

Analisis Dan Penerapan Algoritma *Naïve Bayes* Untuk Klasifikasi Kelayakan Penerimaan Beasiswa PIP (Studi Kasus : SMPN 7 Kota Jambi)

Dimas Yudha Prawira¹, Yovi Pratama², Elvi Yanti³

Ilmu Komputer, Teknik Informatika, Universitas Dinamika Bangsa, Jambi, Indonesia
Email: ¹dimasprawira05@gmail.com, ²yovi.pratama@gmail.com, ³elvote92@gmail.com
Email Penulis Korespondensi: dimasprawira05@gmail.com

Artikel Info :
Artikel History :
Submitted : 19-07-2024
Accepted : 29-09-2024
Published : 30-09-2024

Kata Kunci:

Data mining; Klasifikasi;
Algoritma *Naïve Bayes*;
Rapid Miner; Penerimaan
Beasiswa PIP ;

Abstrak– Beasiswa adalah bantuan keuangan yang diberikan kepada masyarakat dengan tujuan mendukung pendidikan. Program beasiswa didirikan untuk membantu siswa mengatasi masalah keuangan dalam menyelesaikan pendidikan mereka. Beasiswa diberikan secara selektif sesuai dengan kebutuhan individu. Untuk mengetahui akurasi dari hasil prediksi kelayakan penerimaan beasiswa Program Indonesia Pintar (PIP), teknik klasifikasi dapat digunakan sebagai bagian dari data mining. Pengujian dilakukan dengan menggunakan *all attributes* dan atribut terbaik pada *3 options test*, antara lain *Use Training Set*, *5 Fold Cross Validation*, *10 Fold Cross Validation*. Pengujian data mengenai penerimaan beasiswa PIP memiliki akurasi yang tinggi pada *options test Use Training Set* (atribut terbaik) yaitu 93,18% dibandingkan dengan pengujian lainnya. Sedangkan untuk akurasi terendah adalah *5 Fold Cross Validation 10%* (2021 – 2022) yang memperoleh akurasi dengan hasil 81,82%. Algoritma *naïve bayes* dapat dikatakan sebagai salah satu algoritma yang efektif baik dari perhitungan ataupun hasil akhir yang mana pengujian tersebut dapat dijadikan sebagai landasan terkait penerimaan beasiswa.

Abstract– Scholarships are financial aid provided to the community with the aim of supporting education. Scholarship programs were established to help students overcome financial problems in completing their education. Scholarships are awarded selectively according to individual needs. To determine the accuracy of the predicted results of eligibility for the Indonesia Pintar Scholarship Program (PIP), classification techniques can be used as part of data mining. Testing is done by using all attributes and the best attributes on 3 options test, among others *Use Training Set*, *5 Fold Cross Validation*, *10 Fold Cross Validation*. Data on the acceptance of PIP scholarships has a high accuracy in the options test *Use Training Set* (best attribute) which is 93.18% compared to other tests. As for the lowest accuracy is *5 Fold Cross Validation 10%* (2021 – 2022) which obtained an accuracy of 81.82%. *Naïve Bayes* algorithm can be said to be one of the effective algorithms both from the calculation and the final result where the test can be used as a foundation related to scholarship acceptance.

Keywords:

Data Mining;
Classification; *Naïve Bayes* Algorithm; Rapid Miner; PIP scholarship recipients; population characteristics;

1. PENDAHULUAN

Pendidikan dianggap sebagai faktor penting dalam menciptakan sumber daya manusia yang berkualitas melalui sistem pendidikan yang baik. Hal ini akan membantu generasi muda bangsa untuk memiliki motivasi belajar dan mampu bersaing dalam mengembangkan potensi mereka. Untuk mendukung hal tersebut, pemberian beasiswa di sekolah dapat menjadi salah satu cara untuk memfasilitasi akses pendidikan yang lebih baik. Program beasiswa dirancang untuk membantu siswa dalam menanggung biaya pendidikan mereka. Biasanya, pemberian beasiswa bersifat selektif dan diberikan sesuai dengan kebutuhan penerima beasiswa [1].

Beasiswa Program Indonesia Pintar (PIP) adalah program beasiswa yang ditujukan untuk membantu biaya pendidikan bagi siswa-siswa yang kurang mampu di Indonesia. Namun, terdapat ketimpangan yang terjadi dalam penetapan beasiswa PIP sekolah seperti, informasi tentang program beasiswa PIP seringkali hanya disebarkan di sekolah-sekolah tertentu saja, sehingga siswa dari sekolah lain yang juga membutuhkan informasi tersebut kurang mendapatkan kesempatan yang sama dan salah satu ketimpangan yang terjadi dalam penetapan beasiswa PIP sekolah adalah proses seleksi yang tidak transparan. Solusi yang dapat diambil untuk mengatasi masalah beasiswa PIP sekolah adalah dengan meningkatkan akses informasi mengenai program beasiswa PIP, baik melalui media sosial, brosur, atau siaran pers. Algoritma *naïve bayes* dapat membantu dalam mengklasifikasi kelayakan beasiswa terhadap data – data siswa agar dapat mengetahui mana siswa yang memiliki peluang besar untuk mendapatkan beasiswa dengan tepat. Melakukan transparansi dan objektivitas, pemerintah harus mengeluarkan panduan atau petunjuk yang jelas tentang kriteria seleksi dan memastikan bahwa seleksi dilakukan secara adil tanpa diskriminasi.

