

Penerapan Metode *Naive Bayes* Dalam Rekomendasi Strategi Penerimaan Peserta Didik Baru

Afan Hafara Sani, Agus Setiawan, Ardhin Primadewi*

Fakultas Teknik, Program Studi Teknik Informatika, Universitas Muhammadiyah Magelang, Magelang, Indonesia

Email: ¹Affanhafarasaney@gmail.com, ²setiawan@unimma.ac.id, ^{3,*}ardhin@ummgl.ac.id

Email Penulis Korespondensi: ardhin@ummgl.ac.id

Submitted: 29/10/2022; Accepted: 30/11/2022; Published: 30/11/2022

Abstrak—Penerimaan peserta didik baru (PPDB) merupakan kegiatan ujung tombak sekolah dalam mendapatkan siswa khususnya untuk sekolah swasta. SMK Muhammadiyah 1 Borobudur sebagai salah satu SMK swasta di Kabupaten Magelang. SMK Muhammadiyah 1 Borobudur memiliki permasalahan terkait strategi promosi PPDB. Promosi PPDB yang dilakukan belum sepenuhnya menunjukkan pertumbuhan promosi yang signifikan untuk meningkatkan jumlah siswa mendaftar pada setiap tahunnya dan informasi PPDB yang belum sampai ke pelosok Kabupaten Magelang. Dalam menentukan media promosi PPDB SMK Muhammadiyah 1 Borobudur maka penelitian ini menggunakan algoritma *Naive Bayes*. *Naive Bayes* adalah metode klasifikasi dengan menggunakan probabilitas dan statistik untuk memprediksi peluang dimasa depan berdasarkan masa sebelumnya. Dengan menggunakan algoritma *naive Bayes* yang diterapkan dengan perbandingan 70:30 untuk data train dan data test mampu menghasilkan akurasi sebesar 72%. Untuk memaksimalkan hasil akurasi data, penelitian ini menggunakan metode *k-fold cross-validation* dengan k=10 kali menghasilkan akurasi sebesar 75%. Selain itu dapat diketahui 3 informasi PPDB yang menjadi rekomendasi media promosi yaitu mitra siswa, alumni dan instagram.

Kata Kunci: Naive Bayes; Rekomendasi; Strategi; PPDB; Penerimaan siswa

Abstract—New Student Admission (PPDB) is a school spearheading activity in getting students, especially for private schools. SMK Muhammadiyah 1 Borobudur as one of the private vocational schools in Magelang Regency. SMK Muhammadiyah 1 Borobudur has problems related to the PPDB promotion strategy. The PPDB promotion that has been carried out has not fully shown a significant promotion growth to increase the number of students registering each year and PPDB information has not yet reached remote areas of Magelang Regency. In determining the promotion media for PPDB SMK Muhammadiyah 1 Borobudur, this study uses the Naive Bayes algorithm. Naive Bayes is a classification method using probability and statistics to predict future opportunities based on the past. By using the naive Bayes algorithm which is applied with a ratio of 70:30 for train data and test data, it is able to produce an accuracy of 72%. To maximize the results of data accuracy, this study used the k-fold cross-validation method with k=10 times resulting in an accuracy of 75%. In addition, it can be seen that 3 PPDB information is recommended for promotional media, namely student partners, alumni and Instagram.

Keywords: Naive Bayes; Recommendation; Strategy; PPDB; Student Admission

1. PENDAHULUAN

Penerimaan peserta didik baru (PPDB) merupakan kegiatan ujung tombak sekolah dalam mendapatkan siswa khususnya untuk sekolah swasta[1]. Keberhasilan dalam penyelenggaraan lembaga pendidikan (sekolah) bergantung kepada pengelolaan komponen-komponen pendukung pelaksanaan kegiatan seperti kurikulum, penerimaan peserta didik baru, pembiayaan, tenaga pelaksana, dan sarana prasarana [2]. Persaingan antar Sekolah Menengah Kejuruan (SMK) Swasta dalam PPDB semakin meningkat khususnya dalam media promosi baik secara daring maupun luring. Perkembangan teknologi menjadi dasar perkembangan media promosi luring, hal ini selaras dengan perkembangan teknologi 4.0[3]. Sekolah sebagai lembaga penyedia jasa pendidikan khususnya sekolah Swasta, secara tidak langsung menimbulkan persaingan. Dengan adanya persaingan antar SMK jumlah pendaftaran siswa baru mengalami fluktuasi calon peserta didik baru[2]

SMK Muhammadiyah 1 Borobudur sebagai salah satu SMK swasta di Kabupaten Magelang juga menghadapi persaingan yang cukup ketat. Berdasarkan data perolehan PPDB di SMK Muhammadiyah 1 Borobudur pada 3 tahun terakhir yaitu tahun 2020 (118 siswa), tahun 2021 (141 siswa) dan (121 siswa). Hasil observasi diketahui bahwa permasalahan yang dihadapi SMK Muhammadiyah 1 Borobudur terkait PPDB adalah kenaikan biaya operasional promosi yang terus meningkat mengikuti inflasi dan upaya menarik siswa baru dari daerah Kabupaten Magelang dengan radius 20 KM (daerah Sawangan, Srumbung, dan sekitarnya). Namun demikian kegiatan promosi yang terdiri dari kegiatan pemasangan iklan, penyebaran brosur, sosial media, testimoni terhadap sekolah dari para siswa dan alumni (mitra siswa) dan persentasi guru ke SMP sekitar radius 20 KM (daerah Sawangan, Srumbung, dan sekitarnya) belum menunjukkan pertumbuhan promosi yang signifikan pada sekolah.

