

Penerapan Metode Naïve Bayes Untuk Penentuan Penerima Beasiswa Program Indonesia Pintar

Amri Muliawan Nur^{1*}, Weni Nursali², Nurhidayati³, Imam Fathurrahman⁴

^{1,2,3}Program Studi Sistem Informasi, Universitas Hamzanwadi

⁴Program Studi Teknik Informatika, Universitas Hamzanwadi

*muliaamriga@gmail.com

Abstrak

PIP merupakan pemberian bantuan tunai pendidikan kepada anak usia sekolah yang berasal dari keluarga kurang mampu yang ditandai dengan kartu Indonesia pintar (KIP). Tujuan dari penelitian ini untuk mengetahui kinerja dari metode naïve bayes dalam melakukan klasifikasi data siswa yang layak dan tidak layak untuk mendapatkan beasiswa PIP di SMAN 1 Sukamulia, dikarenakan sekolah ini masih mengalami kendala pada proses pengambilan keputusan untuk penentuan calon penerima beasiswa PIP, karena belum adanya sistem yang dapat membantu dalam proses pengolahan data siswa yang layak dan tidak layak untuk mendapatkan beasiswa PIP tersebut. Oleh karena itu, untuk pengolahan data peneliti mencoba menerapkan sebuah sistem baru dengan konsep Data Mining menggunakan metode Naïve bayes, dengan melakukan sebanyak 9 kali pengujian menggunakan Cross Validation mulai dari K-Fold Validation 2 sampai dengan 10, didapatkan hasil accuracy tertinggi pada pengujian ke 9 menggunakan K-Fold Validation 10 yaitu sebesar 92.81%. Serta didapatkan juga nilai Area Under Curve (AUC) sebesar 0.973%, dimana AUC merupakan parameter yang digunakan dalam analisis klasifikasi dalam menentukan model terbaik untuk prediksi suatu kelas atau atribut. AUC sendiri memiliki rentang nilai 0-1, yang artinya semakin nilai AUC mendekati angka 1 maka prediksi atau diagnosa atributnya semakin bagus..

Kata kunci: Data Mining, Naïve Bayes, PIP

Abstract

PIP is the provision of educational cash assistance to school-aged children from underprivileged families who are marked with a smart Indonesia card (KIP). The purpose of this research is to determine the performance of the Naïve Bayes method in classifying data on students who are eligible and who are not eligible to receive a PIP scholarship at SMAN 1 Sukamulia, because this school is still experiencing problems in the decision-making process for determining potential PIP scholarship recipients, because there are no a system that can assist in processing student data that is eligible and not eligible to get the PIP scholarship. Therefore, for data processing, researchers tried to implement a new system with the Data Mining concept using the Naïve Bayes method, by carrying out 9 tests using Cross Validation starting from K-Fold Validation 2 to 10, obtaining the highest accuracy results in the 9th test. using K-Fold Validation 10 which is equal to 92.81%. Also obtained was an Area Under Curve (AUC) value of 0.973%, where AUC is a parameter used in classification analysis to determine the best model for predicting a class or attribute. AUC itself has a value range of 0-1, which means that the closer the AUC value is to 1, the better the prediction or diagnosis of the attribute.

Keywords: Data Mining, Naïve Bayes, PIP

1. Pendahuluan

Jaminan kehidupan yang layak bagi setiap individu di masa yang akan datang salah satunya didukung oleh faktor pendidikan yang baik.

Pendidikan merupakan salah satu hak yang dimiliki oleh setiap insan begitupula di Negara Kesatuan Republik Indonesia, oleh sebab itu negara menjamin hak dari setiap individu untuk

mengenyam Pendidikan. Sesuai dengan ketentuan pasal 31 Undang Undang Dasar 1945, setiap warga negara berhak mendapatkan Pendidikan [1]. Pendidikan sangatlah penting bagi generasi muda, karena kemajuan bangsa dimasa yang akan datang bergantung pada generasi mudanya.

