
PERBANDINGAN ALGORITME NAÏVE BAYES DAN KNN TERHADAP DATA PENERIMAAN BEASISWA (Studi Kasus Lembaga Beasiswa Baznas Jabar)

Muhamad Riyyan¹, Hafiz Firdaus²

^{1,2}Program Studi Teknik Informatika, Universitas Singaperbangsa Karawang

Jl. HS. Ronggo Waluyo, Telukjambe Timur, Karawang, Jawa Barat 41361

¹muhamad.riyyan17006@student.unsika.ac.id, ²hafiz.firdaus17107@student.unsika.ac.id

ABSTRACT

The Indonesian government has launched a 12-year compulsory education program to improve the quality of education in Indonesia. Therefore, the government issues regulations that are made for the authority of local governments and so on to help the cost of education for certain people. The West Java Baznas Scholarship Institute (LBB Jabar) is one of the institutions that provide tuition fees for underprivileged students to meet the requirements, but the application for these scholarships increases excessively and produces quite large data. Therefore, the selection process becomes inefficient and requires certain algorithmic methods. Analysis of large data needs to be done to find out the patterns and knowledge of the data. This study was conducted to compare naive bayes and knn which can be applied properly to the scholarship acceptance data held by LBB Jabar. The classification model created is tested with a confusion matrix. The results showed different values in each algorithm. The test value of the naive Bayes algorithm is that the accuracy value obtained is 80%, the precision is 82.82%, and the recall is 92.47% slightly larger than the KNN algorithm with an accuracy value of 78.97%, a precision of 82.61%, and recall of 91.10% where the value of k used is k=7. Meanwhile, the k=27 knn algorithm has a greater accuracy value with a value of 80.51%, a precision of 81.03%, and a recall of 96.58%.

Keywords : KNN, naive bayes, scholarship, Baznas Jabar

ABSTRAK

Pemerintah Indonesia telah mencanangkan program wajib belajar selama 12 tahun dengan tujuan untuk meningkatkan kualitas pendidikan di Indonesia. Oleh karena itu pemerintah mengeluarkan peraturan-peraturan yang dibuat untuk kewenangan pada pemerintah daerah dan sebagainya agar dapat membantu biaya pendidikan pada orang-orang tertentu. Lembaga Beasiswa Baznas Jawa Barat (LBB Jabar) adalah salah satu lembaga yang memberikan biaya pendidikan bagi pelajar yang kurang mampu secara pembiayaan yang memenuhi persyaratan, namun pendaftar dari beasiswa tersebut masuk secara berlebih dan menghasilkan data yang cukup besar. Oleh karena itu proses seleksi menjadi tidak efisien dan membutuhkan metode algoritme tertentu. Analisa terhadap data yang besar perlu dilakukan untuk bisa mengetahui pola dan pengetahuan pada data tersebut. Penelitian ini dilakukan untuk membandingkan metode *naive bayes* dan knn yang dapat diterapkan dengan baik pada data penerimaan beasiswa yang diselenggarakan oleh LBB Jabar. Model klasifikasi yang dibuat diuji dengan *confusion matrix*. Hasil penelitian menunjukkan nilai yang berbeda pada setiap algoritma. Nilai uji algoritme *naive bayes* yaitu nilai akurasi yang didapat sebesar 80%, presisi sebesar 82,82%, dan recall sebesar 92,47% sedikit lebih besar dibandingkan dengan algoritme knn dengan nilai akurasi 78,97%, presisi sebesar 82,61%, dan recall sebesar 91,10% dimana nilai k yang digunakan adalah k=7. Sedangkan untuk k=27 algoritme knn mempunyai nilai akurasi lebih besar dengan nilai 80,51%, presisi sebesar 81,03%, dan recall sebesar 96,58%.

Kata kunci : KNN, Naive Bayes, Beasiswa, Baznas Jabar

1. PENDAHULUAN

Penyelenggaraan pendidikan yang berkualitas memerlukan biaya yang cukup tinggi, oleh karenanya semua satuan pendidikan pada setiap pelajar yang orang tuanya tidak mampu mendanai pendidikannya berhak mendapatkan dana pendidikan dan bagi mereka yang mempunyai prestasi berhak mendapatkan beasiswa[1]. Tujuan dari beasiswa bantuan adalah mendanai mahasiswa yang kurang beruntung atau kurang mampu tetapi masih mempunyai prestasi. Pemberi beasiswa umumnya memberikan sejumlah perhitungan pada kesulitannya, misalnya, seperti gaji orang tua, jumlah saudara kandung yang juga sedang menempuh pendidikan, kebutuhan target penerima beasiswa, living cost, dan sebagainya[2].

Besarnya biaya pendidikan pada tingkat perguruan tinggi dapat menjadi pandangan khusus bagi pihak pemerintah, karena untuk mengembangkan kualitas sumber daya manusia di Indonesia nyatanya memerlukan biaya yang cukup tinggi. Jumlah keseluruhan biaya pendidikan itu adalah biaya kepada sekolah seperti kebutuhan peserta didik, uang saku dan transport. Dengan adanya beasiswa yang tersedia dari lembaga ataupun komite yang lainnya, kesempatan untuk sekolah di perguruan tinggi menjadi lebih besar khususnya kepada mereka yang terkendala dalam hal biaya.