Hal ini mengakibatkan perlu adanya rekomendasi[4], [5]strategi PPDB SMK Muhammadiyah 1 Borobudur yang dapat membantu panitia PPDB dalam meningkatkan media promosi serta jumlah siswa mendaftar pada setiap tahunnya. Selain itu agar minat masyarakat di pelosok Kabupaten Magelang dapat mengetahui informasi PPDB SMK Muhammadiyah 1 Borobudur secara akurat.

Penelitian terkait rekomendasi pada organisasi baik pada bidang promosi ataupun yang lainnya khususnya di lingkup pendidikan sangat berkembang. Pada tabel 1 di bawah ini sebagai rekam jejak penelitian sebelumnya yang khusus membahas strategi promosi[6], memprediksi potensi pendaftaran siswa[2], dan mahasiswa [7], [8], serta penelusuran minat calon mahasiswa[9].

Dalam menentukan media promosi PPDB SMK Muhammadiyah 1 Borobudur maka penelitian ini menggunakan algoritma *naïve bayes*. *naïve bayes* adalah metode klasifikasi dengan menggunakan probabilitas dan statistik untuk memprediksi peluang dimasa depan berdasarkan masa sebelumnya[10], [11]. Dalam proses klasifikasi *naïve bayes* tidak membutuhkan data train yang besar[12]. Hingga saat ini banyak penelitian yang telah dilakukan dengan menerapkan metode *naïve Bayes*.

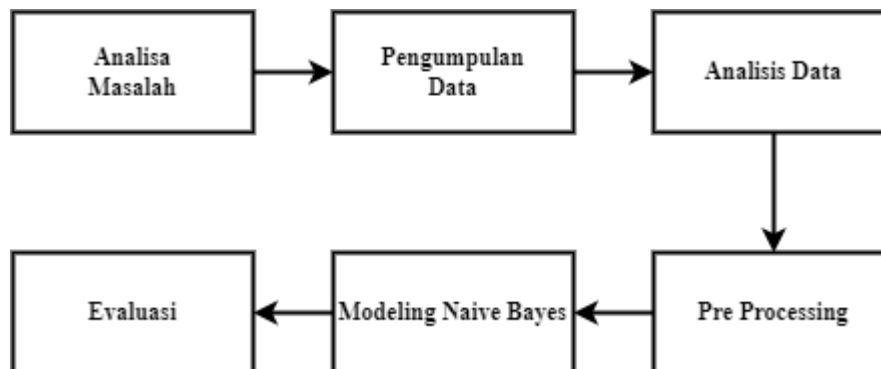
Beberapa penelitian yang telah dilakukan dengan menerapkan metode *naïve Bayes*. Seperti penelitian yang dilakukan Dwiarti Rahma Utami, Adi Kurniawan, dan Ultach Enri pada tahun 2021 tentang *naïve bayes* klasifikasi untuk rekomendasi strategi promosi tindakan pada Aesthetic Dental Clinic Karawang, berhasil menerapkan metode *naïve bayes* menggunakan RapidMiner yaitu 31,7% dengan nilai kappa 0,183 untuk mendukung strategi promosi berpengaruh pada meningkatnya jumlah pelanggan baru [6]. Penelitian yang dilakukan oleh Sherlyn Eka Yuliana pada tahun 2021 tentang penerapan model *naïve bayes* untuk memprediksi potensi pendaftar siswa memberikan hasil klasifikasi untuk memprediksi mendaftar dan tidak mendaftar dengan akurasi 86% untuk prediksi potensi siswa mendaftar [2]. Penelitian yang dilakukan oleh Ahmad Haidar Mirza pada tahun 2017 tentang penerapan algoritma *naïve bayes classifier* dalam menentukan strategi promosi penerimaan mahasiswa baru, hasil dari penelitian ini yaitu memprediksi proses Registrasi dan Tidak Registrasi penerimaan mahasiswa baru berdampak pada efektifitas dan efisiensi promosi peningkatan jumlah mahasiswa baru yang akan mendaftar [13]. Penelitian yang dilakukan oleh Irayori Loelianto, DKK pada tahun 2020 tentang Implementasi teori *naïve bayes* dalam klasifikasi calon mahasiswa baru penelitian ini berhasil menerapkan metode *naïve bayes* dengan 499 data yang dibagi dengan rasio 80:20 menghasilkan akurasi 73% [8]. Penelitian yang dilakukan Budiman dan Zatin Niqotaini pada tahun 2022 tentang analisis perbandingan kinerja algoritma klasifikasi data mining penelusuran minat calon mahasiswa, berhasil menerapkan data klasifikasi untuk penelusuran potensi dan minat calon siswa baru menggunakan data rasio 70:30 menghasilkan klasifikasi J48 dengan kesalahan paling rendah sebesar 0,26% [9].