Berkaca dari kehidupan, banyak generasi muda yang mengalami putus sekolah, disebabkan karena factor biaya. Untuk menyikapi permasalahan tersebut, maka pemerintah telah mengeluarkan berbagai program yang dapat membantu genererasi muda, dari tingkat dasar sampi menengah untuk melanjutkan Pendidikan, salah satunya yaitu beasiswa PIP (Program Indonesia Pintar).

PIP merupakan pemberian beasiswa bantuan tunai pendidikan kepada anak usia sekolah, yang berasal dari keluarga kurang mampu, ditandai dengan kartu Indonesia pintar (KIP). Tujuan pemerintah memberikan beasiswa PIP yaitu untuk membantu biaya pendidikan bagi peserta didik miskin atau rentan miskin yang masih terdaftar sebagai peserta didik pada jenjang pendidikan dasar dan menengah. Akan tetapi, terkadang sasaran penentuan penerima beasiswa PIP tidak tepat sasaran, karena sistem penentuan kelayakan penerima bantuan beasiswa PIP, banyak yang masih mendata siswa secara manual, hal ini juga terjadi pada SMAN 1 Sukamulia, yaitu masih mengalami kendala pada

proses pengambilan keputusan untuk penentuan calon penerima beasiswa PIP, hal ini disebabkan karena belum adanya sistem yang dapat membantu dalam proses pengolahan data siswa yang layak dan tidak layak untuk memperoleh beasiswa PIP tersebut dengan cepat dan tepat. Sehingga, untuk membantu pihak sekolah, peneliti mencoba menerapkan sebuah sistem baru dengan menggunakan konsep Data mining. Data mining merupakan proses pencarian data atau ekstraksi data untuk menentukan pola yang tersembunyi berupa *knowledge* (pengetahuan) yang tidak diketahui sebelumnya dari sekumpulan data di dalam dataset. Data mining juga bisa memudahkan kita dalam proses pengambilan keputusan untuk menyelesaikan suatu permasalahan, dengan melakukan pencarian informasi dalam kumpulan data yang besar [2]. Dalam data mining terdapat beberapa metode yang digunakan untuk melakukan pengolahan data, salah satunya yaitu metode *naïve bayes*.

Naïve bayes merupakan metode yang digunakan untuk keperluan klasifikasi atau pengelompokkan suatu data. *Naïve bayes* dapat memprediksi probabilitas dimasa depan berdasarkan pengalaman sebelumnya. Untuk itu penulis mencoba melakukan pengolahan data menggunakan metode *naïve bayes*, karena metode ini memiliki keunggulan dalam mengklasisfikasi data dengan baik. Sehingga nantinya, hasil yang diperoleh dapat digunakan

sebagai acuan dalam pengambilan keputusan terkait kelayakan penerima PIP (Program Indonesia Pintar).

2. Tinjauan Pustaka

2.1. Penelitian Terkait

Berikut ini terdapat beberapa penelitian yang terkait dengan penelitian yang akan penulis lakukan.

Penelitian pertama yang dilakukan oleh (Nurhidayati dkk) tahun 2023, dengan judul “Implementasi Algoritma Naive Bayes Untuk Klasifikasi Penerima Beasiswa (Studi Kasus Universitas Hamzanwadi)” hasil pengujian yang diperoleh dengan melakukan 10 kali pengujian menggunakan cross validation yaitu *k-fold* 2 sampai *k-fold* 10, didapatkan nilai akurasi tertinggi sebesar 91,43% dari *k-fold* 4 validation. Sedangkan AUC didapatkan 0.996%, algoritma naïve Bayes ini lebih akurat dan sangat bagus digunakan dalam menganalisis serta mengklasifikasi penerima beasiswa bidikmisi dan bisa dijadikan sebagai bahan evaluasi dalam mengambil keputusan untuk proses seleksi beasiswa selanjutnya [3].

