

IMPLEMENTASI *DATA MINING* MENGGUNAKAN METODE *NAIVE BAYES* UNTUK PENENTUAN PENERIMA BANTUAN PROGRAM INDONESIA PINTAR (PIP) (STUDI KASUS : SD NEGERI 9 AIR KUMBANG)

Okta Rini¹, Suzi Oktavia Kunang²

Fakultas Teknik Ilmu Komputer, Universitas Bina Darma

Email: 10oktarini@gmail.com¹, suzi_oktavia@binadarma.ac.id²

ABSTRACT

The Smart Indonesia Program (PIP) through the Smart Indonesia Card (KIP) is the provision of educational cash assistance to school-age children (6-21 years). KIP is an improvement part of the Poor Student Assistance Program (BSM) since late 2014 [1]. SD Negeri 9 Air Kumbang is located on Jl Inpres Nusa Makmur Village, Air Kumbang District, Banyuasin Regency. This elementary school is one of the elementary schools in the Banyuasin area that received funding from the Smart Indonesia Program (PIP). The target of PIP at SDN 9 Air Kumbang is still not on target, due to the lack of criteria for the number of dependents, therefore the authors add criteria for the number of dependents in the study. This research was made based on data that already existed before, namely with 101 training data. Using the Naïve Bayes method and with 5 attributes, namely parents' occupation, parents' income, number of dependents, KIP recipients, and KPS recipients. using Rapid miner support tools in testing the accuracy of the Naïve Bayes method. The accuracy test results obtained using the RapidMiner application and manual calculations obtained an accuracy of 90.00% and the resulting classification is included in the Good Classification group because the AUC value obtained from testing based on the ROC curve using the Naive Bayes method is 0.860. So, it can be concluded that the Naive Bayes Algorithm can be applied to determine the feasibility of accepting the Smart Indonesia program for students at SD Negeri 9 Air Kumbang.

Keywords: Data Mining, PIP, Naive Bayes, Rapid Miner

ABSTRAK

Program Indonesia Pintar (PIP) melalui Kartu Indonesia Pintar (KIP) adalah pemberian bantuan tunai pendidikan kepada anak usia sekolah (6-21 tahun). KIP merupakan bagian penyempurnaan dari Program Bantuan Siswa Miskin (BSM) sejak akhir 2014 [1]. SD Negeri 9 Air Kumbang bertempat di Jl Inpres Desa Nusa Makmur Kecamatan Air Kumbang Kabupaten Banyuasin. SD ini merupakan salah satu SD yang berada di kawasan banyuasin yang mendapatkan dana Program Indonesia Pintar (PIP). Sasaran PIP di SDN 9 Air Kumbang masih kurang tepat sarannya, di karenakan kurangnya kriteria jumlah tanggungan, Oleh karena itu penulis menambahkan kriteria jumlah tanggungan dalam penelitian. Penelitian ini di buat berdasarkan data yang telah ada sebelumnya yaitu dengan 101 data training. Memakai metode *Naive Bayes* dan dengan 5 atribut yaitu Pekerjaan orang tua, Penghasilan orang tua, Jumlah tanggungan, Penerima KIP, dan Penerima KPS. menggunakan tols pendukung *RapidMiner* dalam pengujian keakuratan metode *Naive Bayes*. Hasil pengujian akurasi yang didapatkan menggunakan aplikasi *RapidMiner* maupun perhitungan manual diperoleh akurasi yaitu 90.00% dan klasifikasi yang dihasilkan termasuk ke dalam kelompok *Good Classification* karena nilai AUC yang didapatkan dari pengujian berdasarkan kurva ROC menggunakan metode *Naive Bayes* sebesar 0.860. Jadi, dapat disimpulkan

bahwa *Algoritma Naive Bayes* dapat diterapkan untuk penentuan kelayakan penerimaan program indonesia pintar pada siswa SD Negeri 9 Air Kumbang.

