



KLASIFIKASI PENERIMAAN PESERTA DIDIK BARU MENGGUNAKAN ALGORITMA NAÏVE BAYES DENGAN SMOTE PADA SMKN 1 JAMBLANG

Diding Herudin^{1)*}, Ahmad Faqih²⁾, Agus Bahtiar³⁾

^{1,2,3)} Teknik Informatika, STMik IKMI Cirebon, Jawa Barat
email: didingherudin8@gmail.com¹⁾, faqichkhan@gmail.com²⁾,
agusbahtiar038@gmail.com³⁾

Abstrak

Setiap institusi pendidikan mempunyai cara masing-masing dalam proses penerimaan peserta didik baru. Agar penerimaan peserta didik baru efektif dan efisien maka perlu dilakukan analisis data penerimaan peserta didik baru. Teknik data mining merupakan salah satu cara yang dapat dilakukan untuk mengklasifikasikan data peserta didik baru. Masalah class imbalance biasanya terjadi ketika melakukan klasifikasi, dimana sebuah classifier cenderung mengklasifikasikan class mayoritas dan mengabaikan class minoritas. Untuk mengatasi permasalahan tersebut dapat digunakan dua pendekatan yaitu, pendekatan sample dan algoritma. dalam mengatasi permasalahan data tidak seimbang pada data penerimaan peserta didik baru SMKN 1 Jamblang menggunakan pendekatan Teknik sampling. Teknik sampling yang umumnya digunakan dalam mengatasi permasalahan class imbalance yaitu Over-sampling, Under-sampling, dan kombinasi keduanya. Penyelesaian class imbalance dilihat berdasarkan akurasi, sensitivity, dan specificity dengan metode klasifikasi Naive Bayes yang dikombinasikan algoritma SMOTE pada dataset penerimaan peserta didik baru di SMKN 1 Jamblang. Hasil penelitian yang dilakukan di SMKN 1 Jamblang terhadap 638 data penerimaan peserta didik baru di SMKN 1 Jamblang terklasifikasikan menjadi dua kelas yaitu 274 kelas diterima dan 265 kelas tidak diterima dengan hasil akurasi 92,58%. Hasil penyelesaian class imbalance terhadap 638 data peserta didik baru di SMKN 1 Jamblang menggunakan algoritma Naive Bayes yang dikombinasikan dengan algoritma SMOTE menghasilkan akurasi 92.58%.

Kata Kunci : Data Mining, Algoritma, Naive Bayes, Klasifikasi, SMOTE.

Abstract

Each educational institution has its own way in processes acceptance of new learners. In order for the acceptance of new didi participants effectively and efficiently, it is necessary to analyse the data of new learners' admissions. Data mining techniques are one way that can be done to classify new learners' data. Class imbalance problems usually occur when classifying, where a classifier tends to classify the majority class and ignores the minority class. To overcome this problem can be used two approaches, namely, the sample approach and the algorithm. in overcoming the problem of unbalanced data on the admission data of new learners SMKN1 Jamblang using the sampling technique approach. Sampling techniques that are commonly used in overcoming class imbalance problems are over-sampling, under-sampling, and a combination of both. The completion of class imbalance is seen based on accuracy, sensitivity, and specificity with naive bayes classification method combined with SMOTE algorithm in the new student admission dataset in SMKN 1 Jamblang.



The results of the study conducted at SMKN1 Jamblang on 638 new student admission data in SMKN 1 Jamblang were classified into two classes, namely 274 accepted classes and 265 classes were not received with an accuracy of 92.58%. The results of the completion of class imbalance against 638 new student data in SMK N 1 Jamblang using Naïve Bayes algorithm combined with SMOTE algorithm resulted in 92.58% accuracy..

Keywords: Data Mining, Algorithms, Naïve Bayes, Classification, Smote.

PENDAHULUAN

Pendidikan merupakan suatu hal yang menjadi kebutuhan dasar bagi masyarakat. Baik dan buruknya kualitas hidup masyarakat dan negara tentu tidak terlepas dari kualitas pendidikan yang ada pada sebuah wilayah tersebut. PPDB merupakan proses seleksi yang harus dilalui oleh peserta didik ketika akan memasuki suatu jenjang pendidikan yang baru. Penerimaan peserta didik baru adalah suatu proses seleksi di bidang akademis yang harus dilalui oleh calon peserta didik dalam menempuh jenjang pendidikan yang lebih tinggi.[1]