Penelitian kedua yang dilakukan oleh (Yahya dan Hariman Bahtiar) tahun 2021, dengan judul “Pengaruh Pertumbuhan Ekonomi Terhadap Tingkat Kesejahteraan Masyarakat Kecamatan Selong Kabupaten Lombok Timur – Nusa Tenggara Barat Menggunakan Algoritma Naive

Bayes” Hasil dari penelitian ini adalah dengan melakukan sebanyak 6 kali percobaan menggunakan *K-Fold Validation* yang berbeda-beda. Nilai akurasi tertinggi diperoleh pada percobaan menggunakan *K-Fold Validation* 10 dengan nilai akurasi 93,45%, dapat disimpulkan bahwa masyarakat yang tergolong masyarakat yang sudah sejahtera apabila dilihat dari faktor ekonomi, dengan tingkat akurasi kedekatan dengan keadaan yang sebenarnya sebesar 93,45%. Dari data yang diperoleh 93,45% atau $0,9345 \times \text{jumlah data (kk)} = 0,934 \times 1130 \text{ kk} = 1056 \text{ kk}$ yang menunjukkan masyarakatnya sejahtera dan $6,55\% \times 1130 = 74 \text{ kk}$ yang menyatakan masyarakat yang tidak sejahtera dan dapat dijadikan sebagai acuan dalam pengentasan kemiskinan melalui program yang sudah dicanangkan oleh pemerintah [4].

Penelitian ketiga yang dilakukan oleh (Nurhidayati dan Suhartini) tahun 2021, dengan judul “Probabilitas Pembayaran Kredit Tepat Waktu Menggunakan Algoritma Naive Bayes Pada Koperasi Serba Usaha Daruzzakah Rensing Lombok Timur” Setelah dilakukan pengujian sebanyak 9 kali yaitu dari validation 2 sampai dengan 10 maka didapatkan dua hasil akurasi terbaik dari *K-Fold Validation* 9 dan *K-Fold Validation* 10 yang masing-masing memiliki nilai accuracy sebesar 96,43% pada *K-Fold Validation* 9 dan 96,45% dari *K-Fold Validation* 10. Dan selisih akurasi keduanya adalah 0.02% yang

menunjukkan bahwa terjadi peningkatan 0.02% pada rentang *k-fold validation* 9 dan 10. nilai *accuracy* terbaik pada *K-Fold Validation* 10 dengan akurasi sebesar 96,45% dimana data dibagi menjadi 10 bagian untuk ditraining dan ditesting serta didapatkan juga nilai *AUC* (*Area Under Curve*) sebesar 0.942 yang dimana *AUC* sendiri mempunyai rentang nilai dari 0 sampai 1, yang artinya semakin nilai *AUC* mendekati angka 1 maka prediksi atau diagnosa atributnya semakin bagus. Ini menunjukkan nilai *AUC* sebesar 0.942 yang didapatkan setelah melakukan pengujian dengan *K-Fold Validation* 10 sangat baik karena hampir mendekati angka 1 [5].

Penelitian keempat yang dilakukan oleh (Amri Muliawan Nur dkk) tahun 2020, dengan judul "Implementation of Naive Bayes Algorithm in Analyzing Acceptance of Poor Student Assistance" hasil pengolahan dengan menggunakan metode Naive Bayes dengan melakukan percobaan menggunakan K-Fold Validasi 10, bahwa pengolahan data siswa di MA Hizbul Wathan NW Semaya dapat diketahui hasil yang diperoleh sebesar 95,27% dari 169 data siswa dengan nilai *AUC* 1. Nilai ini termasuk dalam klasifikasi sangat baik. Dari hasil pengolahan data, metode yang digunakan dapat menganalisa penerimaan bantuan siswa miskin dengan sangat baik [6].

Penelitian terakhir yang dilakukan oleh (Muhammad Wasil dan Mahfuz) tahun 2020,

dengan judul penelitian "Analisis Persepsi Masyarakat Terhadap Kinerja Aparatur Desa Dalam Meningkatkan Pelayanan Publik Di Desa Lenek Lauk Menggunakan Metode *Naive Bayes*" Hasil penelitian ini mendapatkan nilai akurasi sebesar 92.26% dan nilai *AUC* pada Kurva *ROC* sebanyak 0.924. Berdasarkan pengujian yang telah dilakukan bahwa algoritma *Naive Bayes* mampu memberikan nilai akurasi yang terbaik [7]

2.2. Landasan Teori

1. Pengertian Implementasi

Implementasi adalah proses membentuk suatu kaitan yang dilaksanakan untuk mencapai suatu tujuan, biasa direalisasikan sebagai dampak dari suatu kegiatan dimana sarana-sarana tertentu telah dirancang dan dijalankan dengan harapan sampai pada tujuan tercapai, sehingga nantinya akan menghasilkan ide, konsep, kebijakan, atau inovasi dalam suatu tindakan [8] [9].