Beasiswa Baznas Jabar adalah beasiswa resmi yang dibuat oleh Lembaga Beasiswa Baznas Jawa Barat (LBB Jabar) dari Divisi Pendistribusian dan Pendayagunaan Badan Amil Zakat Nasional Jawa Barat (Baznas Jabar) yang ditujukan untuk mahasiswa pada perguruan tinggi[3]. Beasiswa ini baru diadakan pada tahun 2020 dimana jumlah pendaftar jauh melebihi dengan jumlah alokasi yang disediakan. Akibatnya jumlah pendaftar masuk berlebihan dan mengakibatkan proses seleksi menjadi tidak efisien dan memakan banyak waktu serta pengambilan keputusan yang subjektif. Dengan menggunakan algoritme klasifikasi data mining terhadap data yang telah ada akan diketahui pola pemetaan karakteristik pemohon ke keputusan yang diambil sehingga proses seleksi akan menjadi lebih mudah, konsisten dan dapat menghindari unsur subjektifitas dari pengambil keputusan.

Beberapa algoritme klasifikasi pada *data mining* yang banyak digunakan untuk menentukan suatu label pada *dataset* diantaranya adalah algoritme *k-nearest neighbor* (knn) dan *naïve bayes*. Kedua algoritme tersebut merupakan algoritme yang marak digunakan pada penelitian pengklasifikasian data[4]. Selain itu juga termasuk

dalam 10 algoritme *data mining* yang banyak digunakan dalam penelitian *machine learning*[5]. Pada penelitian sebelumnya terhadap pengklasifikasian penerima beasiswa, hasil yang didapat dipengaruhi oleh jumlah *data training* dan kategori, juga dalam penelitian tersebut digunakan empat kategori[6]. Sedangkan data penerimaan beasiswa LBB Jabar terdapat tujuh kategori dan dengan atribut yang berbeda.

Oleh karena itu pada penelitian ini akan melakukan perbandingan antara algoritme knn dengan *naïve bayes* disertai dengan analisis pada penerapan terhadap *dataset* penerima Beasiswa Baznas Jabar yang diadakan oleh LBB Jabar. Variabel yang digunakan pada penelitian ini adalah beberapa persyaratan yang sudah ditentukan dan dikeluarkan oleh LBB Jabar sebagai penilaian lolos atau tidaknya para pelamar beasiswa. Dengan adanya penelitian ini didapatkan algoritme mana yang memiliki kinerja terbaik dan menjadi motivasi untuk menerapkan algoritme tersebut pada seleksi penerimaan beasiswa di LBB Jabar.

2. TINJAUAN PUSTAKA DAN TEORI

2.1. Tinjauan Pustaka

Pada penelitian sebelumnya telah dilakukan studi terkait perbandingan kedua metode. Pada penelitian tersebut pengukuran dari tingkat akurasi data untuk algoritme knn diperoleh akurasi sebesar 100% sedangkan pada algoritme *naïve bayes* sebesar 99,89%. Namun pada penelitian tersebut hanya menggunakan empat variabel, dimana jika variabel bertambah maka akan mendapatkan hasil yang lebih optimal[6]. Pada penelitian lain dilakukan analisis akurasi untuk menentukan waktu lulus mahasiswa dengan kelas binomial antara algoritme *naïve bayes*, knn, support vector machine, dan decision tree. Hasil dari penelitian tersebut algoritme *naïve bayes* memiliki tingkat akurasi yang paling tinggi dengan nilai akurasi 76,77% sedangkan algoritme knn berada dibawah *naïve bayes* dengan nilai akurasi 68,05%. Hal yang sama pada penelitian tersebut adalah jumlah variabel yang dinilai bisa meningkatkan nilai akurasi jika variabelnya bertambah[7]. Sedangkan pada penelitian perbandingan klasifikasi Beasiswa Toyota Astra menggunakan *k-nearest neighbor classifier* dan *naïve bayes* sebagai penentu metode klasifikasi pada sistem pendukung keputusan, dari penelitian ini didapatkan algoritma *k-nearest neighbor classifier* mendapatkan nilai akurasi 90,26% yang lebih besar dari *naïve bayes* dengan nilai akurasi sebesar 70,77%[8].

Berdasarkan pada penelitian sebelumnya maka pada penelitian kali ini akan menganalisis perbandingan kinerja antara algoritme knn dan naïve bayes terhadap data penerimaan Beasiswa Baznas Jabar dan mengetahui algoritme mana yang cocok digunakan pada data tersebut. Kebaharuan yang ditekankan pada penelitian ini terletak pada variabel yang lebih variatif dimana terdapat nilai akademik yaitu nilai indeks prestasi kumulatif dan non akademik seperti gaji orang tua, kelengkapan berkas yang nantinya akan ditransformasikan ke dalam bentuk numerik.