Data Mining adalah sebuah proses untuk menggali informasi yang berguna dari suatu kumpulan data besar dengan menerapkan teknik-teknik analisis [2]. *Data mining* membantu dalam menemukan wawasan baru dari data yang sebelumnya tidak diketahui atau tidak terlihat [3]. Dalam *data mining*, diperlukan juga klasifikasi yang digunakan untuk mengidentifikasi pola atau aturan di dalam data dan membuat prediksi tentang kelas atau kategori dari data yang baru [4].

Metode *Naive Bayes* adalah teknik klasifikasi yang menggunakan pendekatan probabilitas untuk memprediksi kategori dari sebuah data. Pendekatan ini menggabungkan probabilitas setiap kata dengan probabilitas kategori yang ada, sehingga dapat menentukan kemungkinan kategori yang tepat untuk data tersebut [5]. Metode klasifikasi data yang dikenal sebagai *Naive Bayes* dapat diandalkan untuk menghasilkan hasil yang akurat. Algoritma ini terkenal karena keakuratannya yang tinggi [6]. Teknik prediksi ini didasarkan pada penerapan teorema Bayes, juga dikenal sebagai aturan Bayes, yang mengasumsikan bahwa semua variabel saling independen [7]. *Naive Bayes* juga dikenal sebagai perhitungan teorema Bayes yang paling sederhana [8], pada metode *Naive Bayes*, terdapat penekanan pada konsep probabilitas hipotesis dan bukti-bukti (*evidence*) [9]. Metode *Naive Bayes* efektif dan efisien dalam menentukan calon penerima beasiswa [10]. Jadi dapat disimpulkan bahwa metode ini merupakan salah satu algoritma klasifikasi untuk memprediksi probabilitas suatu class atau peluang dari data yang sudah ada. Alat bantu yang digunakan yaitu aplikasi *rapid miner* [11]. *Rapid miner* digunakan untuk berbagai tujuan seperti bisnis, penelitian, pendidikan, pelatihan, pembuatan prototipe, dan lain-lain [12].

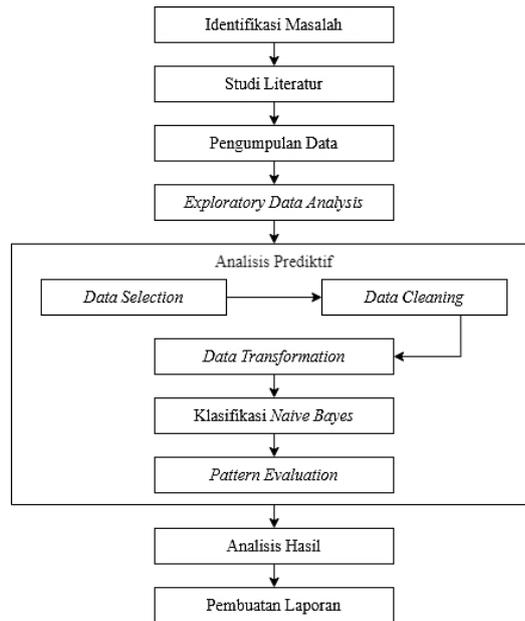
Adapun penelitian sejenis terkait analisis penerimaan beasiswa seperti penelitian yang dilakukan oleh Agus Salim mengenai klasifikasi kelayakan penerima beasiswa dengan menggunakan metode *naive bayes*, diklasifikasikan dalam layak atau tidak layak pada pemberian beasiswa, Klasifikasi diperoleh dari atribut kelayakan seperti alat transportasi, penghasilan, gaji orangtua, dan lainnya. Didapatkan akurasi sebesar 98,01% [13]. Pada penelitian yang dilakukan oleh Alief Ulfa, Agus Setia, dan Purnomo Hadi Susilo dengan menerapkan algoritma *naive bayes*, Tujuan dari program beasiswa adalah untuk memberikan bantuan keuangan kepada mahasiswa yang memenuhi syarat dan kriteria tertentu, seperti prestasi akademik dan keterbatasan finansial. Didapatkan akurasi sebesar 92,7% [14]. Pada penelitian yang dilakukan oleh Debby Alita, Indah Sari, Auliya Rahman, dan Styawati, metode *Naive Bayes* digunakan sebagai bagian dari sistem pendukung keputusan dalam proses seleksi penerima beasiswa., sistem ini bertujuan untuk membantu sekolah dalam menentukan penerima beasiswa secara cepat dan akurat. Didapatkan akurasi sebesar 85,96% [15]. Pada penelitian yang dilakukan oleh Haditsah Annur tentang klasifikasi masyarakat miskin, Berdasarkan uji coba dengan teknik split validasi pada dataset objek penelitian, metode klasifikasi *naive bayes* menghasilkan tingkat akurasi sebesar 73%, yang termasuk dalam kategori *Good*. Selain itu, nilai *Precision* mencapai 92% dan *Recall* mencapai 86%. Hasil ini menunjukkan bahwa sistem klasifikasi yang dikembangkan dapat dijadikan sebagai masukan bagi pengambil keputusan [16]. Pada penelitian yang dilakukan oleh Wiwit Pura, Dinda Ayu Lara Saky, Muhammad Malthuf, Muhammad Gazali, dan Ristu Haiban. Dalam penelitian ini membahas tentang mengklasifikasikan masyarakat miskin dengan metode *naive bayes*, dapat dilihat dari nilai akurasi sebesar 96,63%. Berdasarkan hasil penelitian, dapat disimpulkan bahwa masyarakat desa Lepak tergolong dalam kategori masyarakat miskin. [17].