Kata Kunci: *Data Mining, PIP, Naive Bayes, Rapid Miner*

1. PENDAHULUAN

Program Indonesia Pintar (PIP) melalui Kartu Indonesia Pintar (KIP) adalah pemberian bantuan tunai pendidikan kepada anak usia sekolah (6-21 tahun). KIP merupakan bagian penyempurnaan dari Program Bantuan Siswa Miskin (BSM) sejak akhir 2014 [1]. Sasaran PIP di SDN 9 Air Kumbang masih kurang tepat sarannya, di karenakan kurangnya kriteria dalam penentuan PIP yaitu kriteria jumlah tanggungan. Data yang akurat sangatlah di perlukan, Oleh karena itu penulis menambahkan kriteria jumlah tanggungan dalam penelitian.

Data mining merupakan teknik menggali informasi berharga, terpendam, tersembunyi didalam data (database) dengan jumlah besar dengan begitu ditemukannya pola menarik yang sebelumnya belum diketahui, Data mining merupakan serangkaian proses menentukan hubungan pola dengan memiliki tujuan menyaring data yang sangat besar dan di dapatkan sifat baru, bermanfaat, benar terhadap pemahaman suatu pola atau variasi dari database [2].

Metode Naïve Bayes adalah metode yang hanya membutuhkan jumlah data pelatihan (training data) yang kecil untuk menentukan estimasi parameter yang diperlukan dalam proses pengklasifikasian [3]. Dengan 5 atribut yaitu Pekerjaan orang tua, Penghasilan orang tua, Jumlah tanggungan, Penerima KIP, dan Penerima KPS. Dengan atribut saling berhubungan untuk penentuan kelayakan. Penulis menggunakan tols pendukung Rapid miner dalam pengujian keakuratan metode Naïve Bayes. Pengujian dengan 101 data training

2. METODOLOGI PENELITIAN

2.1 Metode Pengumpulan Data

Metode pengumpulan data dalam penelitian adalah sebagai berikut.

- a. Pengamatan Langsung (Observasi)
Untuk mendapatkan informasi-informasi yang berhubungan dengan suatu kejadian atau peristiwa yang telah terjadi atau sedang berlangsung disekitar. Dalam hal ini yang dilakukan penulis adalah melakukan pengamatan dan mencatatnya menjadi data penelitian.
- b. Wawancara
Metode Wawancara adalah “metode yang dilakukan untuk mendapatkan informasi yang mendalam tentang persepsi, pandangan, wawasan, atau aspek kepribadian para peserta didik yang diberikan secara lesan dan spontan” [7]. Dalam metode ini kegiatan yang dilaksanakan adalah melakukan diskusi serta tanya jawab dengan sumber yang dianggap memiliki pengetahuan yang lebih dalam dari permasalahan penelitian. Penulis melakukan wawancara kepada Kepala Sekolah dan Oprator Sekolah SD Negeri 9 Air Kumbang.
- c. Studi Pustaka
Metode pengumpulan data dari sumber yang di jadikan rujukan yaitu buku, jurnal, internet yang berhubungan dengan penelitian.

2.2 Data Mining

Data *mining* adalah kegiatan menemukan pola yang menarik dari data dalam jumlah besar, data dapat disimpan dalam database, data warehouse, atau penyimpanan informasi lainnya. Data mining berkaitan dengan bidang ilmu-ilmu lain, seperti database system, data warehousing, statistik, machine learning, information retrieval, dan komputasi tingkat tinggi. Selain itu, data mining di dukung oleh ilmu lain seperti neural network, pengenalan pola, spatial data analysisism image database, signal processing [2].

2.3 *Naive Bayes*

Naive Bayes merupakan teknik prediksi berbasis probabilistik sederhana yang berdasar pada penerapan Teorema Bayes (atau aturan Bayes) dengan asumsi independensi (ketidaktergantungan) yang kuat [5]. Definisi lain mengatakan Metode *Naïve Bayes* yang merupakan sebuah pengklasifikasi probabilitas sederhana yang mengaplikasikan Teorema *Bayes* dengan asumsi ketidaktergantungan (independen) yang tinggi. Keuntungan penggunaan Metode *Naïve Bayes* adalah bahwa metode ini hanya membutuhkan jumlah data pelatihan (*training data*) yang kecil untuk menentukan estimasi parameter yang diperlukan dalam proses pengklasifikasian [6].

