

PENERAPAN *DATA MINING* UNTUK PENENTUAN PENERIMA BEASISWA DENGAN METODE *K-NEAREST NEIGHBOR* (K-NN)

Fiqriany Karepesina¹, Lisna Zahrotun²

^{1,2} Teknik Informatika, Fakultas Teknologi Industri, Universitas Ahmad Dahlan

Informasi Makalah

Dikirim, 24 November 2020
Direvisi, 30 September 2022
Diterima, 30 Desember 2022

Kata Kunci:

Data Mining
Pengelolaan Beasiswa
K-Nearest Neighbor

Keyword:

Data Mining
Scholarship Management
K-Nearest Neighbor

INTISARI

Dalam beberapa instansi pengelolaan beasiswa masih menggunakan *microsoft excel* dan pemilihan penerima beasiswa masih menggunakan seleksi administratif secara manual. Salah satu pengolahan data dalam jumlah yang besar adalah *data mining*. Oleh karena itu penelitian ini bertujuan untuk menerapkan data mining dengan metode *k-nearest neighbor* (K-NN) untuk penentuan penerima beasiswa. Metode pengumpulan data dengan metode data private, studi literatur, dan wawancara. Tahapan *data mining* yaitu *cleaning, selection, transformation, knowledge discovery, pattern evaluation, knowledge presentation*. Pengujian sistem yaitu dengan menggunakan *confusion matrix* untuk mengetahui nilai akurasi. Hasil dari penelitian ini adalah klasifikasi untuk menentukan penerima beasiswa menggunakan metode *k-nearest neighbor* dengan jumlah data yang digunakan sebanyak 555 data training dan 115 data testing. Hasil pengujian yang dilakukan dengan *confusion matrix* dan kurva ROC (*Receiver Operation Characteristic*) mendapatkan hasil akurasi terbaik sebesar 77% dengan nilai AUC (*Area Under Curve*) 0,90. Dari hasil pengujian ini dengan nilai AUC yang berada diantara rentang 0.80 – 0.90, maka metode K-NN dalam penentuan beasiswa ini termasuk dalam kategori *good classification* (sangat baik).

ABSTRACT

In some agencies, scholarship management still uses Microsoft Excel and the selection of scholarship recipients still uses manual administrative selection. One of the processing of large amounts of data is data mining. Therefore, this study aims to apply data mining with the k-nearest neighbor (K-NN) method for determining scholarship recipients. Data collection methods using private data methods, literature studies, and interviews. The stages of data mining are cleaning, selection, transformation, knowledge discovery, pattern evaluation, knowledge presentation. System testing is by using a confusion matrix to find out the accuracy value. The results of this study are classifications to determine scholarship recipients using the k-nearest neighbor method with the number of data used as much as 555 training data and 115 testing data. The test results carried out with confusion matrix and ROC (Receiver Operation Characteristic) curves got the best accuracy results of 77% with an AUC (Area Under Curve) value of 0.90. From this award result with an AUC value that is between the range of 0.80 – 0.90, the K-NN method in determining this scholarship is included in the category of good classification (very good).

Korespondensi Penulis:

Lisna Zahrotun
Teknik Informatika, Fakultas Teknologi Industri, Universitas Ahmad Dahlan
Jl. Ringroad Selatan, Kragilan, Tamanan, Kec. Banguntapan, Bantul, Daerah Istimewa Yogyakarta 55191.
Email: lisna.zahrotun@tif.uad.ac.id

1. PENDAHULUAN

Beasiswa merupakan pemberian berupa bantuan keuangan maupun pendidikan yang diberikan perorangan, mahasiswa atau pelajar yang digunakan demi keberlangsungan pendidikan yang ditempuh bagi yang memiliki prestasi dibidang akademik, non-akademik dan kemampuan ekonominya lemah, serta telah memenuhi syarat-syarat yang ditentukan pihak pemberi beasiswa [1].