Berdasarkan penjabaran diatas, maka penulis tertarik untuk melakukan pengklasifikasian penerimaan beasiswa yang Penulis tuangkan dalam judul “**Analisis Dan Penerapan Algoritma Naive Bayes Untuk Klasifikasi Kelayakan Penerimaan Beasiswa PIP (Studi Kasus: SMPN 7 Kota Jambi)**”, Proses seleksi penerima beasiswa memerlukan penelitian dan pengumpulan data untuk menentukan kriteria yang harus dipenuhi oleh calon penerima beasiswa. Hal ini dilakukan untuk memastikan bahwa bantuan dana atau beasiswa diberikan kepada orang yang memenuhi syarat yang telah ditetapkan.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Pada penelitian dibutuhkan kerangka kerja penelitian yang berisi langkah-langkah terstruktur agar penelitian dapat berjalan dengan baik. Adapun kerangka kerja penelitian dapat dilihat pada gambar 1.



Gambar 1. Kerangka Kerja Penelitian

Berdasarkan Gambar 1, berikut adalah uraian dari langkah-langkah penelitian yang akan dilakukan :

2.1 Identifikasi Masalah

Pada tahap awal, peneliti akan melakukan identifikasi masalah sebagai bagian dari proses penelitian yaitu pemberian program beasiswa masih belum optimal dan belum tepat sasaran dan mengenai kelayakan siswa penerima program beasiswa. Sehingga peneliti dapat memilih rencana kerja serta menentukan data yang dibutuhkan dalam penelitian ini.

2.2 Studi Literatur

Pada tahapan ini peneliti melakukan studi literatur yang meliputi buku referensi, artikel, dan hasil penelitian terkait permasalahan yang akan diteliti. Tujuan dari studi literatur ini adalah untuk mendapatkan landasan teoritis tentang teknik *data mining*, proses klasifikasi menggunakan algoritma *naive bayes*, *tool data mining* yang akan digunakan.

2.3 Pengumpulan Data

Pada tahap pengumpulan data, peneliti akan mengumpulkan informasi yang dibutuhkan untuk penelitian, seperti data siswa SMPN 7 Kota Jambi dan data penerima Beasiswa.

a. Pengamatan Langsung (*Observation*)

Peneliti melakukan pengamatan secara langsung kegiatan-kegiatan yang sedang berlangsung pada SMPN 7 Kota Jambi, sehingga peneliti dapat mengumpulkan data - data yang diperlukan dalam penelitian ini.

b. Wawancara (*Interview*)

Peneliti melakukan wawancara dengan menanyakan secara langsung kepada pihak-pihak yang berkepentingan atau pihak-pihak bersangkutan, misalnya Bapak Kepala Sekolah untuk memperoleh data-data yang diperlukan.

2.4 EDA (*Exploratory Data Analysis*)

Pada tahap ini peneliti melakukan *Exploratory Data Analysis* (EDA) yang merujuk dalam proses investigasi awal data, peneliti melakukan analisis kritis untuk menemukan pola, anomali, menguji hipotesis, dan mengecek asumsi. Analisis ini menggunakan teknik statistik deskriptif dan representasi grafis. Melalui analisis data eksploratif, kita dapat mendapatkan pemahaman yang lebih baik mengenai karakteristik dataset yang digunakan..

2.5 Analisis Prediktif

Analisis prediktif merupakan sebuah proses yang melibatkan penggunaan beberapa variabel untuk mendapatkan pola dari data dan memprediksi hasil di masa depan. Teknik klasifikasi sering digunakan dalam analisis ini. Sebelumnya, pada tahap persiapan data, dilakukan pengumpulan dan penyiapan data siswa SMPN 7 Kota Jambi.

Adapun alur proses dari analisis prediktif adalah sebagai berikut :

a. *Data Selection*

Sebelum melakukan tahap penggalian informasi dalam *data mining*, seleksi data dari sekumpulan data operasional perlu dilakukan. Data yang relevan untuk proses data mining dipilih dan disimpan dalam berkas terpisah dari basis data operasional. Pada tahap ini, peneliti melakukan seleksi data mengenai siswa SMPN 7 Kota Jambi.

b. *Data Cleaning*

Cleaning data merupakan *preprocessing* dimana pada tahap ini dilakukan pembersihan data dari *missing value* dan data yang tidak konsisten seperti Siswa SMPN 7 Kota Jambi yang memiliki atribut tidak lengkap.

c. *Data Transformation*

Data Transformation adalah proses transformasi pada data Siswa SMPN 7 Kota Jambi yang telah selesai, sehingga data telah sesuai untuk proses *data mining* dalam klasifikasi kelayakan siswa penerima beasiswa.

d. Klasifikasi *Naive Bayes*

Pada tahap ini dilakukan proses klasifikasi kelayakan siswa penerima beasiswa pada data Siswa SMPN 7 Kota Jambi menggunakan algoritma *naive bayes*. Langkah-Langkah algoritma klasifikasi *naive bayes* untuk kelayakan siswa penerima beasiswa adalah sebagai berikut:

1. Menghitung jumlah kelas atau label dari data siswa.
2. Menghitung jumlah kasus per kelas yang sama dari data siswa.
3. Kalikan semua variabel kelas status kelayakan siswa penerima beasiswa.
4. Bandingkan hasil per kelas status kelayakan siswa penerima beasiswa.