Kelebihan metode Naive Bayes:

- 1) Menangani kuantitatif dan data diskrit
- 2) Kokoh untuk titik noise yang diisolasi, misalkan titik yang dirata – ratakan ketika mengestimasi peluang bersyarat data.
- 3) Hanya memerlukan sejumlah kecil data pelatihan untuk mengestimasi parameter (rata – rata dan variansi dari variabel) yang dibutuhkan untuk klasifikasi.
- 4) Menangani nilai yang hilang dengan mengabaikan instansi selama perhitungan estimasi peluang
- 5) Cepat dan efisiensi ruang
- 6) Kokoh terhadap atribut yang tidak relevan

Kekurangan metode naive bayes :

- 1) Mengamsumsi variabel bebas
- 2) Tidak berlaku jika probabilitas kondisionalnya adalah nol, apabila nol maka probabilitas prediksi akan bernilai nol juga.

Langkah-langkah algoritma *Naïve Bayes* menurut Suntoro (2019), sebagai berikut :

- 1) Siapkan dataset.
- 2) Hitung jumlah kelas pada data training.
- 3) Hitung jumlah kasus yang sama dengan kelas yang sama.
- 4) Kalikan semua hasil sesuai dengan data testing yang akan dicari kelasnya.
- 5) Bandingkan hasil perkelas, nilai tertinggi ditetapkan sebagai kelas baru.

2.4 *Rapid Miner*

Rapid Miner berkembang sejak tahun 2001, sebelumnya disebut dengan nama YALE (*Yet Another Learning Environment*). Software ini dikembangkan oleh Ralf Klinkenberg, Ingo Mierswa, serta Simon Fischer pada *Unit Artificial Intelligence* dari *Technical University of Dortmund*. *Rapid Miner* adalah *platform* analisis modern yang meliputi data mining, machine learning, analisis prediktif, text mining dan analisis bisnis. Software ini digunakan untuk mengukur kinerja algoritma dan untuk menemukan algoritma terbaik yang akan berguna untuk klasifikasi, prediksi dan teknik lainnya di data mining, *Rapid Miner* merupakan software yang user friendly dan memiliki GUI (*Graphic User Interface*) yang efektif yang digunakan untuk bekerja dengan mudah.

Rapid Miner memberikan machine learning dan data prosedur termasuk loading data dan transformasi (Extract, Transform, Load(ELT)), data preprocessing dan pemodelan statistik, visualisasi dan analisis prediktif, penyebaran dan evaluasi.

Terdapat sifat-sifat yang dimiliki oleh *Rapid Miner*, yakni:

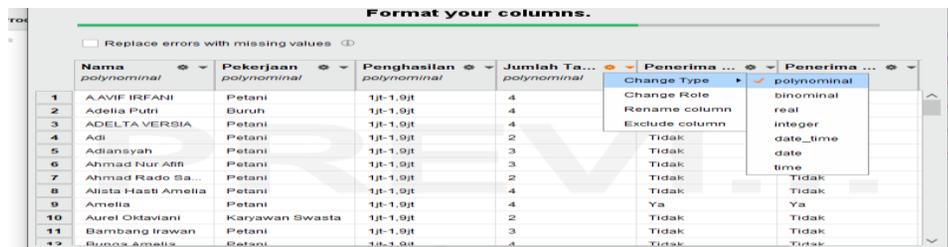
- 1) Penulisan menggunakan bahasa Java, Hal ini memungkinkan *Rapid Miner* bisa berjalan pada sistem operasi yang berbeda-beda.
- 2) Proses menemukan pengetahuan dituangkan model operator trees.
- 3) Merepresentasikan XML internal guna memungkinkan format standar pertukaran data.
- 4) Penggunaan Bahasa *scripting* yang memungkinkan untuk eksperimen dalam skala besar dan pengotomatisasian eksperimen.
- 5) Konsep *multi-layer* yang menjadikan tampilan data menjadi efisien serta memastikan penanganan data.
- 6) Mempunyai GUI(*Graphic User Interface*), command line mode, serta Java API yang bisa dipanggil melalui program lain