2. Pengertian Klasifikasi

Klasifikasi merupakan suatu proses mengelompokkan data berdasarkan kriteria dan kesamaan terhadap variable data dengan tujuan untuk membentuk model yang dapat memprediksi data baru yang belum diketahui kelasnya. Klasifikasi dapat dilakukan dengan beberapa cara yaitu dilakukan dengan cara manual ataupun dengan bantuan teknologi komputasi [2][10].

3. Program Indonesia Pintar (PIP)

Program Indonesia Pintar merupakan bantuan yang diberikan Pemerintah kepada siswa yang kurang mampu berupa uang tunai, yang diperuntukan guna perluasan akses, dan kesempatan belajar yang diberikan kepada peserta didik, yang berasal dari keluarga miskin atau rentan miskin untuk membiayai pendidikan. Tiga syarat Program Indonesia Pintar (PIP) dari kelurahan yaitu data siswa pemegang KIP, Surat Keputusan dan struktur penanggung jawab KIP [11][12].

4. Naïve Bayes

Naïve Bayes adalah pengklasifikasian statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu class. *Bayesian classification* didasarkan pada *teorema Bayes* yang memiliki kemampuan klasifikasi serupa dengan *decision tree* dan *neural network*. *Bayesian classification* terbukti memiliki akurasi dan kecepatan yang tinggi saat diaplikasikan ke dalam database dengan data yang besar.

Berikut Persamaan dari Teorema Bayes:

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

Keterangan:

X : Data dengan class yang belum diketahui

H : Hipotesis data X merupakan suatu class spesifik

P(H|X) : Probabilitas hipotesis H berdasarkan kondisi x (posteriori prob.)

P(H) : Probabilitas hipotesis H

P(X|H) : Probabilitas X berdasarkan kondisi tersebut

P(X) : Probabilitas dari X [13][14].

5. Pengertian Data Mining

Data Mining merupakan proses pengumpulan dan pengolahan data yang bertujuan untuk mengekstrak informasi penting pada data. Proses pengumpulan dan ekstraksi informasi tersebut dapat dilakukan menggunakan perangkat lunak dengan bantuan perhitungan statistika, matematika, ataupun teknologi *Artificial Intelligence (AI)*. Data mining sering disebut juga *Knowledge Discovery in Database (KDD)* [15][16]

3. Metode Penelitian

3.1. Metode Pengumpulan Data

Untuk mendapatkan data sebagai pemenuhan penelitian dan data yang diperoleh benar-benar akurat serta dapat digunakan dalam penelitian nantinya, maka dari itu peneliti menggunakan beberapa metode pengumpulan data. Berikut ini beberapa metode yang penulis gunakan dalam penelitian :

1. Observasi

Teknik observasi yang dilakukan adalah mengumpulkan data dengan melakukan observasi atau pengamatan langsung ke sekolah SMAN 1 Sukamulia untuk melakukan riset / penelitian dan mencari informasi-

informasi yang dibutuhkan dalam menyelesaikan penelitian ini.

2. Wawancara

Wawancara merupakan suatu teknis pengumpulan data melalui tatap muka dan tanya jawab langsung dengan sumber data yaitu Pak Waka Kurikulum, Pembina Osis, Operator Dapodik, dan pihak sekolah yang berkaitan dengan penelitian ini pada sekolah SMAN 1 SUKAMULIA, Kec. Sukamulia, Kab. Lombok Timur.

3. Studi Pustaka

Suatu metode pengumpulan data oleh peneliti dengan cara membaca dan mempelajari buku-buku yang terkait, serta mencari refrensi yang ada diinternet (*browsing*) tentang artikel/ jurnal yang berkaitan dengan penelitian ini.