Dalam melakukan proses klasifikasi data menggunakan *naive bayes*, peneliti menggunakan alat bantu (*tool*) yaitu *Rapid Miner* terhadap data Siswa yang sudah di *testing* menggunakan model *classifier* yang sudah dibentuk.

e. *Pattern Evaluation*

Pattern evaluation bertujuan untuk mengidentifikasi pola dalam *knowledge base* yang ditemukan. Dalam tahap ini, pola khas dan model prediksi dievaluasi untuk menilai apakah hipotesis yang diusulkan tercapai. Pada tahapan ini data Siswa SMPN 7 Kota Jambi yang telah diolah menggunakan metode klasifikasi *naive bayes* akan menghasilkan informasi berupa kelayakan siswa penerima beasiswa.

2.6 Analisis Hasil

Pada tahap ini peneliti menganalisis data Siswa SMPN 7 Kota Jambi untuk dinilai keakuratan data model yang sudah didapatkan dengan perhitungan klasifikasi *naive bayes*. Hasil yang didapatkan berupa kelas “Layak” atau “Tidak Layak” seorang siswa menerima beasiswa.

2.7 Pembuatan Laporan

Pada tahap pembuatan laporan akhir penelitian dilakukan berdasarkan kerangka yang telah dirancang dimulai dari tahap identifikasi masalah, *studi literatur* terkait masalah yang diteliti, pengumpulan data, *exploratory data analysis*, analisis prediktif yang terdiri dari data selection, data processing, data transformation, klasifikasi *naive bayes* untuk kelayakan siswa penerima beasiswa, dan *pattern evaluation*, Kemudian dilakukan analisis hasil keakuratan data dengan perhitungan klasifikasi *naive bayes*. Pembuatan laporan bertujuan sebagai dokumentasi dari penelitian yang telah dilakukan.

3. HASIL DAN PEMBAHASAN

3.1 *Pre-Processing*

Penelitian ini terdapat 2081 data peserta didik SMPN 7 Kota Jambi tahun 2021 – 2022 dan 12 atribut. Berikut ini merupakan profil data penelitian pada Tabel 1.

Tabel 1. Profil Data Penelitian

Kelas	Jumlah
Layak	549
Tidak Layak	1532
Total	2081

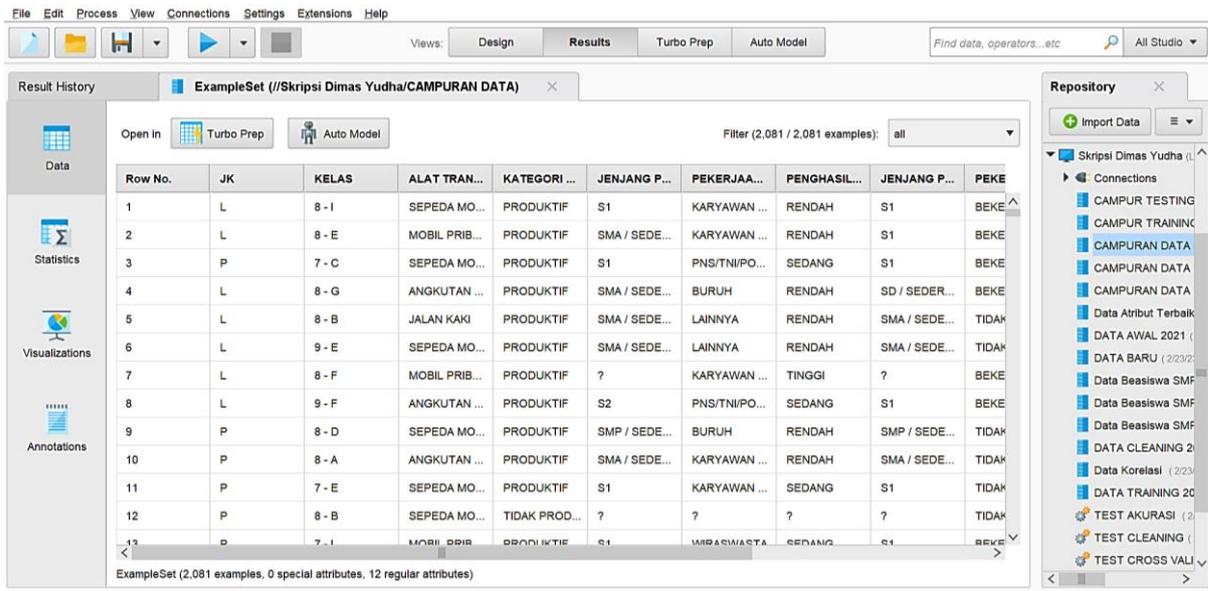
Berdasarkan hasil observasi yang telah dilakukan, penulis mendapatkan 12 atribut data yang akan di analisis dan dilakukan perhitungan diantaranya terdapat atribut adalah jenis kelamin, kelas, alat transportasi, kategori usia ayah, jenjang Pendidikan ayah, pekerjaan ayah, penghasilan ayah, kategori usia ibu, jenjang Pendidikan ibu, pekerjaan ibu, penghasilan ibu, tanggungan orang tua, dan label layak PIP (usulan dari sekolah). Berikut merupakan data peserta didik SMPN 7 Kota Jambi tahun 2021 – 2022 yang akan di input ke dalam *tool* atau alat bantu *rapid miner* pada Tabel 2.

