KLASIFIKASI MENGGUNAKAN METODE NAIVE BAYES UNTUK MENENTUKAN CALON PENERIMA PIP

Angga Pebdika, Ruli Herdiana, Dodi Solihudin
Program Studi Komputerisasi Akuntansi STMIK IKMI Cirebon
Program Studi Teknik Informatika STMIK IKMI Cirebon
Jl. Perjuangan No. 10B Majasem Kec. Kesambi Kota Cirebon Tlp. 0231-490480-490481
anggarahman1502@gmail.com

ABSTRAK

Dalam dunia global saat ini, pendidikan sangat penting untuk lebih meningkatkan sumber daya manusia. Pendidikan akan membantu peserta didik mengenai pengembangkan sikap, keterampilan, serta kecerdasan intelektualnya untuk memberikan manusia yang terampil, cerdas, dan berakhlak mulia. Namun pendidikan seringkali tidak berjalan dengan baik, Ada beberapa faktor yang berkontribusi terhadap hal ini, Contoh yang paling menonjol adalah faktor ekonomi yang menyebabkan banyak anak putus sekolah. Oleh karena itu pemerintah membuat program agar masyarakat miskin dapat melanjutkan pendidikannya melalui program ini. Dalam penyusunan ini dicoba dengan memakai prosedur Naive Bayes. Metode ini merupakan metode mengklasifikasikan data satu atau lebih kategori yang telah diidentifikasi. Operasi Naive Bayes menggunakan perhitungan probabilitas dan statistik yang ditemukan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi probabilitas masa depan berdasarkan pengalaman masa lalu. Tujuan dari penelitian ini adalah untuk menghindari kesalahan dalam menentukan penerimaan bantuan. Perlu diterapkan data mining dengan algoritma Naive Bayes yang bisa mengklasifikasi tingkat kelayakan siswa penerima PIP, sehingga didapat hasil penerimaan program Indonesia Pintar yang lebih akurat. Penelitian ini memberikan informasi baru tentang hasil dari proses analisis yang dilakukan, selain itu dengan menggunakan metode Naive Bayes, proses analisis tersebut dapat membuat model kelayakan untuk menerima program Indonesia Pintar berdasarkan karakteristik yang telah ditentukan sebelumnya. Hasil dari proses penelitian ini diharapkan dapat menciptakan sistem data mining yang dapat memberikan hasil seleksi yang sangat akurat dalam memilih penerimaan PIP.

Kata kunci : *Data Mining,Naïve Bayes, Pendidikan,Program Indonesia Pintar*

1. PENDAHULUAN

Dalam kehidupan pendidikan memiliki peranan yang sangat penting, Pendidikan juga dapat memajukan dan mengembangkan kepribadian seseorang secara mental dan fisik. Itulah sebabnya pendidikan berakibat positif untuk kita, serta pembelajaran pula bisa menyingkirkan buta huruf serta mengarahkan kepiawaian, keterampilan mental, serta lain sebagainya. Namun pendidikan seringkali tidak berjalan dengan baik sebab beberapa faktor problematis menjadi penyebab putus sekolah, seringkali alasan finansial menjadi alasan utama putus sekolah.

Dalam menyikapi permasalahan perekonomian yang membuat banyak anak putus sekolah, pemerintah meluncurkan program Indonesia Pintar. Pada tahun 2015 Kebijakan Program Indonesia Pintar (PIP) didirikan lewat Permendikbud No. 12 Tahun 2015 tanggal 12 Mei 2015. Program Indonesia Pintar ialah program pemerintah yang ditawarkan dalam wujud pembiayaan pendidikan. langsung kepada para siswa. Program ini diperuntukkan bagi individu dari keluarga kurang sanggup secara ekonomi yang ditandai menggunakan Kartu Indonesia Pintar (KIP) [1].

Menurut penelitian terdahulu Kajian oleh Okta Rin dan Suzi Oktavia Kunang (2021) berjudul "Implementasi Data Mining Menggunakan Metode Naive Bayes Untuk Penentuan Penerima Bantuan Program Indonesia Pintar (PIP) (Studi Kasus : Sd Negeri 9 Air Kumbang)". Berdasarkan uraian di atas, Implementasi data mining dengan prosedur Naive Bayes guna memprediksi kelayakan penerima PIP bersumber pada penambahan atribut dataset "Jumlah tanggungan" sangat berguna untuk memastikan kelayakan penerima PIP. Bersumber pada informasi yang diperoleh, penentuan penerima PIP memakai prosedur algoritma Naive Bayes memberikan informasi prognostik yang lebih akurat tentang siswa yang berhak menerima dukungan PIP ketimbang dengan penentuan yang dicoba oleh sekolah. Dengan demikian, metode Naive Bayes berhasil melakukan prediksi dengan presisi 90,00%, presisi 98,57%, recall 87,67%, dan AUC 0,868 menggunakan 101 data [2].

