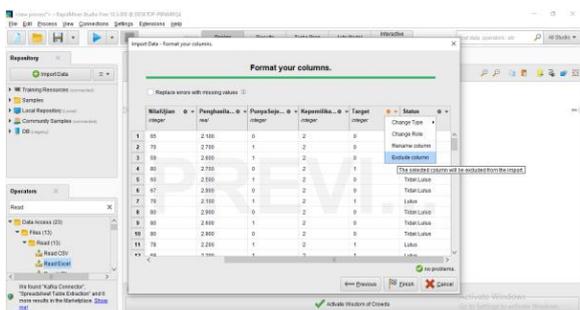


Gambar 3. Pemilihan atribut yang akan di proses

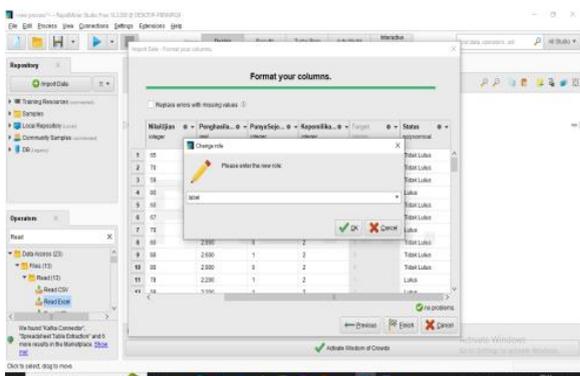
- f. Merubah format kolom



Gambar 4. Dibagian kolom alias klik exclude kolom.



Gambar 5. Dibagian kolom target pilih exclude kolom



Gambar 6. Dibagian kolom status klik *change role* kemudian pilih label

Pada gambar 7 di lakukan data selection atau penyeleksian data dimana attribute attribute yang tidak memiliki arti untuk menentukan penerima kip kuliah di hapus atau dihilangkan. Pada kasus diatas attribute

yang dihilangkan adalah alias dan target. Dimana atribut target diganti menjadi status.

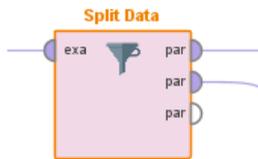
a. Hasil seleksi data

Row No.	Status	predictK.	confidence.	precision	Mislabel	Penghapusan	Penyisipan	Keperluan.	
1	Tidak Lulus	Tidak Lulus	0.837	0.163	1	0	2.000	1	2
2	Tidak Lulus	Tidak Lulus	1.000	0.000	0	67	0.000	0	2
3	Lulus	Lulus	0.104	0.896	0	78	2.100	1	2
4	Tidak Lulus	Tidak Lulus	0.715	0.285	0	79	2.200	0	2
5	Tidak Lulus	Lulus	0.270	0.730	0	80	2.200	1	2
6	Tidak Lulus	Tidak Lulus	0.686	0.314	0	80	2.000	0	2
7	Tidak Lulus	Lulus	0.301	0.699	0	58	2.200	1	2
8	Tidak Lulus	Tidak Lulus	0.987	0.013	1	60	0.000	0	2
9	Tidak Lulus	Tidak Lulus	0.911	0.309	0	79	2.000	1	2
10	Tidak Lulus	Tidak Lulus	1.000	0.000	0	79	2.000	0	2
11	Tidak Lulus	Tidak Lulus	0.689	0.311	0	78	2.000	0	2
12	Tidak Lulus	Lulus	0.301	0.699	0	58	2.200	1	2
13	Lulus	Tidak Lulus	0.957	0.043	0	80	2.700	1	2
14	Lulus	Tidak Lulus	0.702	0.298	0	69	2.200	0	2
15	Tidak Lulus	Tidak Lulus	0.870	0.130	0	70	2.000	0	2

Gambar 7. Hasil seleksi data

Pada gambar 8 menunjukkan bahwasannya hasil dari seleksi data sudah sesuai keinginan dan sudah sesuai apa yang dibutuhkan oleh penulis

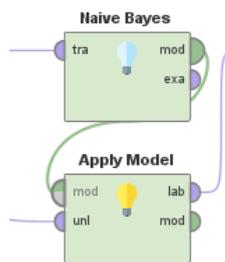
b. Penambahan operator split data



Gambar 8. Pembagian data

Dari read excel kemudian disambungkan dengan operator Split data. Split data merupakan bagian data di mana pada bagian ini data di split menjadi 2 yaitu 80%/20%

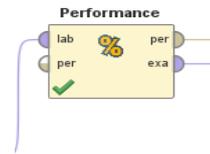
c. Penambahan operator Naïve Bayes dan Apply Model



Gambar 9. Proses klasifikasi data

Setelah data training siap, tahap selanjutnya adalah membangun model menggunakan algoritma *Naive Bayes*. Ini merupakan salah satu teknik klasifikasi data biner yang cukup sederhana dan efektif. Model *Naive Bayes* bekerja dengan menghitung probabilitas suatu data termasuk dalam kelas tertentu berdasarkan hubungan antara data input dan data target.

a. Penambahan operator Performance



Gambar 10. Hasil akurasi

Terakhir ketika performance di klik maka akan menampilkan hasil dari akurasi.

b. Pre-processing Data

Pada *Pre-processing* yaitu pemrosesan data apakah terdapat missing atau tidak. Akan tetapi pada gambar 4.9 menyatakan bahwa tidak ada missing data.