Tabel 2. Data Peserta Didik SMPN 7 Kota Jambi

NO	JK	KELAS	ALAT TRANSPORTASI	KATEGORI USIA AYAH	JENJANG PENDIDIKAN AYAH	PEKERJAN AYAH	PENGHASILAN AYAH	JENJANG PENDIDIKAN IBU	PEKERJAN IBU	PENGHASILAN IBU	TANGGUNG JAWAB ORANG TUA	LAYAK PIP
1	L	8 - I	SEPEDA MOTOR	PRODUKTIF	S1	KARYAWAN	RENDAH	S1	BEKERJA	RENDAH	SEDIKIT	TIDAK LAYAK
2	L	8 - E	MOBIL PRIBADI	PRODUKTIF	SMA / SEC	KARYAWAN	RENDAH	S1	BEKERJA	SEDANG	SEDIKIT	TIDAK LAYAK
3	P	7 - C	SEPEDA MOTOR	PRODUKTIF	S1	PNS/TNI/F	SEDANG	S1	BEKERJA	SEDANG	BANYAK	TIDAK LAYAK
4	L	8 - G	ANGKUTAN UMI	PRODUKTIF	SMA / SEC	BURUH	RENDAH	SD / SEDE	BEKERJA	RENDAH	SEDIKIT	LAYAK
...
2078	L	9 - K	MOBIL PRIBADI	PRODUKTIF	SMA / SEC	KARYAWAN	RENDAH	S1	PNS/TNI/F	SEDANG	SEDIKIT	TIDAK LAYAK
2079	P	9 - K	JALAN KAKI	PRODUKTIF	SMA / SEC	KARYAWAN	SEDANG	SMA / SEC	TIDAK BEK	RENDAH	SEDIKIT	TIDAK LAYAK
2080	P	9 - K	MOBIL PRIBADI	PRODUKTIF	SMA / SEC	WIRASWA	SEDANG	S1	PNS/TNI/F	SEDANG	SEDIKIT	TIDAK LAYAK
2081	L	9 - K	OJEK	PRODUKTIF	S1	PNS/TNI/F	SEDANG	SMA / SEC	TIDAK BEK	RENDAH	SEDIKIT	TIDAK LAYAK

3.2 Features Selection

Setelah itu penulis melakukan *features selection* yakni merupakan proses dari penyeleksian pada data dengan memfokuskan pada atribut yang akan digunakan dalam perhitungan yaitu jenis kelamin, kelas, alat transportasi, kategori usia ayah, jenjang Pendidikan ayah, pekerjaan ayah, penghasilan ayah, jenjang Pendidikan ibu, pekerjaan ibu, penghasilan ibu, tanggungan orang tua, dan label layak PIP dengan menggunakan *tool rapid miner* pada gambar 2.



Gambar 2. Features Selection

3.3 Data Cleaning dan Data Transformation

Setelah melakukan penyeleksian data dengan memfokuskan atribut yang ada pada gambar 2, selanjutnya adalah melakukan tahap *data cleaning*. Pembersihan data atau *data cleaning* digunakan untuk menghilangkan *noise* yang ada pada data dan memastikan bahwa data yang digunakan konsisten dan dapat digunakan dalam analisis. Proses ini melibatkan penghapusan data yang tidak konsisten atau tidak diperlukan menggunakan algoritma *naive bayes*, setelah itu dilakukan tahap transformasi data yaitu proses perubahan data ke dalam format yang sesuai untuk diproses, data diubah dengan *format comma separated files (CSV)* agar dapat mempermudah untuk di aplikasikan ke dalam *rapid miner*.