Data Mining adalah cara untuk menemukan informasi yang berharga, laten, tersembunyi dari sejumlah besar data (database) dan dengan demikian menemukan pola yang menarik, yang sebelumnya tidak diketahui. Data mining adalah sekumpulan proses untuk memastikan korelasi pola dengan tujuan memfilter data yang besar dan mendapatkan fitur terbaru yang berguna untuk memahami suatu model atau varian dari suatu database.

Metode Naïve Bayes merupakan metode yang hanya memerlukan jumlah data latih (training data)

yang kecil untuk memilih parameter yang dibutuhkan selama proses klasifikasian. Dengan menggunakan beberapa atribut Dengan atribut saling berhubungan untuk penentuan kelayakan [3].

2. TINJAUAN PUSTAKA.

2.1. Program Indonesia Pintar

Program Indonesia Pintar (PIP) adalah Sebuah program pemerintah yang diprakarsai oleh Kementerian Pendidikan dan Kebudayaan yang dibentuk dengan tujuan mengatasi permasalahan pengajaran sekolah yang ada, dimana masih banyak masalah siswa putus sekolah karena kesulitan keuangan. Siswa dari keluarga miskin membutuhkan program Indonesia Pintar karena siswa dari keluarga miskin sangat rawan berhenti sekolah di usia dini. Hal seperti ini diakibatkan karena keadaan ekonomi keluarga peserta didik yang kurang mendukung, sehingga kebanyakan peserta didik menetapkan untuk berhenti sekolah serta memilih untuk membantu prekonomian keluarga dengan bekerja. Atas dasar masalah ini, pemerintah mengambil tindakan dalam upaya pemecahan masalah supaya peserta didik yang berasal dari keluarga kurang mampu bisa menuntaskan pendidikannya serta dapat meneruskan persekolahan ke jenjang yang lebih teratas [4]

Kebijakan program Indonesia Pintar melalui Kartu Indonesia Pintar (KIP) ditetapkan pemerintah di bawah Kementerian Pendidikan dan Kebudayaan (Kemendikbud) lewat Tim Percepatan Penanggulangan Kemiskinan Nasional (TNP2K). Program ini bertujuan agar memberikan pendidikan yang layak bagi siswa miskin, mencegah anak putus sekolah dan penuhi kebutuhan dalam pembelajaran. Siswa ditujukan menggunakan bantuan ini untuk memenuhi kebutuhan sekolah seperti: Biaya SPP, perlengkapan sekolah dan uang saku. Dengan adanya Kartu Indonesia Pintar, diharapkan siswa tidak putus sekolah lagi karena kekurangan dana. Bantuan Kartu Indonesia Pintar (KIP) ditujukan pada peserta didik yang kekurangan dari tingkat SD sampai dengan SMA. Salah satu gejala yang muncul ialah pemerataan pendidikan serta akurasi keselarasan berbasis Program Indonesia Pintar (PIP) melalui Kartu Indonesia Pintar (KIP) tidak tepat sasaran. Hal ini dibuktikan dengan masih adanya siswa dari keluarga mampu dengan terdaftar sebagai penerima dana Kartu Indonesia Pintar (KIP), dan terdapat peserta didik yang tergolong kekurangan dan bukan pemeroleh bantuan Indonesia Pintar yang terdaftar [5]**.**

2.2. Data Mining

Data mining merupakan salah satu teknik penggolongan data dengan tujuan mencari kolerasi antar data yang tidak diketahui oleh pengguna dan menyajikannya dalam hasil yang mudah dipahami dan kolerasi data tercatat dapat digunakan sebagai dasar pengambilan keputusan. Penambangan data dapat terbagi dari beberapa kelompok seperti Deskripsi, Evaluasi, Prediksi, Klasifikasi, Pengelompokan dan Pemetaan berdasarkan tugas yang akan dilakukan [6].