Manajemen peserta didik sangat penting dilakukan karena peserta didik merupakan objek dapat terlaksananya sebuah proses belajar mengajar dan proses transfer ilmu. Penerimaan Peserta Didik Baru (PPDB) merupakan sebuah langkah awal bagi sekolah untuk meningkatkan kualitas pendidikan. Dikatakan seperti itu karena jika tidak ada peserta didik maka tidak akan terjadi sebuah proses pembelajaran.[2]

Klasifikasi yang merupakan salah satu permasalahan dalam data mining yang diselesaikan dengan mempergunakan metode supervised learning sangat bergantung pada data latih. Data latih itu sendiri, jumlah distribusi data untuk masing-masing kelas sangat jarang memiliki jumlah yang sama. Dalam kondisi nyata, sangat sering ditemui, jumlah dataset masing-masing kelas berbeda. Kondisi ini

disebut dengan imbalanced data (ketidakseimbangan data). Ketidakseimbangan jumlah data pada masing-masing kelas dalam dataset yang digunakan untuk membuat model klasifikasi dapat menghasilkan model yang memiliki performa kurang baik karena model tersebut hanya mampu memprediksi kelas mayoritas dengan baik namun tidak dapat memprediksi kelas minoritas. Metode SMOTE dapat digunakan untuk menyeimbangkan data tersebut. Hasilnya adalah model klasifikasi yang telah diseimbangkan dengan metode SMOTE memiliki performa lebih baik dibandingkan jika tidak dilakukan penyeimbangan data.[3]

Penerimaan peserta didik baru yang ada di sekolah menengah kejuruan SMK Hasanah Pekanbaru, jalur yang digunakan dalam proses penerimaan peserta didik baru yaitu jalur regular, mandiri dan beasiswa. Metode Klasifikasi algoritma C4.5 digunakan pada penelitian ini untuk memprediksi penerimaan peserta didik baru pada SMK Hasanah. Dari hasil pohon keputusan dan model aturan bahwa jika calon peserta didik baru nilai tes tinggi maka calon peserta didik baru tersebut diprediksi diterima tanpa melihat asal daerah, pekerjaan orang tua dan jurusan.[4]

SMK Negeri 1 Jamblang sebagai lembaga penyedia jasa Pendidikan dalam memberikan layanan Pendidikan berupaya memberikan layanan terhadap peserta didik baru untuk meningkatkan kepuasan terhadap layanan. Pengelolaan pendidikan

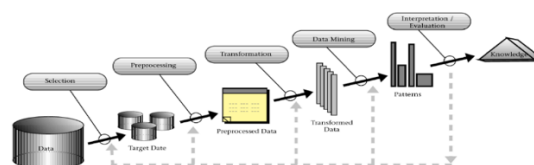


menjadi sangat penting, dimana pertumbuhan dan perkembangan lembaga dipengaruhi oleh kemampuan administrator dalam melakukan scanning lingkungan eksternal, kompetitor Lembaga lain dan memperhitungkan kompetensi internal. Seiring dengan persaingan yang ketat antar pendidikan, kunci utama dalam meningkatkan jumlah calon peserta didik baru. Pihak sekolah harus lebih proaktif dalam mencari calon peserta didik istilahnya adalah “ jemput bola”. Kegiatan penerimaan peserta didik baru SMK Negeri 1 Jombang menghasilkan data profil peserta didik baru. Pengolahan data peserta didik baru dapat menghasilkan informasi yang bermanfaat bagi SMK Negeri 1 Jombang dalam membantu pihak panitia penerimaan peserta didik baru sebagai materi pertimbangan dalam menentukan strategi promosi penerimaan peserta didik baru. Untuk itu pentingnya dilakukan analisis data penerimaan peserta didik baru di SMK Negeri 1 Jombang dengan pendekatan algoritma Naïve bayes dengan SMOTE guna mendapatkan informasi-informasi yang bermanfaat dalam membantu menentukan penerimaan peserta didik baru pada tahun berikutnya. Penelitian terdahulu yang dilakukan oleh Tiara Nur Safitri dengan topik Pada SD Negeri 6 Ketapang. Dalam penelitiannya mengatakan bahwa terjadi ketidak konsistenan pada sistem penentuan penerimaan peserta didik baru di SD Negeri 6 Ketapang, maka perlu dilakukan penelitian menggunakan pendekatan klasifikasi dengan metode algoritma Naïve Bayes. Dari hasil evaluasi dan validasi diketahui bahwa Naïve Bayes memiliki nilai akurasi cukup tinggi yaitu sebesar 76% dengan demikian bahwa naïve bayes dapat dijadikan sebagai penentuan Beasiswa pada Sekolah Dasar Negeri 6 Ketapang[5]