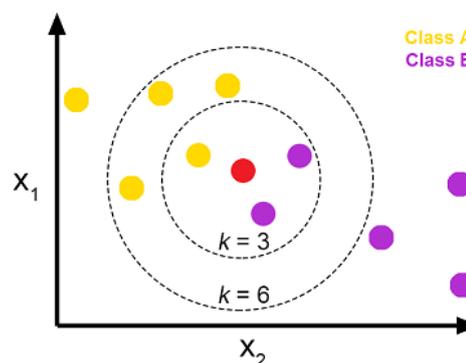
2.2. Metode Klasifikasi

Klasifikasi adalah salah satu teknik imperatif data mining yang menggunakan *training dataset* dan *testing dataset* dan memperkirakan hasilnya. Klasifikasi sebagai teknik penting dari data mining yang memanfaatkan sejumlah contoh yang terdiri dari atribut-atribut yang mungkin bersifat kategoris atau kontinu. Klasifikasi dilakukan pada atribut kelas target, metode klasifikasi juga menjelaskan bahwa metode ini berfungsi untuk mencari *classifier* untuk memprediksi label kelas target[9].

Klasifikasi adalah metode *data mining* yang paling terkenal dan paling banyak digunakan[10]. Proses klasifikasi secara bersamaan dibagi menjadi dua fase yaitu *training* dan *testing*. Proses *testing* adalah ketika model klasifikasi dibangun dari set pelatihan, dan *testing*, ketika model dievaluasi pada set pengujian. Tugas metode klasifikasi adalah menggeneralisasikan struktur yang diketahui untuk bisa diterapkan pada data baru dan algoritme yang termasuk didalamnya adalah *decision tree*, *k-nearest neighbor*, *support vector machine*, *naïve bayes*, dan *neural networks*[11].

2.3. K-Nearest Neighbor

K-Nearest Neighbour ialah suatu metode untuk menerapkan klasifikasi terhadap objek bersumber pada *data training* yang jaraknya sangat dekat dengan titik objek tersebut. *Data training* diproyeksikan ke ruang berukuran banyak, dimana tiap-tiap ukuran merepresentasikan fitur dari data. Ruang ini dipecah jadi bagian-bagian bersumber pada klasifikasi data pendidikan. Nilai *k* yang terbaik untuk algoritme ini bergantung pada data, secara biasanya, nilai *k* yang besar akan mengurangi dampak noise pada klasifikasi, namun membuat batas antara klasifikasi jadi lebih kabur[12]. Gambar 1 merupakan penggambaran dari algoritme *k-nearest neighbor*.



Gambar 1. Ilustrasi Algoritme Knn

K-nearest neighbor merupakan metode yang menggunakan algoritme *supervised*, juga merupakan algoritme yang menggunakan klasifikasi terhadap objek berdasarkan data terjauh (terhadap objek). Secara singkat algoritme knn menghitung jarak antara data uji dan data latih dengan menggunakan pengukuran jarak tertentu[13]. Untuk mencari kedekatan biasanya digunakan beberapa ukuran jarak. Dalam penelitian ini jarak *Euclidean* digunakan sebagai ukuran jarak.

$$d_i = \sqrt{\sum_i^p (x_{2i} - x_{1i})^2} \quad (1)$$

Dimana pada rumus *Euclidean Distance* tersebut :

d = jarak

p = dimensi data

x_{2i} = data latih

x_{1i} = data uji

2.4. Naïve Bayes

Naïve Bayes adalah metode pengklasifikasian probabilistik berdasarkan Teorema Bayes, dimana probabilitas kelasnya ditentukan untuk setiap sampel dan diasumsikan bahwa semua atributnya adalah independen. Pada klasifikasi *naïve bayes* pengklasifikasiannya akan memilih klasifikasi yang paling mirip dengan hasil proses perhitungannya[14].

Naïve bayes menggunakan cabang matematika yang dikenal dengan teori probabilitas untuk menemukan peluang terbesar pada kemungkinan klasifikasinya, dengan melihat frekuensi setiap klasifikasi dalam data latih. Metode ini hanya membutuhkan sedikit data latih untuk menentukan estimasi parameter yang diperlukan dalam proses klasifikasi[15]. Bentuk umum

teorema *bayes* dapat dilihat pada persamaan sebagai berikut :

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)} \quad (2)$$

Dimana pada rumus *Naïve Bayes* tersebut :

- X = Data dengan kelas yang tidak diketahui
- H = Hipotesa data X yang merupakan kelas tertentu
- $P(H|X)$ = Peluang hipotesa H berdasarkan kondisi X
- $P(H)$ = Peluang dari hipotesa H
- $P(X|H)$ = Peluang X berdasarkan kondisi hipotesa H
- $P(X)$ = Peluang dari X

2.5. Confusion Matrix

Confusion matrix merupakan metode yang digunakan untuk mendapatkan akurasi dengan sebuah perhitungan yang digunakan pada teknik penambangan data[16]. *Confusion matrix* digunakan untuk menghitung jumlah observasi pada kelas yang salah dan kelas yang benar dari model klasifikasi masing-masing, kemudian menampilkan hasilnya dalam sebuah tabel[17].