3. HASIL DAN PEMBAHASAN

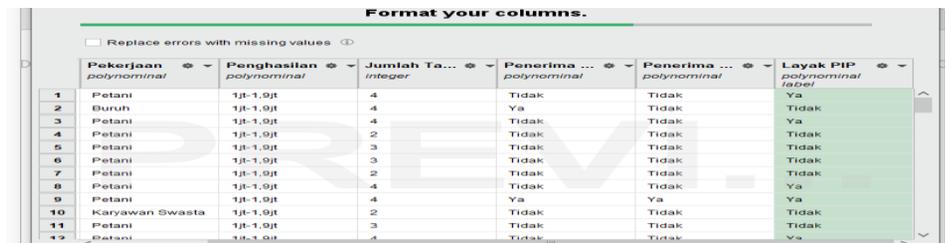
3.1 Proses Pengujian

Klasifikasi pada Rapid Miner menggunakan metode Naive Bayes. Berikut langkah-langkahnya :

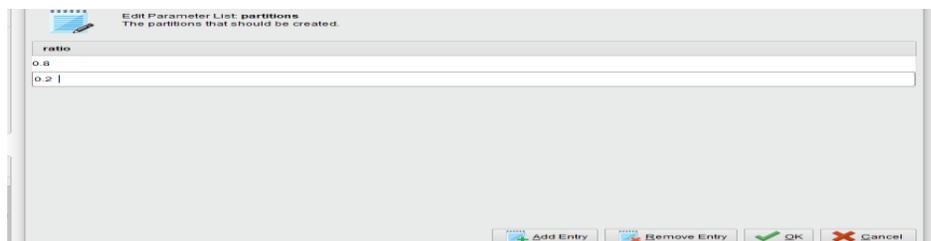
- 1) Buka software RapidMiner, klik New.
- 2) Pada tampilan proses masukan operator retrieve, Drag data pada layar process lalu input data training lalu pada semua atribut pilih change type polynominal, selanjutnya pada atribut layak PIP pilih change role label.
- 3) Masukan operator Split Data lalu buat perbandingan antara data training dan data testing yaitu 80% data training dan 20% data testing, kemudian masukan operator Cross Validation Dan operator Apply Model
- 4) Sambungkan kabel lalu Klik ikon Run berikut tampilan gambarnya:



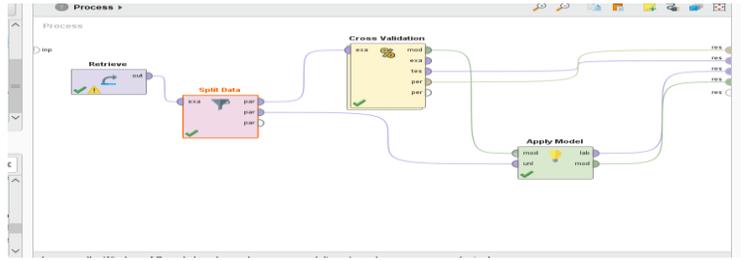
Gambar 1. Change Type



Gambar 2. Change Role



Gambar 3. Split Data



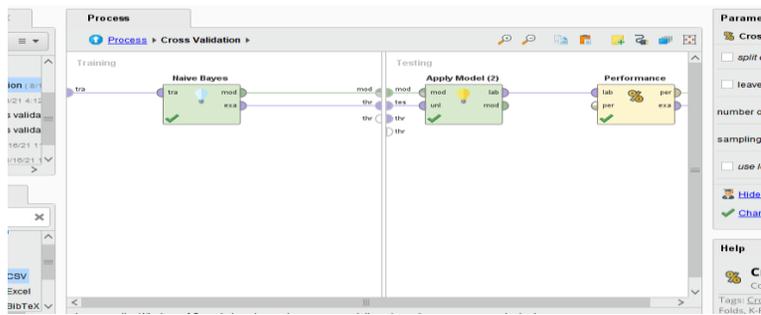
Gambar 4. Proses Pengujian Awal

3.2 Proses Cross Validation 10 Fold

Dalam Cross Validation, jumlah tetap lipatan/partisi dari data ditentukan sendiri. Cara standar untuk memprediksi error rate dari teknik pembelajaran dari sebuah sampel data tetap adalah dengan menggunakan tenfold cross validation. Dengan tenfold cross validation, data akan dibagi secara acak menjadi 10 bagian, dimana class diwakili atau kurang lebih proporsi yang sama seperti pada dataset yang penuh. Setiap bagian mendapatkan giliran dan skema pembelajaran dilatih pada sisa sembilan persepuluh, kemudian error rate dihitung pada holdout set. Dengan demikian, prosedur pembelajaran dilaksanakan sebanyak 10 kali di training set yang berbeda karena setiap set memiliki banyak kesamaan dengan yang lain. Akhirnya, 10 estimasi error dirata-rata untuk menghasilkan perkiraan kesalahan keseluruhan [8].