METODE PENELITIAN

Pada penelitian ini peneliti menggunakan pendekatan deskriptif kuantitatif. Dalam pengertiannya penelitian deskriptif adalah penelitian yang menggambarkan suatu keadaan atau kondisi yang terjadi disuatu daerah, yang pemecahan masalahnya secara tersistem berdasarkan data – data yang bersikap fakta yang ada. dan data kualitatif adalah jenis data berupa angka yang dapat digunakan dalam proses operasi matematika, data ini berasal dari data fakta yang ada. Jenis data yang digunakan dalam penelitian ini bersifat rasio yang merupakan data yang berbentuk angka sebenarnya.[6]

Tahapan penelitian merupakan suatu proses memperoleh atau mendapatkan suatu pengetahuan atau memecahkan permasalahan yang dihadapi, yang dilakukan secara ilmiah, sistematis dan logis. Tahapan penelitian pada penelitian adalah sebagai berikut:



Gambar 1. Tahapan penelitian

HASIL DAN PEMBAHASAN

Data Selection

Dalam penelitian ini data yang digunakan akan diseleksi dengan cara melihat, kesesuaian data dengan topik atau judul penelitian yang akan diteliti, dalam hal ini data yang diperoleh dari SMK Negeri 1 Jombang sudah sesuai dengan format data yang terdiri dari atribut Pilihan, Jalur Pendaftaran, Nama Orang Tua, Nomor Pendaftaran, NISN, NIK, KK, Nama, Asal Sekolah, Pilihan, Skor, Nilai Rapot, Jenis Kelamin, Tempat Lahir,



Tanggal Lahir, Provinsi, Kab/Kota, Kecamatan, Kelurahan, RT, RW, Alamat, Titik koordinat.

Preprocessing Data

Tahap preprocessing data mining yaitu langkah pertama sebelum data diolah dengan metode data mining. Tahap ini penting karena sebagian besar proses data mining lebih banyak berhubungan preprocessing. Lebih dari separuh atau sekitar 60% peneliti menghabiskan waktu pada tahap ini. Preprocessing atau disebut juga preparasi data adalah proses sterilisasi data guna pada saat proses data memiliki keakuratan pola data.

Transpormasi data

Adapun pada tahap ini data akan diubah menjadi bentuk yang sesuai untuk proses data mining. Karena dalam penelitian ini akan dilakukan uji coba secara teoritis dan menggunakan software data mining yaitu RapidMiner, maka data yang telah melalui proses sebelumnya akan di transformasi agar dapat sesuai dengan algoritma yang dipakai yaitu algoritma Naïve Bayes. Pada tahapan ini atribut yang di pakai akan diberi label mengikuti kondisi data-data pada atribut tersebut. data set terdiri dari 638 peserta didik baru yang sudah melalui proses transformasi data.

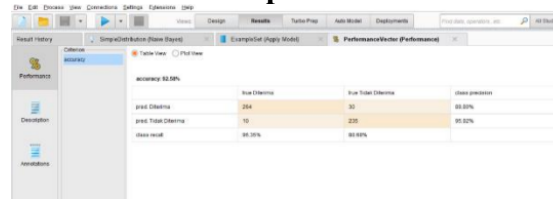
Penerapan Algoritma Naïve Bayes

Pengujian dengan aplikasi yang telah ditentukan. Hasil dari analisis diuji lagi dengan menggunakan software RapidMiner Studio



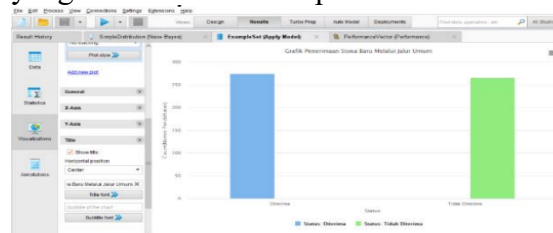
Gambar 2 Model Proses Algoritma Naïve Bayes

Evaluasi dan Interpretasi



Gambar 3. PerformanceVector

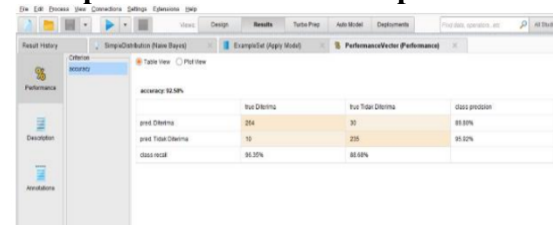
Berdasarkan hasil gambar 3 Performance Vector dengan hasil accuracy 92.58% hasil pred.diterima true diterima 264 dan pred.Tidak Diterima 10 peserta didik baru total yang diterima 274 peserta didik baru. Sedangkan hasil pred.Tidak Diterima true Tidak Diterima 235 dan 30 di pred.diterma true Tidak Diterima, total yang tidak diterima 265 peserta didik baru.