3.2. Lokasi Penelitian

Adapun lokasi/ tempat yang menjadi obyek penelitian yang penulis ambil yaitu di sekolah SMAN 1 Sukamulia, Kecamatan Sukamulia Kabupaten Lombok Timur.

4. Hasil dan Pembahasan

4.1. Hasil Penelitian

Data diolah menggunakan Algoritma naïve bayes dengan bantuan tools rapidminer. Proses pengujian dilakukan sebanyak 9 kali terhadap dataset, dengan tujuan untuk mendapatkan hasil pengolahan yang terbaik. Prose pengujian

dilakukan dari K-Fold Validation 2 sampai dengan K-Fold Validation 10. Data yang digunakan yaitu data *private* tahun 2023 yang berjumlah 320 data, diperoleh dari sekolah SMAN 1 Sukamulia. Ada pun variabel data yang digunakan dalam pengolahan yaitu: Nama Siswa, Pekerjaan Orang Tua, Penghasilan Orang Tua, Tanggungan Orang Tua, Penerima KIP, Penerima KPS, Status Siswa, dan Ranking Siswa. Variabel tersebut merupakan variabel penentu untuk mengetahui kelayakan siswa yang mendapatkan beasiswa PIP.

Dari keseluruhan pengujian diperoleh hasil tertinggi pada pengujian ke 9 menggunakan *K-Fold Validation* 10 dengan nilai *accuracy* sebesar 92.81% dengan nilai *AUC* sebesar 0.973. Hasil ini menunjukkan klasifikasi yang peroleh sangat baik (*excellent classification*).

Berikut adalah hasil akurasi tertinggi yang diperoleh dari pengujian menggunakan *K-Fold Validation* 10 menggunakan tools RapidMiner.

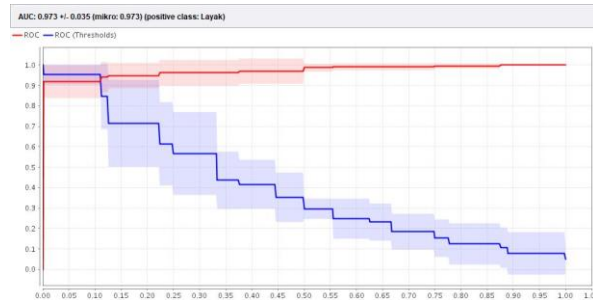
| accuracy: 92.81% +/- 4.43% (mikro: 92.81%) | | | |
|--|------------------|------------|-----------------|
| | true Tidak Layak | true Layak | class precision |
| pred. Tidak Layak | 70 | 7 | 90.91% |
| pred. Layak | 16 | 227 | 93.42% |
| class recall | 81.40% | 97.01% | |

Gambar 1. Akurasi K-Fold Validation 10

Tabel 1. Accuracy, Sensitivity, Specificity, Ppv, dan Npv K-Fold Validation 10

| Nilai (Value) % | |
|--------------------|---------|
| <i>Accuracy</i> | 92,81 % |
| <i>Sensitivity</i> | 97,00 % |

| | |
|-------------|---------|
| Specificity | 81,39 % |
| Ppv | 93,41 % |
| Npv | 90,90 % |



Gambar 2. Hasil AUC K-Fold Validation 10

Hasil AUC (Area Under Curve) yang didapatkan yaitu berjumlah 0,973 sehingga termasuk kedalam klasifikasi sangat baik (*excellent classification*).

Hasil pengujian secara keseluruhan dapat dilihat pada tabel di bawah ini.