Dalam proses pengolahan data dan seleksi penerima beasiswa pada beberapa instansi menggunakan *software Microsoft Office Excel*, sehingga selama ini untuk mendapatkan hasil penerima beasiswa membutuhkan waktu satu bulan karena penyeleksian harus sesuai dengan ketentuan atau syarat dari setiap jenis beasiswa agar penerima beasiswa tepat sasaran, efisien (tepat waktu), dan efektif (akurat). Salah satu pengolahan data dalam jumlah yang besar adalah *data mining*. Data mining merupakan suatu proses yang menggunakan teknik statistik, matematika, kecerdasan tiruan, dan machine-learning untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar [2]. Beberapa metode kerap digunakan dalam data mining salah satunya adalah metode *k-nearest neighbor* (k-nn) yaitu metode yang melakukan klasifikasi berdasarkan kedekatan lokasi (jarak) suatu data dengan data lain. Metode *k-nearest neighbor* (k-nn) memiliki kelebihan antara lain teknik yang sangat sederhana, efisien dan efektif dalam bidang pengenalan pola, kategori teks, pengolahan objek dan lain-lain [3], metode yang cukup sederhana namun memiliki tingkat akurasi yang tinggi [4].

Beberapa penelitian sebelumnya Sistem Pendukung Keputusan Menggunakan metode K-NN dan ELECTRE dengan menggunakan sampel data 12 data [5]. Penelitian terkait penggunaan metode K-NN juga pernah dilakukan dalam penentuan beasiswa [6],[7], [8] dan untuk penentuan kesehatan bayi baru lahir[9]. Penelitian penentuan beasiswa yang dilakukan oleh [8] mendapatkan akurasi 90,90% dengan menggunakan metode K-Nearest Neighbor (K-NN) dengan variabel yang digunakan adalah penghasilan orangtua, tanggungan orangtua, ipk, nim, semester dan jumlah data 22. Penggunaan metode K-NN juga dilakukan oleh [3] dalam melakukan klasifikasi beasiswa peningkatan prestasi akademik (PPA) dan bantuan belajar mahasiswa (BBM) dengan variabel yang digunakan penghasilan orangtua ayah, penghasilan orangtua ibu, tanggungan orangtua, rata-rata ujian nasional, pekerjaan orangtua ayah, rata-rata nilai rapor, tagihan listrik, prestasi akademik, prestasi non akademik, pendidikan orangtua ayah, luas rumah, pajak bumi dan bangunan dan jumlah data 180 data beasiswa BBM dan 353 data beasiswa BBM dan menghasilkan nilai akurasi 88,33% dengan nilai *area under curva* (AUC) 0,925.

Dengan melihat permasalahan dari penerimaan beasiswa maka dalam penelitian ini dilakukan penerapan data mining dengan metode *k-nearest neighbor* (K-NN) untuk penentuan penerima beasiswa dengan variabel yang berbeda dengan penelitian sebelumnya [8],[3] yang digunakan yaitu penghasilan orangtua perbulan, tanggungan, ipk semester I, rata-rata nilai ujian nasional dan 555 data.

2. METODOLOGI

2.1 Data Mining

Data mining adalah suatu proses yang menggunakan teknik statistik, matematika, kecerdasan tiruan, dan *machine-learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar [2].

2.2 Klasifikasi

Klasifikasi atau taksonomi adalah proses menempatkan suatu objek atau konsep kedalam satu set kategori berdasarkan objek atau konsep yang bersangkutan [10] Klasifikasi merupakan proses penemuan model (fungsi) yang menggambarkan dan membedakan kelas data atau konsep yang bertujuan agar bisa digunakan untuk memprediksi kelas dari objek yang label kelasnya tidak diketahui [11]