Data mining adalah suatu proses dimana sejumlah besar data digali dan dianalisis untuk mendapatkan suatu yang valid, baru dan bermanfaat, alhasil bisa menemukan pola atau formula dalam data tersebut. Penambangan data secara umum dapat diklasifikasikan menjadi dua kategori utama, penambangan deskriptif dan penambangan prediktif. Deskriptif Mining adalah proses menemukan informasi penting dari database. Sedangkan prediktif adalah Proses penentuan polaritas data dengan menentapkan beberapa variabel untuk memperkirakan variabel lain di masa mendatang.

Data Mining ialah bagian integral dari pengetahuan pada basis data atau sering disebut dengan Knowledge Discovery in Databases (KDD), yang berarti progres umum transformasi data mentah menjadi model yang berguna, yaitu informasi yang dibutuhkan pengguna sebagai informasi. [7]. Dibawah ini ialah proses terjadinya dalam sebuah KDD :

1. Data selection
   Data selection berasal dari sekumpulan data oprasional yang harus dilakukan sebelum tahap ekstraksi data pada KDD. Hasil Pemilihan data yang akan digunakan dalam tahap data mining ditampilkan dalam satu halaman berdasarkan data operasional.
2. Preprocessing
   Sebelum tahap selanjutnya dilakukan, Diperlukan metode pembersihan untuk data yang menjadi fokus KDD. Dalam tahapan ini mencangkup penghapusan duplikat data, pengecekan data yang tidak konsisten serta koreksi kesalahan dalam data seperti kekeliruan dalam cetak.
3. Transformation
   Coding merupakan transformasi data yang dipilih agar cocok untuk langkah penambangan data. Coding KDD adalah langkah kreatif yang sangat bergantung oleh jenis data atau model yang diambil dari setdata.
4. Data mining
   Dalam langkah ini ialah digunakannya teknik atau metode dengan tujuan menemukan pola atau hasil yang menarik. Teknik, metode, atau algoritma penambangan data yang bervariasi, menentukan metode atau algoritma yang tepat sangat bergantung dalam tujuan serta keseluruhan tahapan dalam KDD.
5. Interpretation/Evaluation
   Model data yang diperoleh dari langkah data mining dapat memberikan hasil yang dapat mudah dimengerti daripada yang berkepentingan. Tujuan dari bagian ini adalah untuk menentukan apakah kebijakan atau informasi yang diperoleh bertentangan dengan informasi.

Knowledge Discovery in Databases (KDD) ialah aktivitas yang mencakup serta memakai data historis dengan tujuan mendapatkan keteraturan, pola serta korelasi pada perpaduan data yang besar [8]. Dalam jurnalnya disampaikan ciri-ciri data mining sebagai berikut :

a. Data Mining mengacu pada inovasi tersembunyi serta model data spesifik yg sebelumnya tidak diketahui.
b. Pada umumnya data mining menggunakan big data.
c. Secara umum, big data digunakan dengan tujuan membentuk hasil yang kredibel.
d. Data mining bermanfaat bagi pengambilan ketentuan krusial terutama pada strategi

2.3. Naïve Bayes

Naive Bayes classifier (NBC) ialah salah satu metode klasifikasi yang sering digunakan. Naive Bayes Classifier (NBC) adalah salah satu metode klasifikasi dan statistik pengklasifikasi yang dapat memprediksi peluang untuk menjadi anggota kelas. NBC berdasarkan pada teori Bayesian memperkirakan bahwasanya nilai atribut tidak bergantung dengan nilai yang lainnya. Keunggulan NBC ialah sederhana namun mempunyai akurasi yang tinggi. Proses klasifikasi data terdiri dari dua langkah. Langkah pertama ialah pelatihan dengan contoh (model training). Langkah kedua ialah mengklasifikasikan data yang kategorinya tidak diketahui [7].

Naive Bayes adalah teknik peramalan probabilistik sederhana sesuai dengan pelaksanaan teorema Bayes (hukum bayes) menggunakan perkiraan independensi yang kuat. Kelebihan dari pada metode ini ialah hanya memerlukan sedikit data latih dengan tujuan memastikan parameter yang diharapkan pada langkah klasifikasi [2].

Kelebihan metode ini ialah :

1. Membenahi kuantitatif serta diskrit
2. Tangguh untuk titik kebisingan terisolasi, seperti yang dirata-rata saat mengevaluasi probabilitas bersyarat data.
3. Hanya dengan jumlah kecil data training untuk memperkirakan parameter (rata – rata serta variansi yang berasal dari variabel) untuk klasifikasi.
4. Skala nilai yang hilang disebabkan oleh pengabaian kejadian selama estimasi kemungkinan peluang.
5. Cepat serta menghemat waktu
6. Kuat terhadap karakteristik yang tidak relevan.