Berdasarkan permasalahan yang telah dijelaskan sebelumnya, maka penulis melakukan penelitian tentang penerapan metode *naïve bayes* dalam rekomendasi strategi PPDB SMK Muhammadiyah 1 Borobudur yang diharapkan mampu memberikan sebuah hasil rekomendasi kepada Panitia PPDB untuk membantu menentukan media promosi.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Tahapan penelitian yang dilakukan dalam rekomendasi strategi PPDB SMK dengan menerapkan *metode naïve bayes* dapat dilihat pada gambar 1. Pada tahap pertama dilakukan proses analisa ini dilakukan proses identifikasi masalah. SMK Muhammadiyah 1 Borobudur memiliki permasalahan terkait strategi promosi PPDB. Promosi PPDB yang dilakukan belum sepenuhnya menunjukkan pertumbuhan promosi yang signifikan untuk meningkatkan jumlah siswa mendaftar pada setiap tahunnya dan informasi PPDB yang belum sampai ke pelosok Kabupaten Magelang. Masalah tersebut mengakibatkan Panitia PPDB kesulitan menentukan rekomendasi strategi promosi PPDB terbaik untuk SMK Muhammadiyah 1 Borobudur.



Gambar 1. Tahapan Penelitian

Pada tahap kedua dilakukan pengumpulan data yang dilakukan dengan melakukan wawancara dan membagikan kuesioner selama 9 bulan (Oktober 2021 – Juni 2022). Selanjutnya tahap Analisis Data dilakukan dengan mendiskripsi dataset dan visualisasi data dalam bentuk grafik. Hal ini akan mempermudah dalam menampilkan informasi detail tentang data dan melihat jumlah frekuensi nilai data. Seperti menampilkan data frekuensi media promosi dan jumlah siswa mendaftar per tahun.

Pada tahap berikutnya dilakukan proses *pre processing* menghapus duplikasi data dan missing value bertujuan untuk menghapus data yang terindikasi memiliki duplikasi data menghilangkan *missing value* pada data. Sehingga pada saat diproses tidak muncul kesalahan atau masalah. Selanjutnya *pre processing* melakukan

penghapusan atribut yang tidak digunakan. Sehingga atribut yang tersisa hanya atribut yang nantinya digunakan pada tahap modeling. Preprocessing selanjutnya *feature engineering* dalam tahap ini melakukan pengelompokan data berdasarkan tipe datanya. Tipe data yang dimaksud berupa tipe data karakter dan numerik. Selain itu tahap ini juga melakukan encoding data untuk tipe data karakter. Tahap berikutnya *feature vector* dalam tahap ini menentukan variabel independen dan variabel dependen. Kemudian proses PreProcessing adalah *feature scaling* yaitu membuat data berada dirange yang sama.

Pada tahap berikutnya dilakukan *modeling* untuk pengambilan kesimpulan hasil rekomendasi strategi PPDB SMK Muhammadiyah 1 Borobudur dengan menggunakan metode *naïve bayes* dan bahasa pemrograman Python. Pada tahap terakhir dilakukan evaluasi Untuk memaksimalkan hasil akurasi pada penelitian ini menggunakan metode *k-fold cross-validation*.

2.2 Metode Naïve Bayes

Naïve bayes adalah pengklasifikasi statistik yang dapat digunakan untuk memprediksi probabilitas milik suatu kelas [14]. Salah satu teknik dalam data mining untuk menyelesaikan masalah menggunakan prinsip peluang (probability)[15]. Algoritma ini menggunakan teorema Bayes bahwa atribut adalah independen atau tidak saling bergantung, diberikan oleh nilai variabel kelas [16]. Cara kerja metode *naïve bayes* yaitu memakai rumus yang digunakan untuk mencari nilai peluang dalam metode *naïve bayes* sebagai berikut :

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)} \quad (1)$$

Keterangan:

X = Sampel data yang memiliki kelas

H = Hipotesa x data kelas

P(H) = Peluang Hipotesa

P(X) = Peluang sampel data yg memiliki kelas

P(X|H) = Peluang dari data sampel X dari data hipotesa data kelas

3. HASIL DAN PEMBAHASAN

Pada penelitian ini menggunakan Algoritma *Naïve Bayes* untuk menyelesaikan klasifikasi Rekomendasi Strategi PPDB SMK Muhammadiyah 1 Borobudur , beberapa hasil dan pembahasannya adalah sebagai berikut :