Confusion matrix dapat diartikan sebagai alat yang memiliki fungsi untuk melakukan analisis apakah *classifier* baik dalam mengenali tupel dari kelas yang berbeda[18]. Perhitungan *Confusion matrix* ditunjukkan pada Tabel 1.

TABEL 1. CONFUSION MATRIX

		<i>True Value</i>	
		<i>True</i>	<i>False</i>
<i>Forecast Value</i>	<i>True</i>	<i>TP (True Positive)</i>	<i>FP (False Positive)</i>
	<i>False</i>	<i>FN (False Negative)</i>	<i>TN (True Negative)</i>

Pada tabel tersebut :

- *TP (True Positive)* : Jumlah data dengan nilai aktual positif dan nilai prediksi positif.
- *FP (False Positive)* : Jumlah data dengan nilai aktual negatif dan nilai prediksi positif.
- *FN (False Negative)* : Jumlah data dengan nilai aktual positif dan nilai prediksi negatif.
- *TN (True Negative)* : Jumlah data dengan nilai aktual negatif dan nilai prediksi negatif.

Untuk pengukurannya adalah *accuracy*, *precision*, dan *recall*. *Accuracy* adalah ukuran kinerja yang paling intuitif untuk menghitung rasio pengamatan yang diprediksi secara tepat pada seluruh dataset, dengan rumus :

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (3)$$

Precision adalah rasio positif yang diprediksi secara tepat dari rasio yang diamati terhadap total prediksi pengamatan positif.

$$Precision = \frac{TP}{(TP + FP)} \quad (4)$$

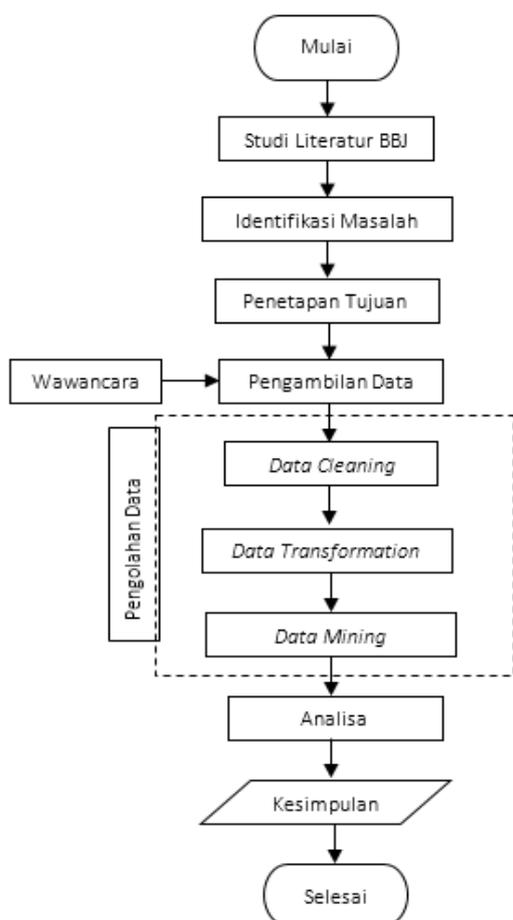
Recall adalah tingkat keberhasilan atau kemampuan sistem dalam menemukan informasi dari *dataset*.

$$Recall = \frac{TP}{(TP + FN)} \quad (5)$$

3. METODOLOGI PENELITIAN

3.1. Skema Alur Penelitian

Proses tahapan pada penelitian ini digambarkan pada Gambar 2. Pada diagram alir tersebut studi literatur dilakukan dengan membaca dan memahami artikel dan informasi umum terkait beasiswa baznas jabar. Studi literatur dilakukan agar dapat mengidentifikasi masalah dan tujuan dengan mudah serta dapat melanjutkan wawancara untuk mendapatkan dataset penelitian. Setelah itu dilakukan tahapan pengolahan data dimana pada proses penambangan data diproses dengan *Rapidminer*.



Gambar 2. Diagram Alir Penelitian

3.2. Pengumpulan Data

Pengumpulan data pada penelitian ini dilakukan dengan dua metode yaitu :

1. Studi pustaka, membaca artikel-artikel, dan literasi terkait beasiswa baznas jabar.
2. Wawancara, pengumpulan data diperoleh setelah wawancara langsung dengan Kepala LBB Jabar sehingga diketahui data para pelamar beasiswa.

Data yang diperoleh adalah data pelamar Beasiswa Baznas Jabar pada tahun 2020. Target data tepatnya adalah data lengkap pelamar beasiswa.

3.3. Analisa Data

Analisa data yang diterapkan pada penelitian ini menggunakan metode analisis deskriptif dimana data yang diperoleh adalah data kuantitatif. Pengetahuan pada data didapatkan

dengan melihat kinerja data di masa lalu. Bentuk data yang didapat yaitu tabel dengan *record* sebanyak 1.139 dan kolom sebanyak 21. Kolom terakhir setelah atribut terdapat label sebagai penentu kelulusan pengaju beasiswa. Pada data tersebut dapat dilihat pengaruh persyaratan beasiswa terhadap lolos atau tidaknya para pengaju beasiswa. Pengolahan data untuk mencari pengetahuan dengan metode klasifikasi dilakukan setelah melalui tahap *preprocessing* dan *transformation*. Metode klasifikasi yang dipakai untuk mengetahui algoritme yang paling baik adalah antara knn dan *naive bayes*. Sedangkan untuk pengukuran kinerja kedua algoritme dengan metode *confusion matrix*, pengukuran dilakukan agar diketahui kinerja masing-masing algoritme.