Kekurangan Metode Naïve Bayes :

1. Memperkirakan bahwa karakteristiknya bebas.
2. Ini tidak berlaku, Jika probabilitas bersyarat adalah 0, probabilitas prediksi juga akan menjadi 0.

Pendekatan Naïve Bayes digunakan untuk analisis data. Ini adalah pengklasifikasi probabilitas dasar yang menggunakan Teorema Bayes dengan asumsi independensi yang tinggi. Pendekatan Nave Bayes memiliki keuntungan karena hanya menggunakan sedikit data training untuk mendapatkan estimasi parameter yang diperlukan dalam proses klasifikasi. Oleh karena itu, terlepas dari variabel independen yang dipertimbangkan, hanya varian dari variabel kelas yang diperlukan untuk menentukan kategorisasinya, bukan seluruh variabel kelas. [9]

2.4. Klasifikasi

Klasifikasi adalah salah satu langkah terpenting pada data mining. Klasifikasi ialah proses mengkategorikan atau memberi label data atau objek baru berdasarkan kualitas tertentu. Teknik klasifikasi melibatkan pemeriksaan variabel yang dihasilkan dari data yang ada. Tujuan dari klasifikasi adalah untuk mengantisipasi kelas yang dihasilkan dari item yang tidak diketahui. Ada tiga tahap klasifikasi yaitu konstruksi model, aplikasi model serta evaluasi. Pembuatan model melibatkan pembuatan contoh memakai data pelatihan yang telah memiliki atribut serta kelas. informasi ini lalu digunakan untuk memilih kelas data atau objek baru. Data tersebut kemudian dievaluasi untuk melihat keakuratan yang didapatkan dari pengembangan dan penerapan model pada data baru [10].

Salah satu metode classifier yang dapat digunakan adalah metode Naive Bayes yang sering disebut dengan Naive Bayes Classifier (NBC). Naive Bayes Classifier (NBC) adalah salah satu metode klasifikasi dan statistik pengklasifikasi yang dapat memprediksi peluang untuk menjadi anggota kelas. Menurut NBC, nilai atribut grup tidak bergantung pada nilai karakteristik lainnya. Keunggulan NBC ialah dasar namun akurat. Prosedur kategorisasi data dibagi menjadi dua tahap. Langkah pertama adalah berlatih dengan contoh (Training Example). Tahap kedua adalah proses pengkategorian data yang belum diketahui kelasnya [7].

3. METODE PENELITIAN

Data penelitian dipisahkan menjadi dua kategori ialah data primer serta data sekunder. Data primer adalah informasi yang dikumpulkan langsung di lapangan, sedangkan data sekunder dikumpulkan secara tidak langsung.

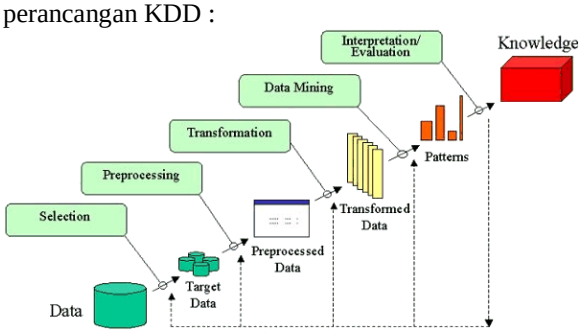
3.1. Teknik Pengumpulan Data

1. Obervasi
   Metode observasi dilakukan dengan pengambilan data calon penerima program indonesia pintar di SMP 1 Negeri Kramatmulya.
2. Metode wawancara
   Hal ini dilakukan melalui wawancara dengan wakil kepala sekolah dan pengurus sekolah untuk

mendapatkan informasi tentang data calon penerima manfaat program Indonesia Pintar.