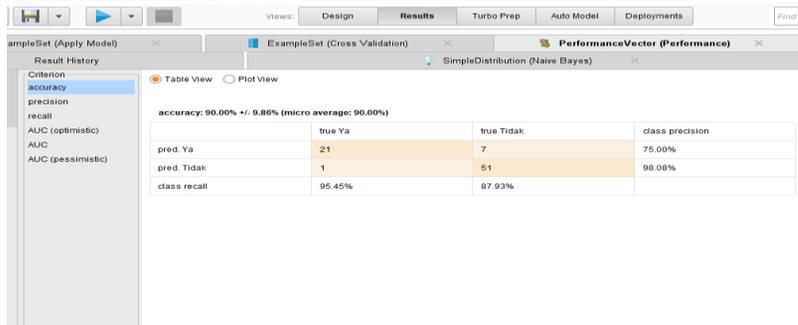
Pada pemodelan Cross Validation digunakan untuk memberikan pelatihan yang di dalamnya terdapat 3 bagian yaitu:

- 1) Naive Bayes digunakan pada bagian training
- 2) Fitur Apply Model untuk mengaplikasikan model pada data testing.
- 3) Fitur Performance untuk menampilkan confusion table yang digunakan untuk menampilkan hasil dari accuracy, precision, recall.



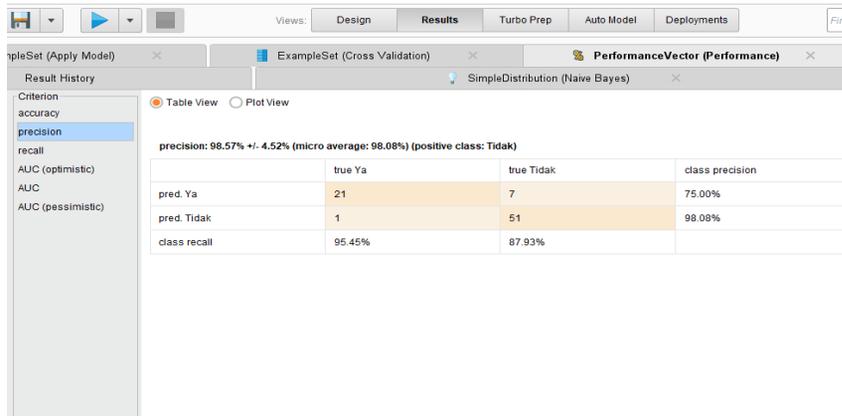
Gambar 5. Desain Pengujian Metode Naive Bayes

3.3 Akurasi dan Prediksi 10 Fold Cross Validation



Gambar 6. Hasil Accuracy

Hasil pengukuran data accuracy yang diperoleh dari data training mencapai 90.00%. Jumlah prediksi Ya yang diklasifikasikan sebagai true Ya (merupakan data benar yang terdeteksi benar) oleh classifier yaitu 21 data, dan jumlah prediksi Ya yang diklasifikasi sebagai true Tidak (merupakan data benar namun terdeteksi sebagai data salah) oleh classifier yaitu 7, dengan pencapaian class precision 75.00%.

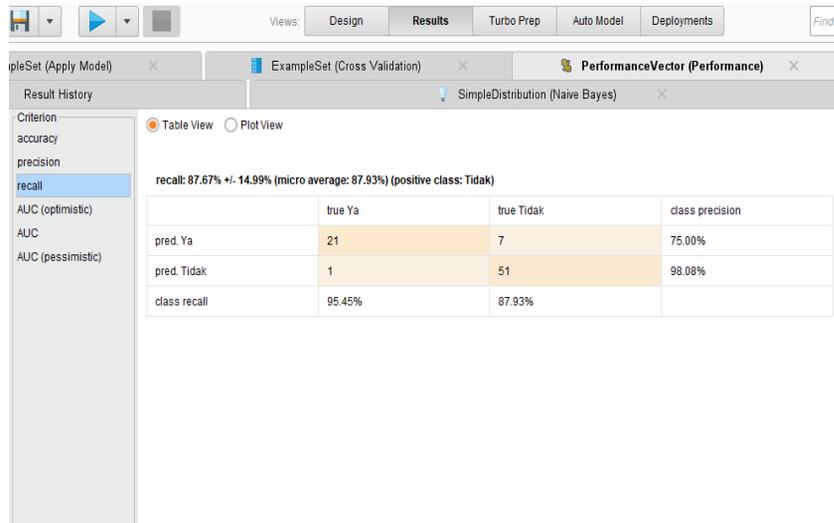


precision: 98.57% +/- 4.52% (micro average: 98.08%) (positive class: Tidak)