3.1 Analisa Masalah

Penelitian ini dilakukan di SMK Muhammadiyah 1 Borobudur yang memiliki 5 jurusan yaitu Akuntansi Dan Keuangan Lembaga (AKL), Otomatisasi Dan Tata Kelola Perkantoran (OTKP), Bisnis Daring Dan Pemasaran (BDP), Tata Busana (TB), Dan Teknik Komputer Dan Jaringan (TKJ). Berdasarkan data perolehan Penerimaan Peserta Didik Baru (PPDB) di SMK Muhammadiyah 1 Borobudur selama 3 tahun terakhir di peroleh yaitu 118 siswa (tahun 2020), 141 siswa (tahun 2021) dan 121 siswa (tahun 2022).

Berdasarkan hasil observasi selama 9 bulan (Oktober 2021 – Juni 2022) diketahui permasalahan yang dihadapi SMK Muhammadiyah 1 Borobudur terkait PPDB adalah kenaikan biaya oprasional promosi yang terus meningkat mengikuti inflasi dan upaya menarik siswa baru dari daerah Kabupaten Magelang dengan radius 20 KM (daerah Sawangan, Srumbung, dan sekitarnya). Namun demikian kegiatan promosi yang terdiri dari kegiatan pemasangan iklan, penyebaran brosur, sosial media, testimoni terhadap sekolah dari para siswa dan alumni (mitra siswa) dan persentasi guru ke SMP sekitar radius 20 KM (daerah Sawangan, Srumbung, dan sekitarnya) belum menunjukkan pertumbuhan promosi yang signifikan pada sekolah.

Hal ini mengakibatkan perlu adanya rekomendasi strategi PPDB SMK Muhammadiyah 1 Borobudur yang dapat membantu panitia PPDB dalam meningkatkan promosi serta jumlah siswa mendaftar pada setiap tahunnya. Selain itu agar minat masyarakat di pelosok Kabupaten Magelang dapat mengetahui informasi PPDB SMK Muhammadiyah 1 Borobudur secara akurat.

3.2 Pengumpulan Data

Data promosi yang digunakan merupakan hasil wawancara dan membagikan kuesioner selama 9 bulan (Oktober 2021 – Juni 2022). Data yang akan digunakan pada proses data mining menggunakan Algoritma Naïve Bayes merupakan data PPDB yang didapatkan dari siswa angkatan tahun 2020, 2021 dan 2022. Data ini merupakan gambaran nyata tingkat keberhasilan promosi panitia PPDB SMK Muhammadiyah 1 Borobudur.

Setelah data di dapat dari kuesioner yang dibagikan ke siswa angkatan tahun 2020, 2021 dan 2022. Selanjutnya data akan diolah menggunakan algoritma Naïve Bayes yang diimplementasikan dengan menggunakan bahasa pemrograman Python. Dengan menggunakan data PPDB dari 3 angkatan terakhir diharapkan dapat diketahui rekomendasi terbaik dalam strategi promosi yang di prioritaskan oleh panitia PPDB SMK Muhammadiyah 1 Borobudur. Pada tabel 1 merupakan contoh dataset yang diolah dari data hasil kuesioner 3

angkatan tahun 2020, 2021 dan 2022. Data yang digunakan berupa tabel dan data menggunakan atribut Nama, Alamat, Asal sekolah, Jurusan Gelombang PPDB, Informasi PPDB, Tahun PPDB, dan Saran Promosi PPDB.

Tabel 1. Data set

Timest amp	Nama_Le ngkap	Alama t	Asal_Se kolah	Juru san	Gelombang _PPDB	Informasi _PPDB	Tahun_ PPDB	Saran_Promo si_PPDB
9/30/2 022 7.55.22	Sasa Maryana	Kecam atan Dukun	SMP N 1 DUKUN	AKL	Gelombang 1	Mitra siswa	2020	Sosialisasi ke SMP
9/30/2 022 7.55.36	Nova Afifatunni 'mah	Kecam atan Borob udur	MTS N 1 MAGEL ANG	AKL	Gelombang 3	Mitra siswa	2022	Menambahkan promosi PPDB di tiktok
9/30/2 022 7.55.42	Maya Wulandari	Kecam atan Borob udur	MTS N 1 MAGEL ANG	AKL	Gelombang 2	Alumni	2022	Menambahkan promosi PPDB di tiktok