2.3 Algoritma *K-nearest neighbor*

Algoritma *Nearest Neighbor* (kadang disebut *K-Nearest Neighbor/KNN*) merupakan algoritma yang melakukan klasifikasi berdasarkan kedekatan lokasi (jarak) suatu data dengan data lain. Prinsip sederhana yang diadopsi oleh algoritma knn adalah “jika suatu hewan berjalan seperti bebek, bersuara kwek-kwek seperti bebek, dan penampilannya seperti bebek, hewan itu mungkin bebek” [12]. Perhitungan jarak (kedekatan lokasi) antara data uji dengan data yang lain, parameter jarak yang digunakan adalah jarak Euclidean yang dihitung menggunakan rumus perhitungan dalam persamaan 1 sebagai berikut

$$d(x_i, x_j) = \sqrt{\sum_{r=1}^n (a_r(x_i) - a_r(x_j))^2} \quad (1)$$

Keterangan :

$d(x_i, x_j)$ = Jarak *Euclidean* (*Euclidean Distance*)

(x_i) = *record* ke- i

(x_j) = *record* ke j

(a_r) = data ke- ri

, $j = 1, 2, 3, \dots n$ (Mitsa, 2010)

2.4 Confusion Matrix

Metode ini menggunakan tabel matriks, jika data set hanya terdiri dari dua kelas, kelas yang satu dianggap sebagai positif dan yang lainnya *negative* [10]. *Confusion matrix* ditunjukkan pada Tabel 1.

Tabel 1. Model *Confusion Matrix*

Classification	Predicted Class	
	Class = YES	Class = No
<i>Observed Class</i> Class = YES	<i>a (True Positive-TP)</i>	<i>b (True Negative TN)</i>
Class = No	<i>c (False Positive-FP)</i>	<i>d (False Negative-FN)</i>

Keterangan :

- True Postive (TP)* : Jumlah record positif yang diklasifikasikan sebagai positif
- True Negative (TN)* : Jumlah record negatif yang diklasifikasikan sebagai negatif
- False Postive (FP)* : Jumlah record negatif yang diklasifikasikan sebagai positif
- False Negative (FN)* : Jumlah record positif yang diklasifikasikan sebagai negative

Berikut adalah persamaan model confusion matrix :

Nilai *accuracy* adalah proposisi jumlah prediksi yang benar. Dapat dihitung dengan menggunakan Persamaan 2.

$$Accuracy = \frac{(TP+TN)}{Total\ Sampel} \times 100\% \tag{2}$$

2.4 Kurva ROC (Receiver Operating Characteristic)

Metode ini menggunakan tabel matriks, jika data set hanya terdiri dari dua kelas, kelas yang satu dianggap sebagai positif dan yang lainnya *negative* [10]. *Confusion matrix* ditunjukkan pada Tabel 2.

Kurva ROC menunjukkan akurasi dan membandingkan klasifikasi secara visual. ROC mengekspresikan *confusion matrix*. ROC adalah grafik dua dimensi dengan *false positives* sebagai garis horisontal dan *true positives* sebagai garis vertikal. ROC memiliki tingkat nilai diagnosa yaitu [10].

- Akurasi bernilai 0.90 – 1.00 = *excellent classification*
- Akurasi bernilai 0.80 – 0.90 = *good classification*
- Akurasi bernilai 0.70 – 0.80 = *fair classification*
- Akurasi bernilai 0.60 – 0.70 = *poor classification*
- Akurasi bernilai 0.50 – 0.60 = *failure*

2.5 Tahapan penelitian

Tahapan dalam penelitian ini ditunjukkan dalam Gambar 1.



Gambar 1. Tahap penelitian

Dalam tahapan penelitian ini meliputi beberapa tahapan diantaranya :

1. *Load Data*
Proses dimana data dimasukkan dalam bentuk file excel.
2. *Data Cleaning*
Data yang di lakukan pembersihan adalah data yang dalam setiap atribut pada tabel yang bernilai ganda atau duplikasi, data kosong maka data tersebut akan di dihapus dari tabel *data mining* yang ada. Data yang dihapus adalah setiap data dari atribut : nim, nama, jenis kelamin, program studi, IPK semester I, rata-rata nilai ujian nasional, nomor hp, alamat, pekerjaan orang tua, tanggungan orangtua, penghasilan orangtua perbulan, keikutsertaan kegiatan mahasiswa baru.
3. *Data Integration*
Pada penelitian ini, data tidak memerlukan adanya integrasi data atau penggabungan data dari beberapa *database*. Karena data yang diperoleh berupa data *excel*.