3.2. Tahap Perancangan

Dalam penelitian ini menggunakan tahap perancangan KDD :

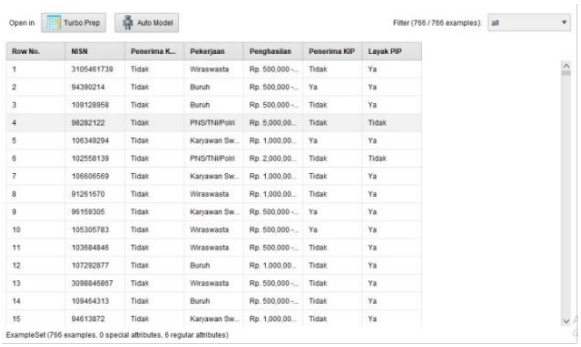


Gambar 1. Tahapan KDD

1. Data selection
   Data selection berasal dari sekumpulan data oprasional yang harus dilakukan sebelum tahap ekstraksi data pada KDD. Hasil Pemilihan data yang akan digunakan dalam tahap data mining ditampilkan dalam satu halaman berdasarkan data operasional.
2. Preprocessing
   Sebelum tahap selanjutnya dilakukan, Diperlukan metode pembersihan untuk data yang menjadi fokus KDD. Dalam tahapan ini mencangkup penghapusan duplikat data, pengecekan data yang tidak konsisten serta koreksi kesalahan dalam data seperti kekeliruan dalam cetak.
3. Transformation
   Coding merupakan transformasi data yang dipilih agar cocok untuk langkah penambangan data. Coding KDD adalah langkah kreatif yang sangat bergantung oleh jenis data atau model yang diambil dari setdata.
4. Data mining
   Dalam langkah ini ialah digunakannya teknik atau metode dengan tujuan menemukan pola atau hasil yang menarik. Teknik, metode, atau algoritma penambangan data yang bervariasi, menentukan metode atau algoritma yang tepat sangat bergantung dalam tujuan serta keseluruhan tahapan dalam KDD.
5. Interpretation/Evaluation
   Model data yang diperoleh dari langkah data mining dapat memberikan hasil yang dapat mudah dimengerti daripada yang berkepentingan. Tujuan dari bagian ini adalah untuk menentukan apakah kebijakan atau informasi yang diperoleh bertentangan dengan informasi

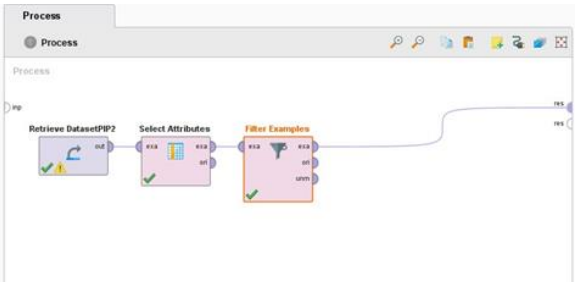
4. HASIL DAN PEMBAHASAN

1. Data Selection
   Data selection mencakup semua atribut dari kumpulan data asli, dipilih untuk mendapatkan atribut yang diperlukan untuk proses penambangan data selanjutnya . Berikut ini menjelaskan proses dimana tolos Rapidminer memilih atribut mana yang akan digunakan dalam tahapan data mining. Pada tahap ini adalah menyeleksi atribut mana saja yang dibutuhkan menggunakan oprator select attributes yang dapat memfilter atribut yang diperlukan. Dengan parameter subset, opsi ini memilih beberapa atribut yang akan digunakan. Jika metadata diketahui, semua atribut dicantumkan dan atribut yang diperlukan dapat dipilih dengan mudah. Berikut dibawah ini adalah hasil seleksi atribut.



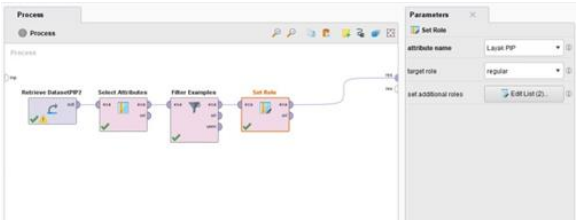
Gambar 2. Hasil Seleksi Atribut

2. Preprocessing
   Dalam tahap ini mencakup hapus duplikat data, periksa data yang tidak konsisten dan perbaiki kesalahan dalam data. Pada langkah ini penulis menggunakan filter sampel rapidminer untuk menghilangkan missing value, data missing value tersebut mengandung dua atribut yaitu pekerjaan dan pendapatan. Berikut adalah proses penghapusan pada rapidminer menggunakan oprator filter examples.



Gambar 3. Penggunaan Oprator Filter Example

3. Tranforsformasi
   Pada tahapan ini, mengubah data menjadi model analisis data dan memodelkan data agar sesuai dengan analisis data mining yang diharapkan, tujuan transformasi adalah mengubah data yang dipilih ke dalam bentuk prosedur penambangan. Mengubah kode NISN menjadi ID dan kode PIP menjadi Label. Atribut dalam peran label berperan sebagai label dan atribut berlabel sebagai operator pembelajaran, label sering disebut sebagai variabel atau kelas. Berikut data yang diolah untuk data mining.