Gambar 4. Visualizations Bar

Dilihat dari hasil gambar 4 bahwa Dataset SMK Negeri 1 Jamblang yang diklasifikasikan menggunakan algoritma Naïve Bayes berdasarkan pendaftaran melalui jalur umum ternyata untuk penerimaan peserta didik baru terklasifikasikan status diterima sebanyak 274 peserta didik baru dan yang tidak diterima 265 peserta didik baru.

Tahapan Evaluasi dan Interpretasi



Gambar 5 PerformanceVector



Berdasarkan hasil gambar 5 Performance Vector dengan hasil accuracy 92.58% hasil pred.diterima true diterima 264 dan pred.Tidak Diterima 10 peserta didik baru total yang diterima 274 peserta didik baru. Sedangkan hasil pred.Tidak Diterima true Tidak Diterima 235 dan 30 di pred.diterma true Tidak Diterima, total yang tidak diterma 265 perseta didik baru.

PerformanceVector

```
PerformanceVector:
accuracy: 92.58%
ConfusionMatrix:
True:   Diterima   Tidak Diterima
Diterima: 264       30
Tidak Diterima: 10   235
```

Gambar 6 Deskripsi PerformanceVector

Dari gambar 6 menunjukkan bahwa yang di prediksi 264 peserta didik true diterima 264 peserta didik dan 10 dipred. Diterima. Sedangkan yang diprediksi tidak diterima 235 true Tidak Diterima dan 30 dipred Tidak Diterima.

Hasil akurasi penerapan penerapan machine learning untuk menentukan klasifikasi penerimaan peserta didik baru menggunakan algoritma Naïve Bayes adalah 92.58%.

Evaluasi Performance

Evaluasi kinerja model klasifikasi yang dihasilkan, dilihat berdasarkan pada hasil pengujian objek yang diprediksi dengan benar dan salah. Model klasifikasi akan ditentukan nilai akurasi. Akurasi dalam klasifikasi menentukan persentase ketepatan record data yang di klasifikasikan secara benar. Hasil hitungan, ditabulasi kedalam bentuk confusion matrix yang memiliki jumlah nilai true positif (TP), false negative (FN) dan true negative (TN).

Perhitungan akurasi menggunakan rumus:

$$: Accuracy = \frac{TP+TN}{TP+TN+FP+FN} * 100\%$$
$$Accuracy = \frac{264 + 235}{264 + 235 + 10 + 30} * 100\%$$
$$Accuracy = \frac{499}{539} * 100\%$$
$$Accuracy = 92.57\%$$

SIMPULAN

Hasil penelitian yang dilakukan di SMK Negeri 1 Jamblang terhadap 638 data penerimaan peserta didik baru di SMKN 1 Jamblang terklasifikasikan menjadi dua klas yaitu 274 klas diterima dan 265 klas tidak diterima dengan hasil akurasi 92,58%. penyelesaian class imbalance terhadap 638 data peserta didik baru di SMK Negeri 1 Jamblang menggunakan algoritma Naïve Bayes yang dikombinasikan dengan algoritma SMOTE menghasilkan akurasi 92.58%

DAFTAR PUSTAKA

- [1] D. A. K. Irfan Nurdiyanto, Odi Nurdiawan, Nining Rahaningsih, Ade Irfma Purnamasari, "Penentuan Keputusan Pemberian Pinjaman Kredit Menggunakan Algoritma C.45," J. Data Sci. dan Inform., vol. 1, no. 1, pp. 16–20, 2021.
- [2] A. S. kaslani, Ade Irma Purnamasari, "Pengembangan Media Pembelajaran Interaktif Berbasis Android Pada Materi Hidrokarbon," J. ICT Infirm. Comun. Technol., vol. 5, no. 1, p. 37, 2021, doi: 10.23887/jjpk.v5i1.33520.
- [3] I. A. Putri Saadah, Odi Nurdiawan , Dian Ade Kurnia, Dita Rizki Amalia, "Klasifikasi Penerima Beasiswa Dengan Menggunakan Algoritma," J. DATA Sci. Inform. (JDSI), vol. 1, no. 1, pp. 11–15, 2021.
- [4] I. A. Erliyana, Odi Nurdiawan, Nining R, Ade Irma Purnamasari,