4. Data Selection

Seleksi data merupakan tahap untuk memutuskan atribut-atribut apa saja yang akan digunakan pada tahap *data mining*. Pada penelitian ini atribut tabel pendaftar beasiswa akan diseleksi, sehingga hanya atribut tertentu yang akan digunakan. Atribut data pendaftar beasiswa antara lain : nim, nama, jenis kelamin, program studi, Ipk semester I, rata-rata nilai ujian nasional, nomor hp, alamat, pekerjaan orang tua, tanggungan orangtua, penghasilan orangtua perbulan, keikutsertaan kegiatan mahasiswa baru. Diseleksi sehingga hanya atribut: Penghasilan orangtua perbulan, tanggungan, ipk semester I, rata-rata nilai ujian nasional, nim yang dipilih.

5. Data Transformation

Pada penelitian ini tranformasi data, merubah data atribut *predictor* yang berbentuk *alphabet* kedalam bentuk *numerik* (angka). Atribut-atrubut yang dilakukan tranformasi pada tabel *data mining* adalah Ipk semester I, rata-rata nilai ujian nasional dan rata-rata penghasilan orangtua perbulan.

6. Teknik Data Mining dengan Metode KNN

Pada penelitian ini menggunakan algoritma *k-nearest neighbor* untuk penentuan penerima beasiswa. Algoritma *k-nearest neighbor* merupakan algoritma klasifikasi yang di bentuk ke kelas-kelas untuk dapat menemukan aturan yang di dapat dari proses algoritma *k-nearest neighbor*.

7. Pattern Evaluation

Pada tahapan ini dilakukan uji model pola untuk mengetahui nilai kesesuaian dan kesalahan atau *error* pada pola yang ditemukan. Pola yang sudah ditemukan kemudian direpresentasikan kepada pengguna agar mudah dimengerti.

8. Knowledge Presentation

Pola aturan algoritma *k-nearest neighbor* yang telah ditemukan kemudian dipresentasikan kepada pengguna agar mudah dipahami.

3. HASIL DAN PEMBAHASAN

3.1 Data Penelitian

Data yang digunakan dalam penelitian ini yaitu pendaftaran beasiswa pada sebuah instansi di Yogyakarta tahun 2015 sebanyak 555 mahasiswa sebagai data training dan 115 sebagai data testing.

3.2 Load Data

Proses dimana data training dimasukkan dalam bentuk file excel.

Tabel 1. Data mahasiswa

No	Nim	Nama	JP	PS	Ipk Sem I	Rata-rata UAN	Alamat	Pekerjaan	Tanggungan	Penghasilan	KKMB	Terima
1	1500004036	Mustafidah	P	PBI	4.00	86.7	Sumberejo Rt 01/02	Petani / Nelayan	5	1300000	Peserta P2K	Tidak
2	1500004042	Lita Dewi Andini	P	PBI	3.60	55.7	Jln. soekarno-hatta rt 08 no 032	Wiraswasta	2	2000000	Peserta P2K	Ya
3	1500004041	Indah Oktavia	P	PBI	3.40	53.1	Jln. Damai rt 08 kelurahan Simpang Perlang,	Petani / Nelayan	1	2000000	Peserta P2K	Tidak
4	1500004049	Etika Sari	P	PBI	3.70	40.1	Desa parangina	Petani / Nelayan	3	500000	Peserta P2K	Ya
5	1500029268	Ika Mita Yunita	P	IKM	3.43	54.0	DUSUN MADUSARI, rt/rw: 017/009,	Petani / Nelayan	2	1000000	Peserta P2K	Ya
...
555	1500019089	Wilda Febriani		TI	3.50	83.5	Tg. Sengkuang blok F No. 31 RT.03/RW. 12	Pegawai Swasta	4	8000000	Peserta P2K	Tidak