Tabel 2. Rekapitulasi Hasil Pengujian Number of Validation 2-10

| Nilai-K | Accuracy (%) | AUC | Precision (%) | Recall (%) |
|---------|--------------|-------|---------------|------------|
| 2 | 89.06 | 0.955 | 90.29 | 95.30 |
| 3 | 91.88 | 0.974 | 92.29 | 97.01 |
| 4 | 91.56 | 0.963 | 92.38 | 96.59 |
| 5 | 92.19 | 0.975 | 93.48 | 96.16 |
| 6 | 91.25 | 0.972 | 92.43 | 96.15 |
| 7 | 91.28 | 0.970 | 91.44 | 97.44 |
| 8 | 91.56 | 0.967 | 92.02 | 97.03 |
| 9 | 91.24 | 0.975 | 92.02 | 96.58 |
| 10 | 92.81 | 0.973 | 93.66 | 97.07 |

4.2. Pembahasan

Algoritma naïve bayes merupakan salah satu algoritma dari metode klasifikasi yang handal dalam memprediksi probabilitas keanggotaan suatu class. Adapun perhitungan untuk mendapatkan nilai kelas dan partisi data dapat dilakukan dengan formula dibawah ini:

$$\frac{P(C|X)=P(x|c)P(c)}{P(x)}$$

Keterangan:

x : Data dengan class yang belum diketahui

c : Hipotesis data merupakan suatu class spesifik

P(c|x): Berdasarkan kondisi (posteriori probability)

P(c) : Probabilitas class yang di cari (prior probability) dari keseluruhan data

P(x|c): Probabilitas berdasarkan kondisi pada class

P(x) : Probabilitas c

Berikut perhitungan probabilitas dari masing-masing kelas dan partisi data :

1. Menghitung jumlah kelas/ label

Kelas/ label yang digunakan dalam penelitian ini adalah kelas layak dan kelas tidak layak. Dari 320 dataset dalam penelitian ini, akan dibagi menjadi 2 partisi yaitu data yang menjadi data training dan data yang di jadikan sebagai data testing.

pengimplementasian rumus di atas untuk mencari nilai probabilitas kelas “Layak” dan kelas “Tidak Layak” dihitung sebagai berikut :

Tabel 3. Probabilitas Kelas

| Kelas Layak | Kelas Tidak Layak |
|--|--|
| $P(\text{Layak}) = \frac{\text{total (Layak)}}{\text{total (keseluruhan data)}}$ | $P(\text{Tidak Layak}) = \frac{\text{total (Tidak Layak)}}{\text{total (keseluruh data)}}$ |
| $P(\text{Layak}) = \frac{115}{160} = 0,718$ | $P(\text{Tidak Layak}) = \frac{45}{160} = 0,28$ |
| $P(\text{Layak}) = 72\%$ | $P(\text{Tidak Layak}) = 28\%$ |

2. Menghitung jumlah kasus perkelas atau menghitung nilai probabilitas dari masing-masing atribut.

Pada penelitian ini yang menjadi atribut adalah Nama Siswa, Pekerjaan Orang Tua, Penghasilan Orang Tua, Tanggungan Orang Tua, Penerima KIP, Penerima KPS, Status Siswa, dan Ranking Siswa. Dengan menggunakan rumus probabilitas yang sama untuk mencari nilai probabilitas masing – masing atribut dapat dihitung dengan cara seperti langkah di atas.

Tabel 4. Probabilitas Atribut Pekerjaan Orang Tua (POT)

| Layak | Tidak Layak |
|--|---|
| $P(\text{POT} \setminus \text{buruh}) = \frac{\text{total(Layak)}}{\text{total (POT layak)}}$ | $P(\text{POT} \setminus \text{buruh}) = \frac{\text{total(Tidak Layak)}}{\text{total (POT tidak layak)}}$ |
| $P(\text{POT} \setminus \text{buruh}) = \frac{33}{115}$ | $P(\text{POT} \setminus \text{buruh}) = \frac{13}{45}$ |
| $P(\text{POT} \setminus \text{buruh}) = 29\%$ | $P(\text{POT} \setminus \text{buruh}) = 29\%$ |

Lakukan hal yang sama untuk mencari nilai atribut / variabel yang lain.