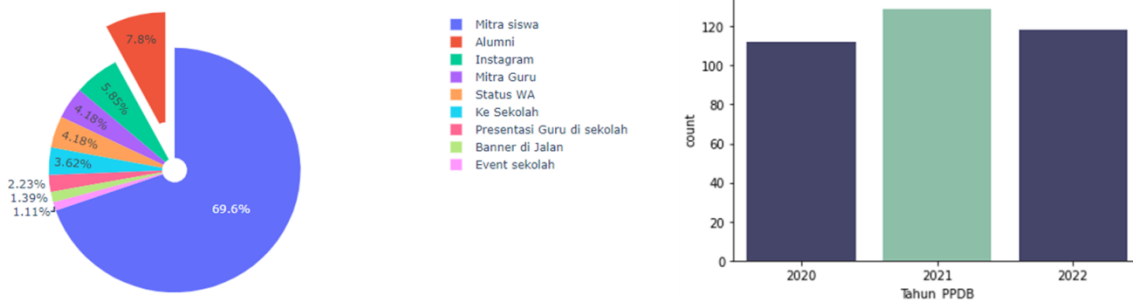
3.3 Analisis Data

Data set berikutnya merupakan data siswa yang mendaftar selama 3 tahun (gambar 2 sisi kiri) dengan distribusi terbanyak pada siswa dengan asal Kecamatan Borobudur (49%), Kecamatan Mungkid (27%), Kecamatan Mertoyudan (9%), Kecamatan Muntilan (6%). Data set jurusan distribusi terbanyak yang mendaftar siswa 3 tahun terakhir (2020, 2021 dan 2022) pada jurusan OTKP (25%), jurusan TB (23%), dan jurusan AKL (22%). Data ini didapatkan dari data pada gambar 2 di sebelah kanan dibagi jumlah populasi data (359 siswa).

Kecamatan Borobudur	0.498607	OTKP	93
Kecamatan Mungkid	0.270195	TB	85
Kecamatan Mertoyudan	0.091922	AKL	79
Kecamatan Muntilan	0.066852	TKJ	54
lain -lain	0.027855	BDP	48
Kecamatan Salaman	0.011142	Name: jurusan, dtype: int64	
Kecamatan Tempuran	0.011142		
Kecamatan Dukun	0.008357		
Kecamatan Sawangan	0.008357		
Kecamatan Secang	0.002786		
Kecamatan Ngluwar	0.002786		
Name: Alamat, dtype: float64			

Gambar 2 Data frekuensi Kecamatan asal siswa mendaftar (kiri) dan jurusan pilihan siswa mendaftar (kanan)

Data set berikutnya merupakan data media promosi yang digunakan panitia PPDB SMK Muhammadiyah 1 Borobudur. Media promosi selama 3 tahun terakhir (2020, 2021 dan 2022) pada gambar 3 dengan distribusi terbanyak pada Mitra siswa (69%), alumni (7%), dan Instagram (5%). Jumlah siswa mendaftar pertahun selama 3 tahun terakhir (2020, 2021 dan 2022) pada gambar 3 sisi kanan tahun 2020 (112 siswa), tahun 2021 (129 siswa), dan tahun 2022 (118 siswa).



Gambar 3. Data frekuensi media promosi (kiri) dan jumlah siswa mendaftar per tahun (kanan)

3.4 PreProcessing

Pada tahap *pre processing* bertujuan untuk membersihkan data mentah menjadi data siap diolah. Dari data semula yang memiliki value tidak konsisten seperti nama asal sekolah yang disingkat, menggunakan huruf besar dan kecil harus disamakan terlebih dahulu. Selain itu pada kolom informasi siswa banyak yang memilih informasi lain –lain karena tidak paham dengan informasi PPDB. Untuk yang memilih informasi PPDB lain – lain dilakukan kuesioner

ulang, memastikan alasan memilih informasi PPDB lain –lain. Yang awalnya berjumlah 56 setelah dibagikan kuesioner menjadi 0, serta menambahkan value baru Alumni. Selanjutnya dilakukan penambahan atribut rekomendasi yang nantinya akan menjadi atribut target. Atribut rekomendasi terdiri atas 2 value yaitu efektif dan tidak efektif. Berikut gambar 4 hasil pre processing tahap awal :

	timestamp	nama	alamat	asal_sekolah	jurusan	gelombang_ppdb	informasi_ppdb	tahun_ppdb	saran	rekomendasi
0	9/30/2022 7:55:22	Sasa Maryana	Kecamatan Dukun	SMP N 1 DUKUN	AKL	Gelombang 1	Mitra siswa	2020	Sosialisasi ke SMP*	efektif
1	9/30/2022 7:55:36	Nova Afifatunnimah	Kecamatan Borobudur	MTS N 1 MAGELANG	AKL	Gelombang 3	Mitra siswa	2022	menambahkan promosi PPDB di tiktok	efektif
2	9/30/2022 7:55:42	MAYA WULANDARI	Kecamatan Borobudur	MTS N 1 MAGELANG	AKL	Gelombang 2	Alumni	2022	Menambah kan promosi PPDB di tiktok	tidak efektif