3.2. Cleaning

Setelah data beasiswa berhasil di *upload* maka tahap selanjutnya adalah tahapan *cleaning* yaitu tahap pembersihan data beasiswa yang dalam setiap atribut. Dalam penelitian ini setelah dilakukan proses *cleaning* tidak terdapat data yang salah sehingga jumlah data tetap 555 data

3.3. Selection

Selection adalah Proses untuk memutuskan atribut-atribut apa yang akan digunakan dalam data beasiswa. Atribut data pendaftar beasiswa antara lain: nim, nama, jenis kelamin, program studi, ipk semester I, rata-rata nilai ujian nasional, nomor hp, alamat, pekerjaan orang tua, tanggungan orangtua, penghasilan orangtua perbulan, keikutsertaan kegiatan mahasiswa baru. Diseleksi sehingga hanya atribut: nim, ipk semester I, rata-rata nilai ujian nasional, tanggungan, rata-rata penghasilan orangtua perbulan.

Tabel 3. Data mahasiswa hasil seleksi

No	Nim	Ipk Sem 1	Rata-rata UAN	Tanggungan	Penghasilan	Terima
1	1500004036	4.00	86.7	5	1300000	Tidak
2	1500004042	3.60	55.7	2	2000000	Ya
3	1500004041	3.40	53.1	1	2000000	Tidak
4	1500004049	3.70	40.1	3	500000	Ya
5	1500029268	3.43	54.0	2	1000000	Ya
6	1500019089	3.50	83.5	4	8000000	Tidak
7	1500011021	3.85	65.6	2	2000000	Ya
8	1500017125	3.65	88.3	3	1000000	Ya
9	1500009010	3.70	80.32	2	12000000	Ya
10	1500023172	4.00	75.67	1	7500000	Ya

3.4 Transformation

Transformation adalah proses merubah data atribut beasiswa yang telah di seleski dalam bentuk alphabet kedalam bentuk numerik (angka). Atribut-atrubut yang dilakukan tranformasi pada tabel data mining adalah IPK semester I, rata-rata nilai ujian nasional dan rata-rata penghasilan orangtua perbulan.

Penentuan variabel dilakukan dengan memilih atribut-atribut yang berpengaruh dalam proses penyeleksian beasiswa, pemberian nilai parameter/variabel disesuaikan dengan data mentah (data *training*). variabel dan hasil tranformasinya ditunjukkan dalam Tabel 4. Dan hasil tranformasi data mahasiswa ditunjukkan pada Tabel 5.

Tabel 4. Penentuan Variabel dan Nilai Variabel (Sumarlin, 2015) dan (Musthafa dkk, 2015)

No	Variabel	Nilai Variabel
1	IPK Semester I	1. 2,50-3,00
		2. 3,01-3,50
		3. 3,51-4,00
2	Rata-rata Nilai Ujian Nasional	1. <6
		2. 6-7
		3. 7,01-8
		4. >8
3	Penghasilan Orangtua per Bulan	1. 100000-500000
		2. 500001-1000000
		3. 1000001-2000000
		4. >2000000