Tabel 3. Data Cleaning dan Data Transformation

JK	KELAS	ALAT TRANSPORTASI	KATEGORI USIA AYAH	JENJANG PENDIDIKAN AYAH	PEKERJA AN AYAH	PENGHASILAN AYAH	JENJANG PENDIDIKAN IBU	PEKERJA AN IBU	PENGHASILAN IBU	TANGGUNG JAWAB ORANG TUA	LAYAK PIP
L	8-I	SEPEDA MOTOR	PRODUKTIF	S1	KARYAWAN	RENDAH	S1	BEKERJA	RENDAH	SEDIKIT	TIDAK LAYAK
L	8-E	MOBIL PRIBADI	PRODUKTIF	SMA / SEDERAJA	KARYAWAN	RENDAH	S1	BEKERJA	SEDANG	SEDIKIT	TIDAK LAYAK
P	7-C	SEPEDA MOTOR	PRODUKTIF	S1	PNS/TNI	SEDANG	S1	BEKERJA	SEDANG	BANYAK	TIDAK LAYAK
L	8-G	ANGKUTAN UML	PRODUKTIF	SMA / SEDERAJA	BURUH	RENDAH	SD / SEDEI	BEKERJA	RENDAH	SEDIKIT	LAYAK
L	8-B	JALAN KAKI	PRODUKTIF	SMA / SEDERAJA	LAINNYA	RENDAH	SMA / SEC	TIDAK BE	RENDAH	SEDIKIT	LAYAK
L	9-E	SEPEDA MOTOR	PRODUKTIF	SMA / SEDERAJA	LAINNYA	RENDAH	SMA / SEC	TIDAK BE	RENDAH	SEDIKIT	LAYAK
L	9-F	ANGKUTAN UML	PRODUKTIF	S2	PNS/TNI	SEDANG	S1	BEKERJA	RENDAH	SEDIKIT	TIDAK LAYAK
P	8-D	SEPEDA MOTOR	PRODUKTIF	SMP / SEDERAJA	BURUH	RENDAH	SMP / SED	TIDAK BE	RENDAH	SEDIKIT	LAYAK
P	8-A	ANGKUTAN UML	PRODUKTIF	SMA / SEDERAJA	KARYAWAN	RENDAH	SMA / SEC	TIDAK BE	RENDAH	BANYAK	LAYAK
P	7-E	SEPEDA MOTOR	PRODUKTIF	S1	KARYAWAN	SEDANG	S1	TIDAK BE	RENDAH	SEDIKIT	TIDAK LAYAK
P	7-I	MOBIL PRIBADI	PRODUKTIF	S1	WIRASWAN	SEDANG	S1	BEKERJA	RENDAH	SEDIKIT	TIDAK LAYAK
P	7-F	SEPEDA MOTOR	PRODUKTIF	S1	KARYAWAN	SEDANG	D3	TIDAK BE	RENDAH	SEDIKIT	TIDAK LAYAK
P	9-K	MOBIL PRIBADI	PRODUKTIF	S1	WIRASWAN	RENDAH	S1	BEKERJA	SEDANG	SEDIKIT	TIDAK LAYAK
P	7-G	MOBIL PRIBADI	PRODUKTIF	S1	PNS/TNI	SEDANG	SMA / SEC	TIDAK BE	RENDAH	SEDIKIT	TIDAK LAYAK
P	8-J	MOBIL PRIBADI	PRODUKTIF	S1	PNS/TNI	SEDANG	S1	BEKERJA	SEDANG	SEDIKIT	TIDAK LAYAK
L	7-A	JALAN KAKI	TIDAK PRODUKTIF	SMA / SEDERAJA	KARYAWAN	RENDAH	D3	TIDAK BE	RENDAH	SEDIKIT	LAYAK
L	9-A	JALAN KAKI	PRODUKTIF	SMA / SEDERAJA	WIRASWAN	RENDAH	D3	TIDAK BE	RENDAH	BANYAK	LAYAK
P	7-I	SEPEDA MOTOR	PRODUKTIF	SMA / SEDERAJA	WIRASWAN	RENDAH	SMA / SEC	TIDAK BE	RENDAH	BANYAK	LAYAK
P	9-C	SEPEDA MOTOR	PRODUKTIF	S2	PNS/TNI	SEDANG	SMA / SEC	TIDAK BE	RENDAH	SEDIKIT	TIDAK LAYAK

Pada proses ini keseluruhan data siswa yang kosong berjumlah 208 data dan akan di *cleaning*, sehingga menyisakan 1873 data dari 2081 data. Setelah itu dilakukan pengujian menggunakan algoritma *naïve bayes* yang akan di terapkan pada penelitian ini dengan menggunakan *rapid miner*, sehingga data tersebut memiliki potensi dalam menentukan “layak” dan “tidak layak” bagi penerima beasiswa.

3.3 Correlation Analysis

Tabel 4. Atribut Correlation Analysis

Nama atribut	No	Bobot
Alat Transportasi	1	0.257
Jenjang Pendidikan Ayah	2	0.182
Pekerjaan Ayah	3	0.229
Penghasilan Ayah	4	-0.599
Pekerjaan Ibu	5	0.323
Penghasilan Ibu	6	-0.343

Beberapa dari atribut memiliki korelasi terbaik yang paling berpengaruh terhadap kelasnya, sehingga pada tabel 4 di atas pula dapat disimpulkan bahwa atribut yang mendekati nilai -1 dan 1 adalah atribut yang mempengaruhi hasil dari perhitungan penelitian.

3.4 Hasil Analisis dan Evaluasi Akurasi Naïve Bayes

Tabel 5. Perbandingan Evaluasi Akurasi Rapid Miner

Model Evaluasi	Akurasi	Jumlah Kelas Data	Presentasi Data
Use Training Set 2022 (All Attributes)	Correctly Classified Instances	599	92,87%
	Incorrectly Classified Instances	46	7,13%
5 Fold Cross Validation 2022 (All Attributes)	Correctly Classified Instances	598	92,71%
	Incorrectly Classified Instances	47	7,29%
10 Fold Cross Validation 2022 (All Attributes)	Correctly Classified Instances	597	92,56%