Gambar 4. Pre processing tahap awal

- a. Menghapus duplikasi data dan missing value

Pada tahap *pre processing* selanjutnya bertujuan untuk menghapus data yang terindikasi memiliki duplikasi data menghilangkan missing value pada data. Sehingga pada saat diproses tidak muncul kesalahan atau masalah. Tahap ini menghapus data yang duplikat. Awalnya data berjumlah 359 baris serta 10 kolom selanjutnya dilakukan penghapusan duplikasi yang ternyata hasilnya tidak berubah sehingga dapat terlihat bahwa data tersebut tidak memiliki duplikasi data. Begitu pula bahwa setelah data dicek tidak memiliki missing value. Berikut gambar 5 proses penghapusan duplikasi dan missing value sebagai berikut:

```
[ ] df.shape
      (359, 10)

[ ] df=df.drop_duplicates()

[ ] df.shape
      (359, 10)

timestamp      0
nama           0
alamat         0
asal_sekolah   0
jurusan        0
gelombang_ppdb 0
informasi_ppdb 0
tahun_ppdb     0
saran          0
rekomendasi    0
dtype: int64
```

Gambar 5. Proses penghapusan duplikasi (kiri) dan proses missing value (kanan)

- b. Menghapus atribut yang tidak digunakan

Preprocessing selanjutnya melakukan penghapusan atribut yang tidak digunakan pada tahap selanjutnya. Sehingga atribut yang tersisa hanya atribut yang nantinya digunakan pada tahap *modeling*. Pada tabel 2 merupakan Tahap melakukan penghapusan terhadap 3 atribut yaitu *timestamp*, *nama* dan *saran*. Berikut data setelah dilakukan penghapusan atribut sebagai berikut :

Tabel 2. Setelah penghapusan atribut

alamat	asal_sekolah	jurusa n	gelombang_ppdb	Informasi_ppd b	Tahun _ppdb	rekomen dasi
Kecamatan Dukun	SMP N 1 DUKUN	AKL	Gelombang 1	Mitra siswa	2020	efektif
Kecamatan Borobudur	MTS N 1 MAGELANG	AKL	Gelombang 3	Mitra siswa	2022	efektif
Kecamatan Borobudur	MTS N 1 MAGELANG	AKL	Gelombang 2	Alumni	2022	Tidak efektif
Kecamatan Borobudur	SMP N 2 BOROBUDUR	AKL	Gelombang 1	Mitra siswa	2022	efektif
Kecamatan Borobudur	SMP N 2 SALAMAN	AKL	Gelombang 3	Mitra siswa	2022	efektif

- c. Feature Engineering

Pre processing dalam tahap ini melakukan pengelompokan data berdasarkan tipe datanya. Kelompok data numerikal terdiri atas data yang memiliki tipe data numerik yaitu *tahun_ppdb* sedangkan kelompok data kategorikal memiliki tipe data berupa karakter yaitu *alamat*, *asal_sekolah*, *jurusan*, *gelombang_ppdb*, *informasi_ppdb*, dan *rekomendasi*. Selain itu tahap ini melakukan encoding data sehingga data yang memiliki tipe data karakter diubah menjadi angka. Berikut gambar 6 hasil pre processing pada tahap *feature engineering*:

	alamat	asal_sekolah	jurusan	gelombang_ppdb	informasi_ppdb	tahun_ppdb	rekomendasi
0	1	43	0	0	6	2020	1
1	0	11	0	2	6	2022	1
2	0	11	0	1	0	2022	0
3	0	48	0	0	6	2022	1
4	0	57	0	2	6	2022	1

Gambar 6. Hasil feature engineering

d. Feature vector

Pada penelitian ini untuk variable dependen (y) yaitu atribut rekomendasi. Lalu untuk variable independen (x) yaitu variable alamat, asal_sekolah, jurusan, gelombang_ppdb, dan informasi_ppdb. Kemudian membagi data training dan data testing, dengan perbandingan 70:30. X_train terdiri atas sebanyak 251 baris dan 6 kolom. X_test sebanyak 108 baris dan 6 kolom.

e. Feature scaling

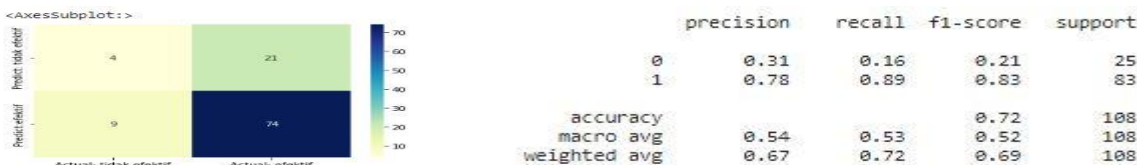
Pada tahap *feature scaling* penelitian ini menggunakan metode *min-maxscaler*. *min-maxscaler* mengubah data sehingga distribusinya menghasilkan nilai rata-rata 0 dan standar 1. Sehingga hasil dari tahap ini dapat dilihat pada gambar 7 berikut :

	alamat	asal_sekolah	jurusan	gelombang_ppdb	informasi_ppdb	tahun_ppdb
0	0.3	0.125000	0.50	0.000000	0.750	0.5
1	0.0	0.833333	0.00	0.333333	0.000	0.0
2	0.3	0.444444	0.75	0.666667	0.750	0.5
3	1.0	0.236111	0.50	0.666667	0.375	0.0
4	0.0	0.152778	1.00	0.000000	0.750	0.0