3.4. Data Cleaning

Untuk menyediakan akses ke data yang akurat dan konsisten serta meningkatkan kualitas data, *data cleaning* perlu dilakukan untuk pendeteksian dan penghapusan kesalahan dan inkonsistensi dari data. Proses yang dilakukan pada penelitian ini adalah untuk menghapus :

- a. Atribut yang tidak relevan

Pada tabel 2 data mentah yang diperoleh terdapat 20 atribut dimana beberapa atribut tersebut banyak yang tidak relevan dengan dengan kelulusan pelamar beasiswa dan juga tidak diperhitungkan sebagai penilaian kelulusan beasiswa.

TABEL 2. ATRIBUT DATA MENTAH

No	Atribut	Keterangan
1	Timestamp	Waktu pengisian formulir
2	Pas foto	Foto pengaju beasiswa
3	Tempat lahir	Tempat lahir pengaju
4	Tanggal Lahir	Tanggal lahir pengaju
5	Alamat	Alamat pengaju
6	Asal Kampus	Asal kampus pengaju
7	Angkatan	Tahun awal masuk studi
8	Jurusan/Fakultas	Jurusan studi pengaju
9	IPK	Besaran ipk pengaju
10	Nama Ayah	Nama ayah pengaju
11	Nama Ibu	Nama ibu pengaju
12	Alamat Orang tua	Alamat orang tua pengaju
13	Gaji Orangtua	Besaran gaji orang tua pengaju
14	Uang saku perbulan	Uang saku pengaju
15	Prestasi	Prestasi pengaju
16	Riwayat Organisasi	Riwayat organisasi pengaju
17	Keahlian	Keahlian yang dimiliki
18	Beasiswa Lain	Beasiswa lain yang dimiliki
19	Nama Beasiswa Lain	Nama beasiswa lain yang dimiliki
20	Kelengkapan berkas	Kelengkapan berkas syarat pengajuan

Setelah melalui proses wawancara hanya ada 7 atribut yang dijadikan penilaian untuk kelulusan pengajuan beasiswa, atribut tersebut adalah :

1. Alamat
2. Indeks Prestasi Kumulatif (IPK)
3. Gaji Orang Tua
4. Uang Saku Perbulan
5. Prestasi
6. Beasiswa Lain
7. Kelengkapan Berkas

b. Duplikasi data

Langkah kunci dalam aktivitas pembersihan data adalah mengidentifikasi duplikat terlepas banyak sekali manifestasi dari contoh yang sama di sumber yang berbeda. Dari sebanyak 1.139 record ditemukan banyak data yang sama pada setiap pengaju beasiswa. Total hasil record setelah penghapusan terdapat 975 record.

3.5. Data Transformation

Data transformation pada tahap ini, data akan ditransformasikan ke dalam format yang sesuai untuk kebutuhan proses data mining. Data akan ditransformasikan kedalam bentuk integer berdasarkan range yang sudah ditentukan sesuai penilaian dari tim LBB Jabar. Beberapa algoritme data mining berperforma lebih baik pada data dengan bentuk *integer*, sementara yang lain lebih baik pada data *boolean*. Pada ketujuh atribut terpilih akan ditransformasikan sebagai berikut :

a. Atribut Alamat

Pada atribut ini akan ditransformasikan ke dalam bentuk binomial dimana jika pengaju bukan warga jawa barat maka akan mempengaruhi target *class*.

TABEL 3. TRANSFORMASI ATRIBUT ALAMAT

No	Kategori	Keterangan
1	1	Warga Jawa Barat
2	0	Bukan warga Jawa Barat

b. Atribut IPK

Atribut IPK akan ditransformasikan dari besaran ipk pengaju kedalam bentuk kategori integer.

TABEL 4. TRANSFORMASI ATRIBUT IPK

No	Kategori	Keterangan
1	5	IPK 3,70 – 4,00
2	4	IPK 3,50 – 3,69
3	3	IPK 3,30 – 3,49
4	2	IPK 3,00 – 3,29
5	1	IPK <3,00 (kurang dari)

c. Atribut Gaji Orang Tua

Pada atribut ini akan ditransformasikan ke dalam bentuk kategori integer dari rentang gaji orang tua.

TABEL 5. TRANSFORMASI ATRIBUT GAJI ORANG TUA

No	Kategori	Keterangan
1	5	Gaji <500.000 (kurang dari)
2	4	Gaji 500.001-1.000.000
3	3	Gaji 1.000.001-3.000.000
4	2	Gaji 3.000.001-5.000.000
5	1	Gaji >5.000.000 (lebih dari)

- d. Atribut Uang Saku Perbulan
 Pada atribut ini akan ditransformasikan ke dalam bentuk kategori integer dari rentang uang saku perbulan pengaju beasiswa.