	true Ya	true Tidak	class precision
pred. Ya	21	7	75.00%
pred. Tidak	1	51	98.08%
class recall	95.45%	87.93%	

Gambar 7. Hasil Precision

Jumlah prediksi Tidak yang diklasifikasikan sebagai true Ya (merupakan data benar yang terdeteksi benar) oleh classifier yaitu 1 data, dan jumlah prediksi Tidak yang diklasifikasikan sebagai true Tidak (merupakan data benar namun terdeteksi sebagai data salah) oleh classifier yaitu 51 data, dengan pencapaian class precision 95.45%.

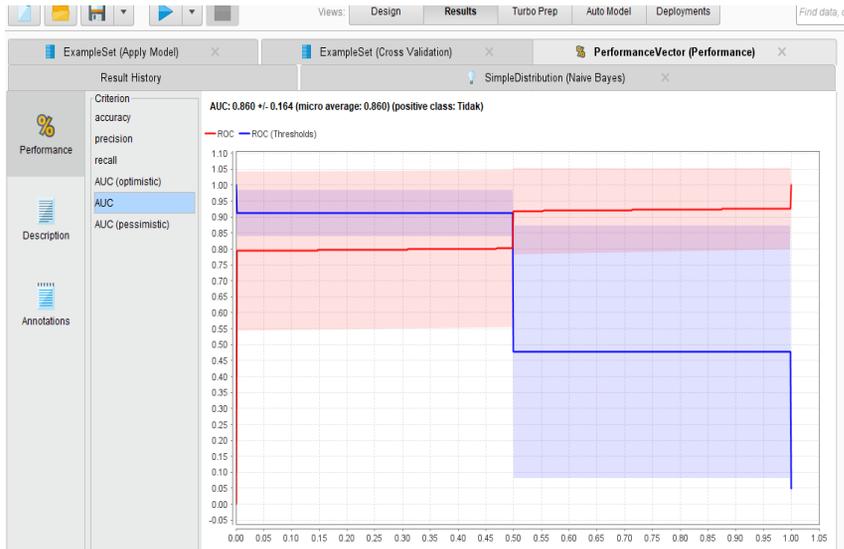


recall: 87.67% +/- 14.99% (micro average: 87.93%) (positive class: Tidak)

	true Ya	true Tidak	class precision
pred. Ya	21	7	75.00%
pred. Tidak	1	51	98.08%
class recall	95.45%	87.93%	

Gambar 8. Hasil Recall

Untuk class recall dengan true Ya (merupakan data benar yang terdeteksi benar) mencapai 95.45% sedangkan untuk class recall dengan true Tidak (merupakan data benar namun terdeteksi sebagai data salah) mencapai 87.93%.



Gambar 9. Diagram Kurva ROC/AUC

Tingkat keakurasian kurva ROC/AUC dapat diklasifikasikan menjadi 5 kelompok, yaitu [4]:

- 1) 0.90 - 1.00 = Excellent Clasification
- 2) 0.80 - 0.90 = Good Classification
- 3) 0.70 - 0.80 = Fair Classification
- 4) 0.60 - 0.70 = Poor Classification
- 5) 0.50 - 0.60 = Failure

Berdasarkan kriteria di atas menunjukan klasifikasi yang dihasilkan termasuk ke dalam kelompok Good Classification (klasifikasi baik) karena nilai AUC yang didapatkan dari pengujian berdasarkan kurva ROC menggunakan metode Naive Bayes sebesar 0.860.