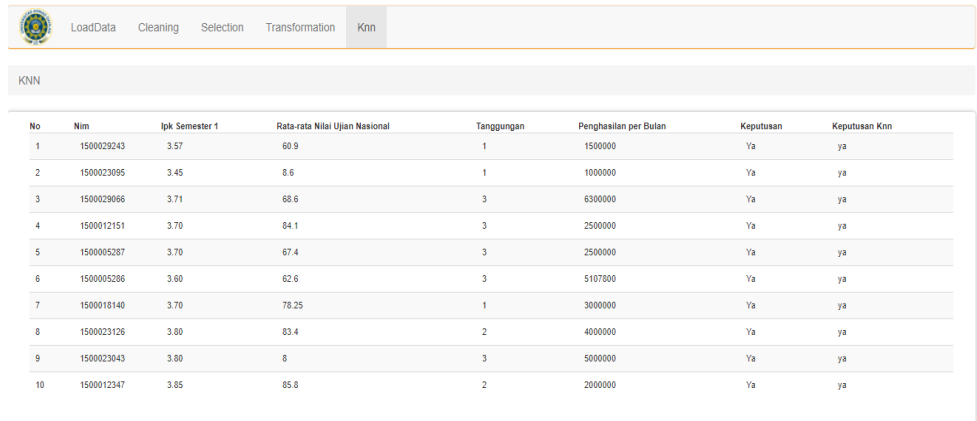
Tabel 5. Data mahasiswa hasil transformasi

No	NIM	IPK Semester I	Rata-rata Nilai Ujian Nasional	Tanggungan	Penghasilan Orangtua per Bulan	Keputusan di terima Beasiswa
1	1500004036	3	4	5	3	Tidak
2	1500004042	3	1	2	3	Ya
3	1500004041	2	1	1	3	Tidak
4	1500004049	3	1	3	1	Ya
5	1500029268	2	1	2	2	Ya

6	1500019089	2	4	4	4	Tidak
7	1500011021	3	2	2	3	Ya
8	1500017125	3	4	3	2	Ya
9	1500009010	3	4	2	4	Ya
10	1500023172	3	3	1	4	Ya

3.5 Implementasi K-NN

Hasil data pada Tabel 5 kemudian digunakan untuk proses training. Hasil training akan digunakan untuk melakukan testing data mahasiswa. Tampilan aplikasi testing menggunakan K-NN ditunjukkan pada Gambar 1.



No	Nim	Ipk Semester 1	Rata-rata Nilai Ujian Nasional	Tanggungan	Penghasilan per Bulan	Keputusan	Keputusan Knn
1	1500029243	3.57	60.9	1	1500000	Ya	ya
2	1500023095	3.45	8.6	1	1000000	Ya	ya
3	1500029066	3.71	68.6	3	6300000	Ya	ya
4	1500012151	3.70	84.1	3	2500000	Ya	ya
5	1500005287	3.70	67.4	3	2500000	Ya	ya
6	1500005286	3.60	62.6	3	5107000	Ya	ya
7	1500018140	3.70	78.25	1	3000000	Ya	ya
8	1500023126	3.80	83.4	2	4000000	Ya	ya
9	1500023043	3.80	8	3	5000000	Ya	ya
10	1500012347	3.85	85.8	2	2000000	Ya	ya