Gambar 4. Penggunaan Operator Set Role

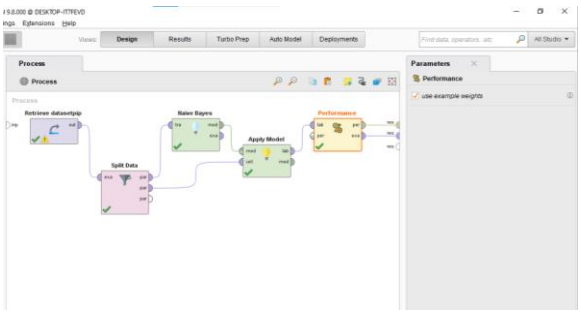
Set-Role merupakan operator yang mengklasifikasikan atribut menjadi atribut spesifik atau atribut standar . Operator Set Role memisahkan baris yang mengambil atribut koordinat serta prediksi posisi yang diberikan ke kelas label. Dibawah ini merupakan hasil transformation data dengan mengganti NISN sebagai ID serta Layak PIP sebagai label.

Row No.	NISN	Layak PIP	Penerima K...	Pekerjaan	Penghasilan	Penerima KIP
1	3105461739	Ya	Tidak	Wiraswasta	Rp. 500,000 -...	Tidak
2	94390214	Ya	Tidak	Buruh	Rp. 500,000 -...	Ya
3	109128958	Ya	Tidak	Buruh	Rp. 500,000 -...	Tidak
4	98282122	Tidak	Tidak	PNS/TNI/Polri	Rp. 5,000,00...	Tidak
5	106349294	Ya	Tidak	Karyawan Sw...	Rp. 1,000,00...	Ya
6	102558139	Tidak	Tidak	PNS/TNI/Polri	Rp. 2,000,00...	Tidak
7	106606569	Ya	Tidak	Karyawan Sw...	Rp. 1,000,00...	Tidak
8	91261670	Ya	Tidak	Wiraswasta	Rp. 1,000,00...	Tidak
9	96159305	Ya	Tidak	Karyawan Sw...	Rp. 500,000 -...	Ya
10	105305783	Ya	Tidak	Wiraswasta	Rp. 500,000 -...	Ya
11	103684846	Ya	Tidak	Wiraswasta	Rp. 500,000 -...	Tidak
12	107292877	Ya	Tidak	Buruh	Rp. 1,000,00...	Tidak
13	3098846867	Ya	Tidak	Wiraswasta	Rp. 500,000 -...	Tidak
14	109464313	Ya	Tidak	Buruh	Rp. 500,000 -...	Tidak
15	94613872	Ya	Tidak	Karyawan Sw...	Rp. 1,000,00...	Tidak

ExampleSet (763 examples, 2 special attributes, 4 regular attributes)

Gambar 5. Hasil Transformasi Data

4. Data Mining
   Pada tahapan data mining ini ialah menggunakan algoritma naïve bayes yang berfungsi memecahkan masalah klasifikasi. aplikasi yang digunakan pada proses ini ialah rapidminer versi 9.8. Dengan menggunakan operator Retrieve, Split Data, Naive Bayes, Apply Model dan Perpormance. Pada gambar dibawah ini ialah proses Data mining dengan memakai algoritma naïve bayes untuk mengklasifikasikan tingkat kelayakan peserta didik calon penerima Program Indonesia Pintar.



Gambar 6. Proses Algoritma Naïve Bayes

Berdasarkan gambar di atas, Operator Split Data digunakan untuk membagi data menjadi data latih dan data uji. pada subproses training menggunakan operator naive bayes sebagai penerapan model algoritme. sementara operator Apply model terapkan menggunakan subproses Pengujian untuk menguji data yang dihasilkan dari operator Naive Bayes, operator Kinerja dipergunakan untuk mengevaluasi kinerja model, secara otomatis menyampaikan daftar nilai kriteria kinerja yang disediakan untuk tugas tertentu.

5. Interpretation/Evaluation
   Pada fase ini dilakukan proses evaluasi hasil data mining untuk memudahkan pendataan , salah satunya adalah penggunaan Confusion Matrix. Confusion matrix ialah alat untuk menganalisis seberapa baik sebuah klasifikasi. True positive serta true negative menyampaikan hasil ketika klasifikasi benar, sedangkan false positive serta false negative menyampaikan hasil ketika klasifikasi salah.