TABEL 6. TRANSFORMASI ATRIBUT UANG SAKU PERBULAN

No	Kategori	Keterangan
1	5	Uang saku <200.000 (kurang dari)
2	4	Uang saku 200.000-500.000
3	3	Uang saku 500.001-800.000
4	2	Uang saku 800.001-1.000.000
5	1	Uang saku >1.000.000 (lebih dari)

- e. Atribut Prestasi
 Pada atribut ini akan ditransformasikan ke dalam bentuk kategori integer dari kategori tingkatan prestasi pengaju beasiswa.

TABEL 7. TRANSFORMASI ATRIBUT PRESTASI

No	Kategori	Keterangan
1	5	Tingkat internasional
2	4	Tingkat nasional
3	3	Tingkat provinsi
4	2	Tingkat kabupaten/kota
5	1	Lainnya

- f. Atribut Beasiswa Lain
 Pada atribut ini akan ditransformasikan ke dalam bentuk binomial. Apabila pengaju memiliki beasiswa lain akan mempengaruhi target class.

TABEL 8. TRANSFORMASI ATRIBUT BEASISWA LAIN

No	Kategori	Keterangan
1	1	Ada beasiswa lain
2	0	Tidak ada beasiswa lain

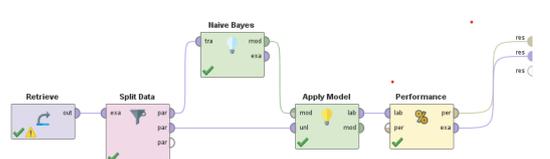
- g. Atribut Kelengkapan Berkas
 Pada atribut ini akan ditransformasikan ke dalam bentuk binomial. Apabila pengaju melengkapi berkas pendaftaran akan mempengaruhi target class.

TABEL 9. TRANSFORMASI ATRIBUT KELENGKAPAN BERKAS

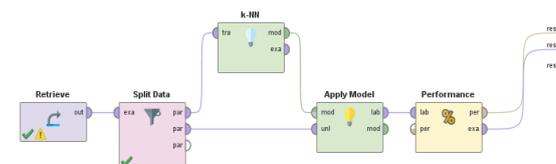
No	Kategori	Keterangan
1	1	Warga Jawa Barat
2	0	Bukan warga Jawa Barat

3.6. Data Mining

Pada proses *data mining* atau penambangan data dilakukan pada *tools* rapidminer dengan menerapkan algoritme knn dan *naive bayes*. Data yang diolah sebelumnya dilakukan proses pelabelan pada kolom kelulusan pengaju beasiswa dan perubahan jenis atribut binomial pada data *boolean*.



Gambar 3. Pemodelan Dengan Naive Bayes



Gambar 4. Pemodelan Dengan KNN

1. *Retrieve* digunakan untuk pemanggilan data yang akan diuji oleh rapidminer.
2. *Split Data* digunakan untuk memisahkan antara data uji dan data latih. Dengan rasio 4 banding 1, 80% pada dataset dijadikan data latih dan 20% dijadikan data uji dimana jumlah data uji sebanyak 195 record.
3. *Naive Bayes* merupakan algoritme yang diuji dengan rapidminer dimana sebelumnya label sudah disesuaikan menjadi binomial untuk bisa dibaca oleh algoritme *naive bay*
4. *KNN* merupakan algoritme kedua yang diuji oleh rapidminer. Pada algoritme knn akan diuji dengan nilai K yang berbeda.
5. *Apply Model* ditujukan untuk mendapatkan prediksi pada data yang tidak terlihat atau untuk mengubah data dengan menerapkan model.
6. *Performance* adalah salah satu operator pada rapidminer yang bisa digunakan untuk evaluasi kinerja model klasifikasi.

4. HASIL DAN PEMBAHASAN

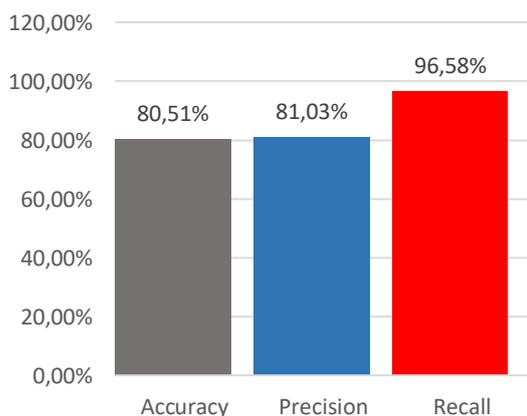
Hasil uji dan evaluasi algoritme akan di tampilkan dengan hasil perhitungan *confusion matrix* pada setiap algoritme yang diuji. Tahap evaluasi dilakukan agar diketahui seberapa baik kedua algoritme diterapkan pada dataset BBJ 2020. Dengan begitu perbandingan antara algoritme dapat dilakukan dengan mudah.