	<i>Incorrectly Classified Instances</i>	48	7,44%
	<i>Correctly Classified Instances</i>	601	93,18%
Use Training Set 2022 (Atribut Terbaik)	<i>Incorrectly Classified Instances</i>	44	6,82%
	<i>Correctly Classified Instances</i>	599	92,87%
5 Fold Cross Validation 2022 (Atribut Terbaik)	<i>Incorrectly Classified Instances</i>	46	7,13%
	<i>Correctly Classified Instances</i>	597	92,56%
10 Fold Cross Validation 2022 (Atribut Terbaik)	<i>Incorrectly Classified Instances</i>	48	7,44%
	<i>Correctly Classified Instances</i>	1480	87,83%
Use Training Set 90% (2021 – 2022)	<i>Incorrectly Classified Instances</i>	205	12,17%
	<i>Correctly Classified Instances</i>	1473	87,42%
5 Fold Cross Validation 90% (2021 – 2022)	<i>Incorrectly Classified Instances</i>	212	12,58%
	<i>Correctly Classified Instances</i>	1473	87,42%
10 Fold Cross Validation 90% (2021 – 2022)	<i>Incorrectly Classified Instances</i>	212	12,58%
	<i>Correctly Classified Instances</i>	173	92,51%
Use Testing Set 10% (2021 – 2022)	<i>Incorrectly Classified Instances</i>	14	7,49%
	<i>Correctly Classified Instances</i>	153	81,82%
5 Fold Cross Validation 10% (2021 – 2022)	<i>Incorrectly Classified Instances</i>	34	18,18%
	<i>Correctly Classified Instances</i>	155	82,89%
10 Fold Cross Validation 10% (2021 – 2022)	<i>Incorrectly Classified Instances</i>	32	17,11%

Setelah dilakukan analisis klasifikasi *naïve bayes* pada *tool rapid miner*, dari pengujian antara all attributes tahun 2022, atribut dengan korelasi terbaik tahun 2022, dan *dataset* 2021 – 2022, dapat di simpulkan bahwa, atribut dengan korelasi terbaik tingkat persentasinya lebih tinggi dibandingkan all attributes tahun 2022 dan dataset 2021 – 2022 yaitu berjumlah 93,18%.

3.5 Peforma Klasifikasi Menggunakan *Rapid Miner*

Tabel 6. Hasil Uji Algoritma *Naïve Bayes*

Model Evaluasi	Akurasi	Precision	Recall	F-Measure
<i>Use Training Set 2022 (All Attributes)</i>	92,87%	96,67%	93,36%	94,99%
<i>5 Fold Cross Validation 2022 (All Attributes)</i>	92,71%	96,26%	93,58%	94,90%
<i>10 Fold Cross Validation 2022 (All Attributes)</i>	92,56%	96,04%	93,58%	94,79%
<i>Use Training Set 2022 (Atribut Terbaik)</i>	93,18%	96,26%	94,19%	95,21%
<i>5 Fold Cross Validation 2022 (Atribut Terbaik)</i>	92,87%	95,84%	94,19%	95,01%
<i>10 Fold Cross Validation 2022 (Atribut Terbaik)</i>	92,56%	95,42%	94,19%	94,83%
<i>Use Training Set 90% (2021 – 2022)</i>	87,83%	94,85%	88,40%	91,51%
<i>5 Fold Cross Validation 90% (2021 – 2022)</i>	87,42%	94,59%	88,08%	91,22%
<i>10 Fold Cross Validation 90% (2021 – 2022)</i>	87,42%	94,51%	88,16%	91,22%
<i>Use Testing Set 10% (2021 – 2022)</i>	92,51%	82,35%	89,36%	85,71%
<i>5 Fold Cross Validation 10% (2021 – 2022)</i>	81,82%	65,85%	57,45%	61,36%
<i>10 Fold Cross Validation 10% (2021 – 2022)</i>	82,89%	67,44%	61,70%	64,44%

Dari pengujian menggunakan *rapid miner* pada data peserta didik 2021 – 2022 didapatkan hasil layak PIP sebesar 473 data dan tidak layak sebesar 1400 data dari jumlah keseluruhan 1873 data. Kelas 7 – 9 yang mendapatkan beasiswa PIP terbanyak tahun 2021 terdapat pada kelas 7 sejumlah 77 orang dengan penerima terbanyak pada kelas 7C sejumlah 11 orang, kelas 8 sejumlah 66 orang dengan penerima terbanyak pada kelas 8G sebesar 12 orang, dan kelas 9 sejumlah 70 orang dengan penerima terbanyak pada kelas 9A sebesar 11 orang. Sedangkan tahun 2022 pada kelas kelas 7 sejumlah 116 orang dengan penerima terbanyak 7I sebesar 12 orang, kelas 8 sejumlah 112 orang dengan penerima terbanyak 8D sebesar 12 orang, dan kelas 9 sejumlah 94 orang dengan penerima terbanyak 9C sebesar 11 orang. Maka diperoleh akurasi tertinggi yaitu dengan menggunakan *Use Training Set (Atribut Terbaik)* dengan persentasi akurasi yaitu 93,18% untuk *Correctly Classified Instances* dan 7,44% untuk *Incorrectly Classified Instances* pada *10 Fold Cross Validation (All Attributes)* dan *10 Fold Cross Validation (Atribut terbaik)*.

4. KESIMPULAN

Penelitian ini menggunakan data peserta didik dari SMP N 7 Kota Jambi pada tahun 2021 – 2022 berjumlah 2081 data siswa dengan atribut jenis kelamin, kelas, alat transportasi, kategori usia ayah, jenjang Pendidikan ayah, pekerjaan ayah, penghasilan ayah, jenjang Pendidikan ibu, pekerjaan ibu, penghasilan ibu, tanggungan orang tua, dan layak PIP. Setelah di *cleaning* menjadi 1873 data, Perbandingan 3 *options test* pada proses pengujian *all attributes* dan atribut terbaik menggunakan *rapid miner* menghasilkan persentase hasil akurasi dari klasifikasi *Naïve Bayes*, yaitu *use training set*, *5 fold cross validation*, dan *10 fold cross validation*. Hasil uji klasifikasi *Naïve*

Bayes dengan akurasi terbesar yang diperoleh pada *all attributes* 2022 adalah *Use Training Set* sebesar 92,87%, dan pada atribut terbaik adalah *Use Training Set* sebesar 93,18%, dapat dilihat akurasi terbesar ada pada atribut terbaik yang berarti dengan menggunakan metode *Correlation Analysis* kita bisa mengetahui atribut mana yang memiliki korelasi terbaik.