Gambar 7. Hasil *feature scaling*

3.5 Modeling Naïve Bayes

Pada gambar 8 merupakan hasil dari akurasi berdasarkan algoritma *naïve bayes* , penelitian ini menghasilkan akurasi keseluruhan bernilai 72%, dari atribut serta hasil untuk perhitungan *precision, recall, f1-score*



Gambar 8. Hasil *Confusion Matrix* (kiri) dan Hasil Akurasi berdasarkan algoritma *naïve bayes* (kanan)

Ukuran kinerja klasifikasi machine learning berdasarkan nilai prediksi dan aktual adalah apa yang dimaksud dengan *Confusion Matrix*. Metode *Naive Bayes* yang memiliki empat kombinasi nilai prediksi dan nilai aktual, empat diantaranya adalah *True Positive (TP)*, *True Negative (TN)*, *false positif (FP)* dan *negatif palsu (FN)*[17].

Berdasarkan gambar 10 (kiri) hasil dari performa *confusion matrix* data X_test yang digunakan dengan label tidak efektif berjumlah 21 dan label efektif berjumlah 74. Memiliki prediksi nilai benar sebanyak 4 dan prediksi nilai salah sebanyak 9. Hasil algoritma *naïve bayes* pada gambar 10 (kanan) menghasilkan value yang tidak efektif memiliki *precession* sebesar 31%, *recall* sebesar 16%, dan *f1-score* sebesar 21%. Untuk value yang efektif memiliki *precession* sebesar 78%, *recall* sebesar 89%, dan *f1-score* sebesar 83%. Sehingga akurasi model *naïve bayes* yang didapatkan sebesar 72%.

3.6 Evaluasi

Berikutnya yaitu hasil prediksi berdasarkan data X_test nya. Pada hasil klasifikasi menggunakan algoritma *naïve bayes* dan menggunakan dataset X_test nya maka menghasilkan urutan informasi PPDB. Urutan tersebut dihasilkan dari jumlah informasi PPDB terbanyak. Urutan tersebut dapat dijadikan rekomendasi media promosi PPDB pada SMK Muhammadiyah 1 Borobudur. Sehingga apabila diambil dari 3 urutan teratas maka mitra siswa, alumni dan *instagram* menjadi rekomendasi media promosi untuk PPDB selanjutnya. Pada gambar 9 dibawah ini merupakan hasil akhir penelitian untuk klasifikasi dengan *naïve bayes*.

	informasi	total_num
0	Mitra siswa	72
1	Alumni	9
2	Instagram	8
3	Mitra Guru	7
4	Presentasi Guru di sekolah	4
5	Ke Sekolah	3
6	Status WA	3
7	Banner di Jalan	2

Gambar 9. Hasil Prediksi berdasarkan data X_test

Untuk memaksimalkan hasil akurasi pada penelitian ini menggunakan metode *k-fold cross-validation*. Dimana pada satu model dengan value yang sama akan di uji cobakan sebanyak k kali. Pengujian menggunakan 10 kali percobaan dan diambil nilai rata-ratanya sehingga menghasilkan akurasi data sebesar 75%.

4. KESIMPULAN

Maka diketahui bahwa data test yang didapat tidak mempunyai duplikasi data dan *missing value*. Selain itu kolom yang digunakan dalam penelitian ini adalah jurusan, gelombang, informasi, dan tahun sebagai variabel independen. Sedangkan untuk variabel dependennya adalah rekomendasi. Dengan menggunakan algoritma *naïve bayes* yang diterapkan dengan perbandingan 70:30 untuk data train dan data test mampu menghasilkan akurasi sebesar 72%. Upaya untuk memaksimalkan hasil akurasi data, penelitian ini menggunakan metode *k-fold cross-validation* dengan k=10kali menghasilkan akurasi sebesar 75%. Selain itu dapat diketahui 3 informasi PPDB yang menjadi rekomendasi media promosi yaitu mitra siswa, alumni dan *instagram*.