name	Type	Missing	Statistics
Status	Preprocessing	0	Tidak Lulus (175) / Tidak Lulus (175) / Tidak Lulus (175) / Tidak Lulus (175)
predictK(Status)	Preprocessing	0	Lulus (75) / Tidak Lulus (127) / Tidak Lulus (127)
confidence(Tidak Lulus)	Prep	0	0.008 / 1.000 / 0.456
confidence(Lulus)	Prep	0	0.000 / 0.962 / 0.144
Precision	Integer	0	0
Mislabel	Integer	0	58 / 80 / 69.875
PenghapusanOris	Prep	0	0.000 / 2.400 / 2.400
PenyisipanOris	Integer	0	0

Gambar 11. Pre-processing data statistics

a. Data Transformation

Gambar 12. Penambahan label

Row No.	Status	predictK.	confidence.	precision	Mislabel	Penghapusan	Penyisipan	Keperluan.	
1	Tidak Lulus	Tidak Lulus	0.837	0.163	1	0	2.000	1	2
2	Tidak Lulus	Tidak Lulus	1.000	0.000	0	67	0.000	0	2
3	Lulus	Lulus	0.104	0.896	0	78	2.100	1	2
4	Tidak Lulus	Tidak Lulus	0.715	0.285	0	79	2.200	0	2
5	Tidak Lulus	Lulus	0.270	0.730	0	80	2.200	1	2
6	Tidak Lulus	Tidak Lulus	0.686	0.314	0	80	2.000	0	2
7	Tidak Lulus	Lulus	0.301	0.699	0	58	2.200	1	2
8	Tidak Lulus	Tidak Lulus	0.987	0.013	1	60	0.000	0	2
9	Tidak Lulus	Tidak Lulus	0.911	0.309	0	79	2.000	1	2
10	Tidak Lulus	Tidak Lulus	1.000	0.000	0	79	2.000	0	2
11	Tidak Lulus	Tidak Lulus	0.689	0.311	0	78	2.000	0	2
12	Tidak Lulus	Lulus	0.301	0.699	0	58	2.200	1	2
13	Lulus	Tidak Lulus	0.957	0.043	0	80	2.700	1	2
14	Lulus	Tidak Lulus	0.702	0.298	0	69	2.200	0	2
15	Tidak Lulus	Tidak Lulus	0.870	0.130	0	70	2.000	0	2

Gambar 13. Atribut yang diberi label

Pada gambar 13 merupakan penentuan label pada dataset yang akan digunakan pada proses ini. Sedangkan pada gambar 14 merupakan atribut yang akan di beri label setelah label dipilih maka simpan dataset yang akan digunakan untuk pemrosesan.



lulus, (false lulus), dan (false tidak lulus). Dari hasil penyeleksian, terdapat 112 siswa dengan prediksi tidak lulus yang sesuai (true tidak lulus) dan 15 siswa dengan prediksi lulus yang sesuai (true lulus). Namun, terdapat 23 siswa dengan prediksi lulus yang ternyata tidak lulus (false lulus) dan 50 siswa dengan prediksi tidak lulus yang ternyata lulus (false tidak lulus). Algoritma Naïve Bayes mencapai tingkat akurasi sebesar 81.00%, yang mengindikasikan seberapa baik algoritma dapat memprediksi hasil yang benar. Class recall untuk kategori true tidak lulus mencapai angka tertinggi sebesar 82.96%, menunjukkan kemampuan algoritma untuk mendeteksi dengan baik siswa yang seharusnya tidak lulus. Sementara itu, class recall untuk kategori lulus sebesar 76.92%, menunjukkan tingkat keberhasilan yang sedikit lebih rendah dalam mengidentifikasi siswa yang seharusnya lulus. Dengan demikian, dapat disimpulkan bahwa algoritma Naïve Bayes cenderung lebih baik dalam memprediksi atau mengidentifikasi.

Berdasarkan hasil pengukuran tingkat akurasi dari model prediktif yang dikembangkan menggunakan algoritma *Naive Bayes* pada dataset 1000 siswa, mencapai tingkat akurasi sebesar 81.00%. Penelitian ini masih memiliki banyak kekurangan adapun saran yang dapat dikemukakan yaitu: untuk peneliti selanjutnya dalam mengklasifikasi Penentuan Penerima KIP sebaiknya menggunakan metode lain selain metode algoritma *Naive Bayes*.

#### DAFTAR PUSTAKA

- [1] Afdul Hadi, I. A. (2023). *MENENTUKAN KELAYAKAN PENERIMA KIP MENGGUNAKAN KLASIFIKASI DENGAN METODE ALGORITMA NAIVE BAYES*.
- [2] Cahyaningsih, R. I. (2018). *PENDISTRIBUSIAN KARTU INDONESIA PINTAR (KIP)*.
- [3] Hakam Febtadianrano Putro, R. T. (2020). *Penerapan Metode Naive Bayes Untuk Klasifikasi Pelanggan*.
- [4] Jefi, H. A. (2022). *Penerapan Algoritma Naive Bayes Dalam Memprediksi Penyakit*.
- [5] kurniawan, I. (2022). *Penentuan Penerima Bantuan Siswa Miskin dengan Metode Naive Bayes*.
- [6] Naomi Chatrina Siregar, R. R. (2020). *Implementasi Metode Naive Bayes Classifier (NBC) Pada Komentar*.
- [7] Nyoman Tri Anindia Putra, G. I. (2023). *Penerapan Algoritma Copeland Score Sebagai Penunjang Penerimaan Beasiswa KIPDiKampus Institut Bisnis dan Teknologi Indonesia*.
- [8] Rika Nofitri, N. I. (2019). *ANALISIS DATA HASIL KEUNTUNGAN MENGGUNAKAN SOFTWARE RAPIDMINER*.
- [9] Setiawan, R. (2021). <https://www.dicoding.com/blog/apa-itu-data-mining/>. *Apa itu data mining*.
- [10] Widiyanto, M. H. (2019). *Algoritma Naive Bayes*.