4.1. Hasil Uji Algoritme Naïve Bayes

Hasil uji yang telah dibuat menggunakan algoritme *naïve bayes* adalah nilai akurasi yang didapat sebesar 80%, nilai presisi sebesar 82,82%, dan nilai recall sebesar 92,47%.

accuracy: 80,00%			
	true 0	true 1	class precision
pred. 0	135	28	82,82%
pred. 1	11	21	65,62%
class recall	92,47%	42,86%	

Gambar 5. Hasil Uji *Naïve Bayes* Pada Rapidminer



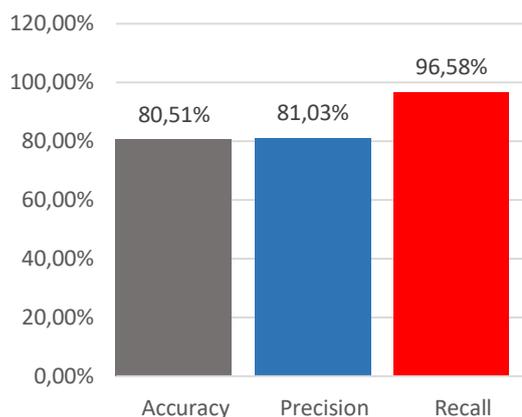
Gambar 6. Visualisai Hasil Uji *Naïve Bayes*

4.2. Hasil Uji Algoritme KNN

Untuk hasil uji yang telah dibuat menggunakan algoritme knn dengan nilai k=7 didapat hasil akurasi yang sebesar 78,97%, nilai presisi sebesar 82,61%, dan nilai recall sebesar 91,10%.

accuracy: 78,97%			
	true 0	true 1	class precision
pred. 0	133	28	82,61%
pred. 1	13	21	61,76%
class recall	91,10%	42,86%	

Gambar 7. Hasil Uji Knn k=7 Pada Rapidminer

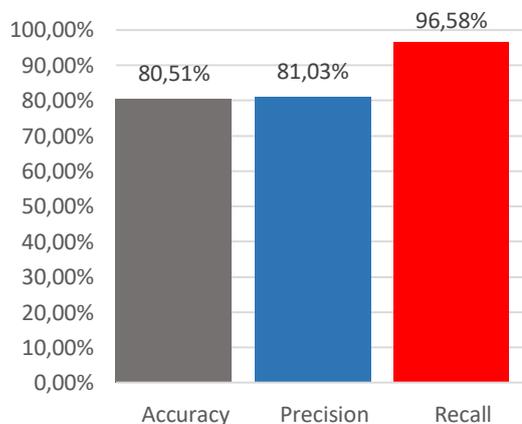


Gambar 8. Visualisasi Hasil Uji Knn K=7

Sedangkan pada algoritme knn hasil uji yang dibuat dengan menerapkan nilai k=27 mendapatkan nilai akurasi yang lebih besar dari hasil uji algoritme *naïve bayes* dengan nilai akurasi sebesar 80,51%, nilai presisi sebesar 81,03%, dan nilai recall sebesar 96,58%.

accuracy: 80,51%			
	true 0	true 1	class precision
pred. 0	141	33	81,03%
pred. 1	5	16	76,19%
class recall	96,58%	32,65%	

Gambar 9. Hasil Uji Knn K=27 Pada Rapidminer



Gambar 10. Visualisasi Hasil Uji Knn K=27

5. Kesimpulan Dan Saran

5.1. Kesimpulan

Pada penelitian kali ini berhasil menarapkan algoritme knn dengan *naïve bayes* dengan menggunakan *tools* rapidminer terhadap data penerimaan beasiswa BBJ 2020. Hasil yang diperoleh adalah algoritme *naïve bayes* mempunyai akurasi yang sedikit lebih besar dari knn, dengan penerapan nilai k=7 pada algoritme knn. Tetapi untuk penerapan nilai k=27 pada algoritme knn didapatkan hasil akurasi yang lebih

besar dari *naïve bayes*. Pada algoritme knn banyaknya record akan mempengaruhi relevannya nilai k yang diterapkan. Untuk dataset dengan jumlah *record* mencapai 1000 lebih, algoritme knn bisa menjadi alternatif yang baik untuk digunakan.

5.2. Saran

Untuk mendapatkan hasil yang lebih baik perlu ada beberapa hal yang bisa diterapkan pada penelitian selanjutnya.

1. Algoritme yang diterapkan pada dataset BBJ 2020 perlu percobaan pada algoritme lain selain *naïve bayes* dan knn.
2. Pemodelan pada *tools* rapidminer dapat mempengaruhi hasil uji, misalkan diterapkannya operator validasi lain.
3. Seleksi atribut pada pemodelan akan sangat mendapatkan hasil uji yang berbeda dari algoritme *naïve bayes* dan knn. Dengan adanya seleksi atribut maka akan ditemukan atribut yang lebih relevan dan signifikan terhadap uji dataset.

6. UCAPAN TERIMA KASIH

Terimakasih kepada pihak LBB Jabar atas izin penggunaan data serta ikut berpartisipasi dan mendukung penelitian ini agar bisa terlaksana dengan baik.