Tabel 1. Analisa Kesalahan Prediksi

Nama	Layak PIP	Prediction Layak PIP
Nur Azizah	Tidak	Ya
Rehan Depriansyah	Tidak	Ya
Juhadi	Tidak	Ya
Deri Apriansyah	Tidak	Ya
Winda Sari	Tidak	Ya
M.Jofansyah	Ya	Tidak
Suci Rahayu	Tidak	Ya

Jumlah data yang diuji : 101

Jumlah data yang diprediksi benar : 94

Jumlah data yang diprediksi salah : 7

Performance Vector:

Accuracy: 90.00% +/- 9.86% (micro average: 90.00%)

Confusion Matrix:

True	Ya	Tidak
Ya	21	7
Tidak	1	51

Precision: 98.57% +/- 4.52% (micro average: 98.08%) (positive class: Tidak) Confusion Matrix:

True	Ya	Tidak
Ya	21	7
Tidak	1	51

Recall: 87.67% +/- 14.99% (micro average: 87.93%) (positive class: Tidak)

Confusion Matrix:

True	Ya	Tidak
Ya	21	7
Tidak	1	51

AUC (optimistic): 0.868 +/- 0.160 (micro average: 0.868) (positive class: Tidak)

AUC: 0.860 +/- 0.164 (micro average: 0.860) (positive class: Tidak)

AUC (pessimistic): 0.852 +/- 0.172 (micro average: 0.852) (positive class: Tidak)

4. Kesimpulan

Berdasarkan penelitian yang dilakukan, dapat ditarik kesimpulan bahwa, penerapan Data Mining menggunakan metode Naive Bayes untuk memprediksi kelayakan penerima bantuan PIP berdasarkan dataset dengan menambahkan atribut jumlah tanggungan sangat membantu dalam menentukan kelayakan penerima bantuan PIP.

Berdasarkan data yang telah diperoleh, proses penentuan penerima bantuan PIP menggunakan metode Naive Bayes menghasilkan informasi prediksi siswa yang layak mendapatkan bantuan PIP lebih akurat dibandingkan dengan penentuan yang dilakukan pihak sekolah. Dengan demikian metode Naive Bayes ini berhasil memprediksi dengan presentase accuracy sebesar 90.00%, precision 98.57%, recall 87.67% dan nilai AUC 0.868 dengan menggunakan data sebanyak 101.

DAFTAR PUSTAKA

- [1] Ningsih, S. R., Damanik, I. S., Gunawan, I., & Saputra, W., 2017, "Electre Dalammenentukan Penerima Program Indonesia Pintar (Pip) Melalui Kartu Indonesia Pintar (KIP) (Studi Kasus : Sd Swasta Al – Washliyah Moho Kabupaten Simalungun)", pp.264–275..
- [2] Han, J.,&Kamber, M., 2006, *Data Mining Concept and Tehniques*. San Fransisco: Morgan Kauffman.
- [3] Ningsih, E. W., & Hardiyani., 2020, "Penerapan Algoritma Naïve Bayes Dalam Penentuan Kelayakan Penerima Kartu Jakarta Pintar Plus", *Jurnal Teknik Komputer AMIK BSI*, VI(1), 15–20.
- [4] A. Andriani, 2013, "Sistem Pendukung Keputusan Berbasis Decision Tree Dalam Pemberian Beasiswa Studi Kasus : Amik „ Bsi Yogyakarta", *Semin. Nas. Teknol. Inf. dan Komun. 2013 (SENTIKA 2013)*, vol. 2013, no. Sentika, pp. 163–168, 2013.
- [5] Prasetyo, Eko., 2012, *Data Mining Konsep dan Aplikasi Menggunakan Matlab*. Yogyakarta: Andi.

- [6] Ningsih, E. W., & Hardiyanto, 2020, "Penerapan Algoritma Naïve Bayes Dalam Penentuan Kelayakan Penerima Kartu Jakarta Pintar Plus". *Jurnal Teknik Komputer AMIK BSI*, VI(1), 15–20.
- [7] Purnomo, B. H., 2011, "Metode dan teknik pengumpulan data dalam penelitian tindakan kelas (Classroom Action Research)", *Jurnal Pengembangan Pendidikan*, 8(1), 251–256.
- [8] A. B. Rahardjo, 2015, "Penerapan Data Mining Untuk Mengklasifikasi Penerima dan Bukan Penerima Kartu Identitas Miskin (KIM) Kelurahan Sumurrejo Gunungpati dengan Metode Naive Bayes Classifier," pp. 1–8.