Gambar 1. Hasil implementasi K-NN

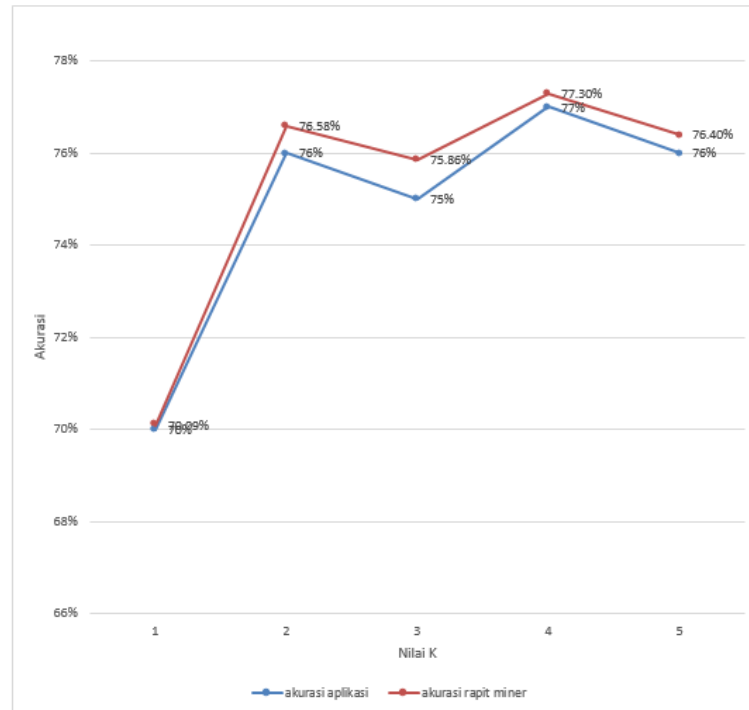
3.6 Pengujian

Hasil pengujian perhitungan dengan akurasi antara aplikasi yang dibuat dengan *tools RapidMiner* diperlihatkan oleh Tabel 4. Hasil akurasi dengan *k-nearest neighbor* pada aplikasi memperlihatkan bahwa dengan menggunakan data training 115, data testing 555 dengan mengkombinasikan nilai K dihasilkan nilai akurasi terbaik sebesar 77% dengan nilai K (jarak tetangga terdekat) adalah 4. Hal ini juga diperlihatkan pada *tools RapidMiner* dengan menggunakan data *training* 115, data testing 555 dengan mengkombinasikan nilai K dihasilkan nilai akurasi terbaik sebesar 77,30% dengan nilai K (jarak tetangga terdekat) adalah 4. Hasil akurasi dengan menggunakan aplikasi dengan menggunakan data training 115, data *testing* 555 dan nilai terbaik K=4 memiliki selisih sebesar 0,40% dengan hasil pada *tools RapidMiner*. Daftar akurasi ditunjukkan dalam Tabel 6.

Tabel 6. Daftar Akurasi

Input Ke	Training	Testing	K	Akurasi pada Aplikasi	Akurasi pada Rapid Miner
1	115	555	5	76%	76,40%
2	115	555	4	77%	77,30%
3	115	555	3	75%	75,86%
4	115	555	2	76%	76,58%
5	115	555	1	70%	70,09%

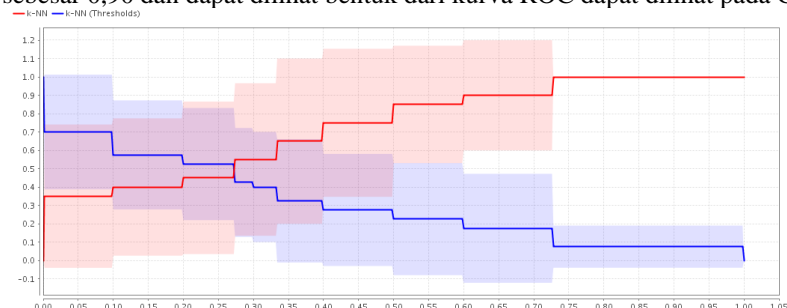
Berdasarkan pada Tabel 4. didapatkan perbandingan hasil akurasi berupa gambar grafik yang dapat dilihat pada Gambar 2



Gambar 2. Perbandingan Hasil Akurasi

Dari hasil pengujian diatas kemudian ditampilkan hasil dari kurva ROC. Kurva ROC menunjukkan akurasi dan membandingkan klasifikasi secara visual. ROC mengekspresikan *confusion matrix*. ROC adalah grafik dua dimensi dengan *false positives* sebagai garis horisontal dan *true positives* sebagai garis vertikal. Terdapat dua titik garis pada kurva ROC yaitu garis *false postifive rate* atau *specificity* dan *garis true positive rate* atau *sensitivity*. *Specificity* didefinisikan sebagai kemampuan alat tes untuk mendeteksi data negatif sedangkan *sensitivity* didefinisikan sebagai kemampuan alat tes untuk mendeteksi data positif.