3. Langkah selanjutnya, kelompokkan dan kalikan semua variabel kelas yang bernilai “Layak dan “Tidak Layak” dan Bandingkan hasil perkelas.

$$\begin{aligned}
 & \text{Kelas Layak} = (\text{Layak PIP} / \text{Layak}) \times (\text{Pekerjaan Orang Tua} | \text{Buruh} / \text{Layak}) \times (\text{Penghasilan Orang Tua} | \text{Rendah} / \text{Layak}) \times (\text{Tanggungan Orang Tua} | \text{Banyak} / \text{Layak}) \times (\text{Penerima Kip} | \text{Ya} / \text{Layak}) \times (\text{Penerima KPS} | \text{Ya} / \text{Layak}) \times (\text{Status Siswa} | \text{Tidak} / \text{Layak}) \times (\text{Rangking} | \text{Dua} / \text{Layak}) \\
 & = (115 / 160) \times (33 / 115) \times (42 / 115) \times (38 / 115) \times (69 / 115) \times (55 / 115) \times (84 / 115) \times (5 / 115) \\
 & = 0,000226 = 0,23 \% \text{ untuk nilai kelas prediksi "layak"}.
 \end{aligned}$$

$$\begin{aligned}
 & \text{Kelas Tidak Layak} = (\text{Layak PIP} / \text{Tidak Layak}) \times (\text{Pekerjaan Orang Tua} | \text{Buruh} / \text{Tidak Layak}) \times (\text{Penghasilan Orang Tua} | \text{Rendah} / \text{Tidak Layak}) \times (\text{Tanggungan Orang Tua} | \text{Banyak} / \text{Tidak Layak}) \times (\text{Penerima Kip} | \text{Ya} / \text{Tidak Layak}) \times (\text{Penerima KPS} | \text{Ya} / \text{Tidak Layak}) \times (\text{Status Siswa} | \text{Tidak} / \text{Tidak Layak}) \times (\text{Rangking} | \text{Dua} / \text{Tidak Layak}) \\
 & = (45 / 160) \times (13 / 45) \times (6 / 45) \times (13 / 45) \times (0 / 115) \times (2 / 45) \times (43 / 45) \times (1 / 45) \\
 & = 0 \% \text{ untuk nilai kelas prediksi "tidak layak"}.
 \end{aligned}$$

Karena nilai Layak 0,23 % > nilai Tidak Layak 0 %, maka dapat disimpulkan bahwa data yang

Pekerjaan orang tua | Buruh, Penghasilan Orang Tua | Rendah, Tanggungan Orang Tua | Banyak, Penerima Kip | Ya, Penerima KPS | Ya, Status Siswa | Tidak, Rangking | Dua, data tersebut diklasifikasikan / dikelompokkan kedalam kelas data Layak untuk mendapatkan beasiswa PIP

5. Kesimpulan

Berdasarkan hasil pengolahan data yang dilakukan di SMAN 1 Sukamulia menggunakan 320 data pendaftar beasiswa PIP tahun 2023, dapat disimpulkan bahwa dalam penentuan klasifikasi siswa yang layak dan tidak layak untuk mendapatkan beasiswa Program Indonesia Pintar (PIP) dapat dievaluasi dengan sangat baik menggunakan algoritma naïve bayes. Dari pengujian yang telah dilakukan sebanyak 9 kali pengujian menggunakan Cross Validation, mulai dari K-Fold Validation 2 sampai dengan 10 didapatkan hasil accuracy tertinggi pada pengujian ke 9 menggunakan K-Fold Validation 10 yaitu sebesar 92.81 % dengan Area Under Curve (AUC) sebesar 0.973%, yang menjelaskan bahwa kasifikasi algoritma naïve bayes sangat baik (excellent classification), dimana AUC merupakan parameter yang digunakan dalam analisis klasifikasi dalam menentukan model terbaik untuk prediksi suatu kelas atau atribut. AUC sendiri memiliki rentang nilai 0-1, yang artinya semakin nilai AUC mendekati angka 1 maka klasifikasi atau pengelompokan atributnya

semakin bagus. Ini menunjukkan bahwa nilai AUC sebesar 0.973% yang didapatkan setelah melakukan pengujian K-Fold Validation 10 sangat baik (excellent classification) karena mendekati angka 1..