REFERENCES

- [1] R. Ramadhon, R. Jaenudin, and S. Fatimah, "PENGARUH BEASISWA TERHADAP MOTIVASI BELAJAR MAHASISWA PENDIDIKAN EKONOMI UNIVERSITAS SRIWIJAYA," *Jurnal Profit*, vol. 4, no. 2, pp. 203–213, May 2018.
- [2] O. Soleh, O. Devega, P. L. Asoti, and K. Olivia, "ANALISA KEMUNGKINAN CALON MAHASISWA BARU MEMILIH PRODI TEKNIK INFORMATIKA MELALUI PENERAPAN DATA MINING DECISION TREE," 2014.
- [3] F. Firdaus and A. Mukhlis, "Implementasi Algoritma Naive Bayes Pada Data Set Kualitatif Prediksi Kebangkrutan," *JURIKOM (Jurnal Riset Komputer)*, vol. 7, no. 1, p. 15, Feb. 2020, doi: 10.30865/jurikom.v7i1.1757.
- [4] D. Firdaus, "Penggunaan Data Mining dalam Kegiatan Sistem Pembelajaran Berbantuan Komputer," *Jurnal*, vol. 6, pp. 2089–5615, 2017.
- [5] D. Dahri *et al.*, "METODE NAIVE BAYES UNTUK PENENTUAN PENERIMA BEASISWA BIDIKMISI UNIVERSITAS MULAWARMAN," *Jurnal Informatika Mulawarman*, vol. 11, no. 2, p. 29, 2016.
- [6] A. Hadianto, B. S. Setya Rintyarna, L. Ali Muharom, and S. Bisnis Cerdas, "KLASIFIKASI MAHASISWA PENERIMA BEASISWA DENGAN METODE NAIVE BAYES," 2015.
- [7] E. Prasetyo, *Data mining : konsep dan aplikasi menggunakan MATLAB*. Yogyakarta: CV Andi Offset.
- [8] D. Sartika and D. I. Sensuse, "Perbandingan Algoritma Klasifikasi Naive Bayes, Nearest Neighbour, dan Decision Tree pada Studi Kasus Pengambilan Keputusan Pemilihan Pola Pakaian," *Jatsi*, vol. 1, no. 2, 2017.
- [9] S. Adi, "IMPLEMENTASI ALGORITMA NAÏVE BAYES CLASSIFIER UNTUK KLASIFIKASI PENERIMA BEASISWA PPA DI UNIVERSITAS AMIKOM YOGYAKARTA," *Jurnal Mantik Penusa*, vol. 22, Aug. 2018.
- [10] R. A. Saputra and S. Ayuningtias, "PENERAPAN ALGORITMA NAÏVE BAYES UNTUK PENENTUAN CALON PENERIMA BEASISWA PADA SMK PASIM PLUS SUKABUMI," 2016.
- [11] M. Faid, M. Jasri, and T. Rahmawati, "Perbandingan Kinerja Tool Data Mining Weka dan Rapidminer Dalam Algoritma Klasifikasi," *Teknika*, vol. 8, no. 1, pp. 11–16, Jun. 2019, doi: 10.34148/teknika.v8i1.95.
- [12] H. Rizqifaluthi and M. A. Yaqin, "Process Mining Akademik Sekolah menggunakan RapidMiner," *MATICS: Jurnal Ilmu Komputer dan Teknologi Informasi (Journal of Computer Science and Information Technology)*, vol. 10, no. 2, pp. 47–51, 2018, doi: 10.18860/mat.v10i2.5745.
- [13] A. Salim, "Klasifikasi Kelayakan Penerima Beasiswa SD N 07/1 Desa Pulau Menggunakan Metode Naïve Bayes," Aug. 2020.
- [14] A. Ulfa Kurnia, A. Setia Budi, and P. H. Susilo, "SISTEM PENDUKUNG KEPUTUSAN PENERIMAAN BEASISWA MENGGUNAKAN METODE NAIVE BAYES," *JOUTICA*, vol. 5, no. 2, 2020.
- [15] D. Alita, I. Sari, and A. Rahman Isnain, "PENERAPAN NAÏVE BAYES CLASSIFIER UNTUK PENDUKUNG KEPUTUSAN PENERIMA BEASISWA," *JDMSI*, vol. 2, no. 1, p. 702022, 2021.
- [16] H. Annur, "KLASIFIKASI MASYARAKAT MISKIN MENGGUNAKAN METODE NAÏVE BAYES," *ILKOM Jurnal Ilmiah*, vol. 10, Aug. 2018.
- [17] W. P. Nurmayanti, "Penerapan Naive Bayes dalam Mengklasifikasikan Masyarakat Miskin di Desa Lepak," *Geodika: Jurnal Kajian Ilmu dan Pendidikan Geografi*, vol. 5, no. 1, pp. 123–132, Jun. 2021, doi: 10.29408/geodika.v5i1.3430.