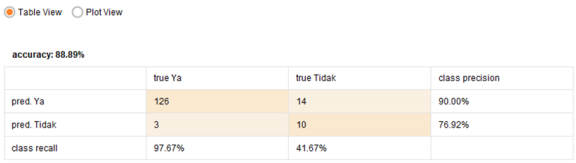


Table View Plot View

accuracy: 88.89%

	true Ya	true Tidak	class precision
pred. Ya	126	14	90.00%
pred. Tidak	3	10	76.92%
class recall	97.67%	41.67%	

Gambar 7. Hasil Klasifikasi Naïve Bayes

Dari hasil klasifikasi algoritma naive bayes didapatkan hasil akurasi mencapai 88,89%. Lebih spesifiknya, bilangan true positive (TP) 126, true negative (TN) 10, false positive (FP) 14, dan false negative (FN) 3. True Yadan true Tidak adalah kelas asli atau nilai real. prediksi YA dan prediksi tidak adalah kelas atau nilai prediktif. Class Precision ialah class yang mengukur tingkat presisi antara informasi yang diminta oleh pengguna dengan hasil prediksi yang diberikan oleh sistem. Class reccal ialah kelas yang mengevaluasi hasil yang diberikan oleh sistem saat memprediksi informasi.

Algoritma Naïve bayes, bisa dihitung dengan menggunakan rumus sebagai berikut :

1. Accuracy

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} x\ 100\%$$

$$Accuracy = \frac{126 + 10}{126 + 10 + 14 + 3}\ x\ 100\%$$

$$Accuracy\ = 0.8888 = 88.89\%$$

Dari perhitungan tersebut, nilai akurasi hasil pengujian klasifikasi menggunakan algoritma Naive Bayes adalah sebesar 88.89%.

2. Precision

$$Precision = \frac{TP}{TP + FP} x\ 100\%$$

$$Precision = \frac{126}{126 + 14} x\ 100\%$$

$$precision = 0.9 = 90.00\%$$

Dari perhitungan tersebut, nilai precision hasil pengujian klasifikasi Algoritma Naive Bayes memiliki tingkat keberhasilan 90,00%.

3. Recal

$$Recall = \frac{TP}{TP + FP} x\ 100\%$$

$$Recall = \frac{126}{126 + 3} x\ 100\%$$

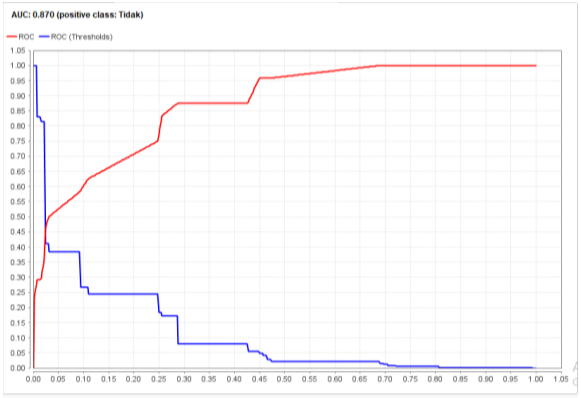
$$\text{Recall} = 0.9767 = 97.67\%$$

Dari perhitungan tersebut, nilai recali hasil pengujian klasifikasi Algoritma Naive Bayes memiliki tingkat keberhasilan 97.00%

4. AUC

Dari analisis data diatas menggunakan software Rapid Miner dengan pengukuran Naive Bayes, hasil AUC-0.807 masuk dalam kategori baik (klasifikasi baik). Tingkat akurasi AUC dibagi menjadi lima bagian, yaitu :

- 0.90 – 1.00 = Excellent Classification
- 0.80 – 0.90 = Good Classification
- 0.70 – 0.80 = Fair Classification
- 0.60 – 0.70 = Poor Classification
- 0.50 – 0.60 = Failure



Gambar 8. Hasil AUC

5. Perpormance Vector

PerformanceVector accuracy 88.89%

Confusion Matrix:

True	Ya	Tidak
Ya	126	14
Tidak	3	10

Precision: 76.92% (positive class: Tidak)

ConfusionMatrix:

True	Ya	Tidak
Ya	126	14
Tidak	3	10

recall: 41.67% (positive class: Tidak)

ConfusionMatrix:

True	Ya	Tidak
Ya	126	14
Tidak	3	10

AUC (optimistic): 0.889 (positive class: Tidak)
AUC: 0.807 (positive class: Tidak)
AUC (pessimistic): 0.851 (positive class: Tidak)