Daftar Pustaka:

- [1] Rismayanti, R., *IMPLEMENTASI ALGORITMA C4. 5 UNTUK MENENTUKAN PENERIMA BEASISWA DI STT HARAPAN MEDAN*. JURNAL MEDIA INFOTAMA, 2016. **12**(2).
- [2] ROHMAWATY, I.M., *PENGARUH PENDISTRIBUSIAN DANA ZAKAT DALAM BENTUK BEASISWA BAZNAS TERHADAP PRESTASI MAHASISWA Studi: Mahasiswa Penerima Beasiswa BAZNAS KAB. Serang di UIN Sultan Maulana Hasanuddin Banten Tahun 2016*. 2018, Universitas Islam Negeri" Sultan Maulana Hasanuddin" Banten.
- [3] Jabar, L.B.B. *Apa Itu Lembaga Beasiswa BAZNAS*. 2020; Available from: <https://beasiswa.baznas.go.id/>.
- [4] Maricar, M.A. and D. Pramana, *Perbandingan Akurasi Naïve Bayes dan K-Nearest Neighbor pada Klasifikasi untuk Meramalkan Status Pekerjaan Alumni ITB STIKOM Bali*. Jurnal Sistem Dan Informatika (JSI), 2019. **14**(1): p. 16-22.
- [5] Kamila, V.Z. and E. Subastian, *KNN vs Naive Bayes Untuk Deteksi Dini Putus Kuliah Pada Profil Akademik Mahasiswa*. Jurnal Rekayasa Teknologi Informasi (JURTI), 2019. **3**(2): p. 116-122.
- [6] Sumiah, A. and N. Mirantika, *Perbandingan Metode K-Nearest Neighbor dan Naive Bayes untuk Rekomendasi Penentuan Mahasiswa Penerima Beasiswa pada Universitas Kuningan*. Buffer Informatika, 2020. **6**(1): p. 1-14.
- [7] Widaningsih, S., *Perbandingan Metode Data Mining Untuk Prediksi Nilai Dan Waktu Kelulusan Mahasiswa Prodi Teknik Informatika Dengan Algoritma C4, 5, Naïve Bayes, Knn Dan Svm*. Jurnal Tekno Insentif, 2019. **13**(1): p. 16-25.
- [8] Lestari, D.R.D., *Perbandingan Klasifikasi Beasiswa Toyota Astra Menggunakan K-Nearest Neighbor Classifier Dan Naïve Bayes Sebagai Penentu Metode Klasifikasi Pada Sistem Pendukung Keputusan Penerimaan Beasiswa Toyota Astra (Studi Kasus: Institut Teknologi Sepuluh Nopember)*. 2017, Institut Teknologi Sepuluh Nopember.
- [9] Abro, M., H. Nawaz, and W. Abro, *Performance Analysis of Dissimilar Classification Methods using RapidMiner*. Sindh University Research Journal-SURJ (Science Series), 2016. **48**(1).
- [10] Çiğşar, B. and D. Ünal, *Comparison of data mining classification algorithms determining the default risk*. Scientific Programming, 2019. **2019**.
- [11] Neelamegam, S. and E. Ramaraj, *Classification algorithm in data mining: An overview*. International Journal of P2P Network Trends and Technology (IJPTT), 2013. **4**(8): p. 369-374.
- [12] Dina, N.Z. and R.S. Marjianto, *PREDIKSI PENENTUAN PENERIMA BEASISWA DENGAN METODE KNEAREST NEIGHBOURS (Studi Kasus: Program Studi Sistem Informasi Fakultas Vokasi Universitas Airlangga)*. InfoTekJar: Jurnal Nasional Informatika dan Teknologi Jaringan, 2018. **2**(2): p. 135-139.
- [13] Windana, F., A. Handaratri, and M.T.A. Zaen, *IMPLEMENTASI METODE K-NEAREST NEIGHBOR UNTUK PENGENALAN BUAH MURBEI DENGAN KORELASI KANDUNGAN ANTOSIANIN*. Jurnal Informatika dan Rekayasa Elektronik, 2021. **4**(1): p. 78-86.

- [14] Rasjid, Z.E. and R. Setiawan, *Performance comparison and optimization of text document classification using k-NN and naïve bayes classification techniques*. Procedia computer science, 2017. **116**: p. 107-112.
- [15] Gerhana, Y., et al. *Comparison of naïve Bayes classifier and C4. 5 algorithms in predicting student study period*. in *Journal of Physics: Conference Series*. 2019. IOP Publishing.
- [16] Mutawalli, L., M.T.A. Zaen, and W. Bagye, *Klasifikasi Teks Sosial Media Twitter Menggunakan Support Vector Machine (Studi Kasus Penusukan Wiranto)*. Jurnal Informatika dan Rekayasa Elektronik, 2019. **2**(2): p. 43-51.
- [17] Chen, Z., et al., *The Lao text classification method based on KNN*. Procedia Computer Science, 2020. **166**: p. 523-528.
- [18] Yulianto, A., P. Sukarno, and N.A. Suwastika. *Improving adaboost-based intrusion detection system (IDS) performance on CIC IDS 2017 dataset*. in *Journal of Physics: Conference Series*. 2019. IOP Publishing.