6. Daftar Pustaka

- [1] J. J. Lanawaang and R. Mesra, "Faktor Penyebab Anak Putus Sekolah di Kelurahan Tuutu Analisis Pasal 31 Ayat 1, 2, dan 3 UUD 1945," *J. Ilm. Mandala Educ.*, vol. 9, no. 2, 2023.
- [2] Z. Setiawan *et al.*, *Buku Ajar Data Mining*. PT. Sonpedia Publishing Indonesia, 2023.
- [3] N. I. Nurhidayati, Y. Y. Yahya, F. F. Fathurrahman, and L. M. S. S. L. M. Samsu, "Implementasi Algoritma Naive Bayes Untuk Klasifikasi Penerima Beasiswa (Studi Kasus Universitas Hamzanwadi)," *Infotek J. Inform. dan Teknol.*, vol. 6, no. 1, pp. 177–188, 2023.
- [4] Y. Yahya and H. Bahtiar, "Pengaruh Pertumbuhan Ekonomi Terhadap Tingkat Kesejahteraan Masyarakat Kecamatan Selong Kabupaten Lombok Timur – Nusa Tenggara Barat Menggunakan Algoritma Naive Bayes," *Infotek J. Inform. dan Teknol.*, vol. 4, no. 1, pp. 20–28, 2021, doi: 10.29408/jit.v4i1.2981.
- [5] N. Hidayati and S. Suhartini, "Probabilitas Pembayaran Kredit Tepat Waktu Menggunakan Algoritma Naive Bayes Pada Koperasi Serba Usaha Daruzzakah Rensing Lombok Timur," *Infotek J. Inform. dan Teknol.*, vol. 4, no. 2, pp. 194–204, 2021.
- [6] A. M. Nur, M. F. Wazdi, B. Harianto, and M. F. Zaini, "Implementation of Naive Bayes Algorithm in Analyzing Acceptance of Poor Student Assistance," *J. Phys. Conf. Ser.*, vol. 1539, no. 1, 2020, doi: 10.1088/1742-6596/1539/1/012018.

- [7] M. Wasil, "Infotek : Jurnal Informatika dan Teknologi Analisis Persepsi Masyarakat Terhadap Kinerja Aparatur Desa Dalam Meningkatkan Pelayanan Publik Di Desa Lenek Lauk Menggunakan Metode Naive Bayes Pelayanan publik adalah kegiatan atau rangkaian kegiatan dalam ran," vol. 3, no. 2, pp. 140–146, 2020.
- [8] B. T. Haji, "Pengertian Implementasi," *Lap. AKHIR*, vol. 31, 2020.
- [9] Uin-suska, "Pengertian Implementasi dalam Penelitian".
- [10] A. P. Wibawa, M. Guntur, A. Purnama, M. F. Akbar, and F. A. Dwiyanto, "Metode-metode klasifikasi," in *Prosiding Seminar Ilmu Komputer dan Teknologi Informasi*, 2018.
- [11] D. A. Astuti, D. Febriyanti, and M. Q. Kariem, "Evaluasi Kebijakan Program Kartu Indonesia Pintar (Pip) Di Kelurahan Tuan Kentang Tahun 2020," *TheJournalish Soc. Gov.*, vol. 4, no. 3, pp. 249–256, 2023.
- [12] H. Retnaningsih, "Program Indonesia Pintar: Implementasi Kebijakan Jaminan Sosial Bidang Pendidikan," *Aspir. J. Masal. Sos.*, vol. 8, no. 2, pp. 161–177, 2019.
- [13] H. Annur, "Klasifikasi Masyarakat Miskin Menggunakan Metode Naive Bayes," *Ilk. J. Ilm.*, vol. 10, no. 2, pp. 160–165, doi: 10.33096/ilkom.v10i2.303.160-165.
- [14] J. Han, J. Pei, and H. Tong, *Data mining: concepts and techniques*. Morgan kaufmann, 2022.
- [15] S. Suhirman, "Data Warehouse dan Mining," *GAES-PACE B. Publ.*, pp. 14–25, 2023.
- [16] P. Bhatia, *Data mining and data warehousing: principles and practical techniques*. Cambridge University Press, 2019.