5. KESIMPULAN DAN SARAN

Kesimpulan berikut dapat dibentuk berdasarkan penjelasan yang diberikan : Data mining dengan metode naive bayes dapat memprediksi tingkat kelayakan pengguna bantuan PIP berdasarkan data set yang sudah dikumpulkan dengan menggunakan bantuan tolls rapidminer serta menggunakan beberapa operator diantaranya Oprator Retrieve, Split Data, Naïve Bayes, Apply Model, dan Performance. Berdasarkan hasil penelitian serta pengujian dengan teknik data mining memakai algoritma klasifikasi naive bayes KDD maka nilai akurasi keseluruhan 88.89% dan Class recall YA 97.67%, Class recall tidak 41.67%, Class precision YA 90.00% dan Class precision tidak 76.92%.

DAFTAR PUSTAKA

[1] Herlinawati, E. Heriyati, Sudiyono, and A. B. Susanto, *Kajian Program Indonesia Pintar (PIP): Strategi Penjangkauan Anak Tidak Sekolah (ATS) Untuk Mengikuti Pendidikan Melalui Program Indonesia Pintar (PIP)*. 2018.

[2] O. Rini and S. O. Kunang, "Implementasi Data Mining Menggunakan Metode Naive Bayes Untuk Penentuan Penerima Bantuan Program Indonesia Pintar (Pip) (Studi Kasus : Sd Negeri 9 Air Kumbang)," *Bina Darma Conf. Comput. Sci.*, vol. 3, no. 4, pp. 714–722, 2021, [Online]. Available: https://conference.binadarma.ac.id/index.php/BDCCS/article/view/2450

[3] A. Saleh, "Implementasi Naive Bayes," *J. Informatics, Inf. Syst. Softw. Eng. Appl.*, vol. 2, no. 3, pp. 207–2017, 2015, doi: 10.24076/citec.2015v2i3.49.

[4] Yudi Agusman, "Public Inspiration : Jurnal Administrasi Publik Implementasi Program Indonesia Pintar di Sekolah Dasar Negeri 1 Kolakaasi Kabupaten Kolaka," vol. 4, no. 2, pp. 105–113, 2019, doi: 10.22225/pi.4.2.2019.105-113.

[5] N. E. Rohaeni and O. Saryono, "Implementasi Kebijakan Program Indonesia Pintar (PIP) Melalui Kartu Indonesia Pintar (KIP) dalam

Upaya Pemerataan Pendidikan," *J. Educ. Manag. Adm. Rev.*, vol. 2, no. 1, pp. 193–204, 2018, doi: 10.4321/ijemar.v2i1.1824.

[6] A. Syafii, G. Dwilestari, and A. Ajiz, "KOMPARASI ALGORITMA NAÏVE BAYES DAN ALGORITMA C4.5 DALAM KLASIFIKASI PELANGGAN PRODUK INDIHOME," *JURSIMA J. Sist. Inf. dan Manaj.*, vol. 10, no. 2, pp. 60–70, 2022, doi: 10.47024/js.v10i2.414.

[7] S. Adi, "Implementasi Algoritma Naïve Bayes Classifier Untuk Klasifikasi Penerima Beasiswa PPA Di Universitas Amikom Yogyakarta," *J. Mantik Penusa*, vol. 22, no. 1, pp. 11–16, 2018, [Online]. Available: https://e-jurnal.pelitanusantara.ac.id/index.php/mantik/article/view/342

[8] R. Saputra and A. J. P. Sibarani, "Implementasi Data Mining Menggunakan Algoritma Apriori Untuk Meningkatkan Pola Penjualan Obat," vol. 7, no. 2, 2020, doi: 10.35957/jatisi.v7i2.195.

[9] E. W. Ningsih and Hardiyan, "Penerapan Algoritma Naïve Bayes Dalam Penentuan Kelayakan Penerima Kartu Jakarta Pintar Plus," *J. Tek. Komput. AMIK BSI*, vol. 6, no. 1, pp. 15–20, 2020, doi: 10.31294/jtk.v6i1.6680.

[10] D. A. Nasution, H. H. Khotimah, and N. Chamidah, "Perbandingan Normalisasi Data untuk Klasifikasi Wine Menggunakan Algoritma K-NN," *Comput. Eng. Sci. Syst. J.*, vol. 4, no. 1, pp. 78–82, 2019, doi: 10.24114/cess.v4i1.11458.