REFERENCES

- [1] A. Saputro and A. Primadewi, "Implementation of the Code Igniter Framework in the Online New Student Admission Application at SMK Muhammadiyah 2 Mertoyudan Magelang," *Urecol Journal. Part E Eng.*, vol. 2, no. 1, pp. 37–46, 2022.
- [2] S. Eka *et al.*, "PENERAPAN MODEL NAIVE BAYES UNTUK MEMPREDIKSI POTENSI PENDAFTARAN SISWA DI SMK TAMAN SISWA TELUK BETUNG BERBASIS WEB," *J. Eng. Comput. Sci. Inf. Technol.*, vol. 1, no. 1, pp. 82–87, 2021.
- [3] Reflianto and Syamsuar, "Pendidikan dan Tantangan Pembelajaran Berbasis Teknologi Informasi di Era Revolusi Industri 4.0," *J. Ilm. Teknol. Pendidik.*, vol. 6, no. 2, pp. 1–13, 2018.
- [4] D. J. Lubis and A. I. Ningtias, "Penerapan Metode Naïve Bayes Untuk Rekomendasi Pemilihan Asisten Laboratorium Komputer Di Perguruan Tinggi," *Teknois J. Ilm. Teknol. Inf. dan Sains*, vol. 12, no. 2, pp. 127–138, 2022.
- [5] F. Farid, U. Enri, and Y. Umaidah, "Sistem Pendukung Keputusan Rekomendasi Topik Skripsi Menggunakan Naïve Bayes Classifier," *JOINTECS (Journal Inf. Technol. Comput. Sci.)*, vol. 6, no. 1, p. 35, 2021.
- [6] D. R. Utami, A. Kumiawan, and U. Enri, "Naïve Bayes Klasifikasi untuk Rekomendasi Strategi Promosi Tindakan pada Aesthetic Dental Clinic Karawang," *J-SAKTI (Jurnal Sains Komput. Inform.)*, vol. 5, no. September, pp. 601–610, 2021.
- [7] A. H. Mirza, "Application of Naive Bayes Classifier Algorithm in Determining New Student Admission Promotion Strategies," *J. Inf. Syst. Informatics*, vol. 1, no. 1, pp. 14–28, 2019.
- [8] I. Loelianto, M. S. S. Thayf, and H. Angriani, "Implementasi Teori Naive Bayes Dalam Klasifikasi Calon Mahasiswa Baru Stmik Kharisma Makassar," *SINTECH (Science Inf. Technol. J.)*, vol. 3, no. 2, pp. 110–117, 2020.
- [9] Budiman and Z. Niqotaini, "Analisis Perbandingan Kinerja Algoritma Klasifikasi Data Mining Penelusuran Minat Calon Mahasiswa," *Sist. J. Sist. Inf.*, vol. 11, pp. 271–290, 2022.
- [10] A. Senika, R. Rasiban, and D. Iskandar, "Implementasi Metode Naive Bayes Dalam Penilaian Kinerja Sales Marketing Pada PT. Pachira Distrinusa," *J. Media Inform. Budidarma*, vol. 6, no. 1, p. 701, 2022.
- [11] I. Riswanto and R. H. Laluma, "Klasifikasi Kelayakan Pinjaman Pada Koperasi Karyawan Menggunakan Metode Naive Bayes Classifier Berbasis Web," *Infotronik J. Teknol. Inf. dan Elektron.*, vol. 5, no. 1, pp. 11–16, 2020.
- [12] R. H. Sukarna and Y. Ansori, "Implementasi Data Mining Menggunakan Metode Naive Bayes Dengan Feature Selection Untuk Prediksi Kelulusan Mahasiswa Tepat Waktu," *J. Ilm. Sains dan Teknol.*, vol. 6, no. 1, pp. 50–61, 2022.
- [13] D. Wahyudi, A. H. Mirza, and P. H. Merriayu, "Implementasi Data Mining Dengan Naive Bayes Classifier Untuk Mendukung Strategi Promosi (Studi Kasus Universitas Bina Darma Palembang)," *Univ. Bina Darma Palembang*, vol. 1, no. 1, 2017.
- [14] Vianti Widyasari and Arief Senja Fitriani, "Application of Data Mining with Classification Methods for Promotion of New Student Admissions at Muhammadiyah University of Sidoarjo Using Web-Based Naïve Bayes Algorithm," *Procedia Eng. Life Sci.*, vol. 1, no. 2, 2021.
- [15] A. Carol, "Penerapan Naive Bayesian Classifier Dalam Penyeleksian Beasiswa PPA," *J. Comput. Syst. Informatics*, vol. 2, no. 2, pp. 194–198, 2021.
- [16] S. Wulandari, W. Saputra, S. Tunas Bangsa Pematangsiantar, and A. A. Tunas Bangsa Pematangsiantar JJenderal Sudirman Blok No, "Penerapan Metode Naive Bayes dalam Menentukan Pengaruh Penasihat Akademik pada Kelulusan Mahasiswa Tingkat Akhir," *Pros. Semin. Nas. Ris. Inf. Sci.*, vol. 1, no. September, pp. 661–669, 2019.
- [17] P. A. Winata, "Klasifikasi Naive Bayes Keparahan Trauma Pasien Menggunakan Data Neuro Cognitive Dan Data Physiologic Dengan Python," *eminar Nas. Mat. Geom. Stat. dan Komputas*, pp. 99–109, 2022.