- “Klasifikasi Penerima Beasiswa Dengan Menggunakan Algoritma,” *J. DATA Sci. Inform. (JDSI)*, vol. 1, no. 1, pp. 11–15, 2021.
- [5] D. Anggarwati, O. Nurdiawan, I. Ali, and D. A. Kurnia, “Penerapan Algoritma K-Means Dalam Prediksi Penjualan,” *J. DATA Sci. Inform. (JDSI)*, vol. 1, no. 2, pp. 58–62, 2021.
- [6] T. Hadi, N. Suarna, A. I. Purnamasari, O. Nurdiawan, and S. Anwar, “Game Edukasi Mengenal Mata Uang Indonesia ‘ Rupiah ’ Untuk Pengetahuan Dasar Anak-Berbasis Android,” *JURIKOM (Jurnal Ris. Komputer)*, vol. 8, no. 3, pp. 89–98, 2021, doi: 10.30865/jurikom.v8i3.3609.
- [7] O. Nurdiawan, R. Herdiana, and S. Anwar, “Komparasi Algoritma Naïve Bayes dan Algoritma K-Nearst Neighbor terhadap Evaluasi Pembelajaran Daring,” *Smatika J.*, vol. 11, no. 02, pp. 126–135, 2021, doi: 10.32664/smatika.v11i02.621.
- [8] A. rinaldi D. Subandi, Husein Odi Nuriawan, “Augmented Reality dalam Mendeteksi Produk Rotan menggunakan Metode Multimedia Development Life Cycle (MDLC),” *Means (Media Inf. Anal. dan Sist.*, vol. 6, no. 2, pp. 135–141, 2021.
- [9] H. S. Mr Agis, O. Nurdiawan, G. Dwilestari, and N. Suarna, “Informasi Penjualan Motor Bekas Berbasis Android Untuk Meningkatkan Penjualan di Mokascirebon.com,” *JURIKOM (Jurnal Ris. Komputer)*, vol. 8, no. 6, pp. 205–212, 2021, doi: 10.30865/jurikom.v8i6.3629.
- [10] D. Teguh, A. Ade, B. Riyan, T. Hartati, D. R. Amalia, and O. Nurdiawan, “Smart School Sebagai Sarana Informasi Sekolah di SDIT Ibnu Khaldun Cirebon,” *JURIKOM (Jurnal Ris. Komputer)*, vol. 8, no. 6, pp. 284–293, 2021, doi: 10.30865/jurikom.v8i6.3681.
- [11] I. Kepuasan, P. Informa, A. Febriyani, G. K. Prayoga, and O. Nurdiawan, “Index Kepuasan Pelanggan Informa dengan Menggunakan Algoritma C.45,” *JURIKOM (Jurnal Ris. Komputer)*, vol. 8, no. 6, pp. 330–335, 2021, doi: 10.30865/jurikom.v8i6.3686.
- [12] K. S. H. K. Al Atros, A. R. Padri, O. Nurdiawan, A. Faqih, and S. Anwar, “Model Klasifikasi Analisis Kepuasan Pengguna Perpustakaan Online Menggunakan K-Means dan Decision Tree,” *JURIKOM (Jurnal Ris. Komputer)*, vol. 8, no. 6, pp. 323–329, 2022, doi: 10.30865/jurikom.v8i6.3680.
- [13] F. Febriansyah, R. Nining, A. I. Purnamasari, O. Nurdiawan, and S. Anwar, “Pengenalan Teknologi Android Game Edukasi Belajar Aksara Sunda untuk Meningkatkan Pengetahuan,” *JURIKOM (Jurnal Ris. Komputer)*, vol. 8, no. 6, pp. 336–344, 2021, doi: 10.30865/jurikom.v8i6.3676.
- [14] E. S. Nugraha, A. R. Padri, O. Nurdiawan, A. Faqih, and S. Anwar, “Implementasi Aplikasi Pengaduan Masyarakat Berbasis Android Pada Gedung DPRD,” *JURIKOM (Jurnal Ris. Komputer)*, vol. 8, no. 6, pp. 360–366, 2021, doi: 10.30865/jurikom.v8i6.3679.
- [15] R. Nurcholis, A. I. Purnamasari, A. R. Dikananda, O. Nurdiawan, and S. Anwar, “Game Edukasi Pengenalan Huruf Hiragana Untuk Meningkatkan Kemampuan

ISSN 2338-1523
E-ISSN 2541-576X

Volume 10 No. 2
Agustus 2022



Berbahasa Jepang,” Build.
Informatics, Technol. Sci., vol. 3, no.
3, pp. 338–345, 2021, doi:
10.47065/bits.v3i3.1091.