Hasil perhitungan kurva ROC untuk algoritma k-nearest neighbor (knn) dengan menggunakan data *training* hasilnya sebesar 0,90 dan dapat dilihat bentuk dari kurva ROC dapat dilihat pada Gambar 3.



Gambar 3. Kurva ROC Algoritma K-Nearest Neighbor

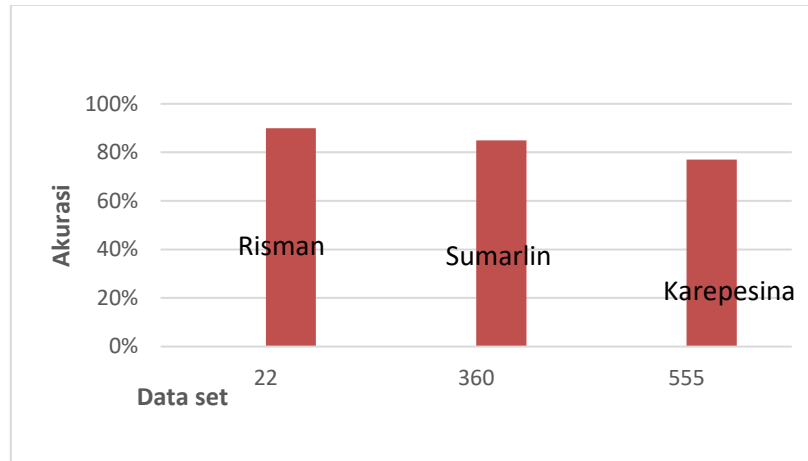
Berdasarkan pada Gambar 3 menunjukkan bahwa hasil algoritma k-nearest neighbor (knn) mendapatkan nilai 0,90 dimana nilai tersebut menunjukkan nilai AUC (*Area Under Curve*) *good classification* (sangat baik).

3.6 Evaluasi Pola dan Representasi Pengetahuan

Dalam penelitian ini dengan menggunakan metode k-nearest neighbor (k-nn) didapatkan hasil akurasi terbaik sebesar 77% dengan menggunakan data beasiswa sebanyak 770 kemudian dilakukan proses *cleaning* sehingga data menjadi 670 kemudian digunakan data *training* sebanyak 115, data *testing* sebanyak 555. Hal ini berbeda dengan dari penelitian sebelumnya yang diperoleh akurasi sebesar 90,90% karena data yang digunakan hanya 22 data sampel dimana data training sebanyak 21 dan 1 data testing [8]. Hasil dari penelitian ini juga berbeda dengan penelitian sebelumnya yang dihasilkan akurasi sebesar 85,96% dengan menggunakan 360 data [3]. Perbedaan data set dan akurasi dari metode *k-nearest neighbor* dapat dilihat pada Tabel 5

Peneliti	Risman dkk (2013)	Sumarlin (2015)	Karepesina (2016)
Data set	22 data	360 data	555 data
Akurasi	90,90%	85,96%	77%

Berdasarkan pada Tabel 5 terdapat perbedaan dari sisi jumlah data set yang digunakan serta akurasi yang diperoleh. Hasil dari tabel tersebut didapatkan perbandingan berupa gambar grafik perbandingan yang dapat dilihat pada Gambar 4



Gambar 4 Perbandingan Penelitian

4. KESIMPULAN

Berdasarkan hasil dan penelitian yang dilakukan pada program data *mining* untuk penentuan penerima beasiswa dengan menggunakan metode *k-nearest neighbor* (KNN) dapat ditarik kesimpulan sebagai berikut :

1. Telah diterapkan data mining dengan metode *k-nearest neighbor* (KNN) untuk penentuan penerima beasiswa.
2. Hasil dari penelitian ini adalah klasifikasi untuk menentukan penerima beasiswa menggunakan metode *k-nearest neighbor* dimana hasil pengujian dengan *confusion matrix* dan kurva ROC (*Receiver Operation Characteristic*) di peroleh hasil akurasi terbaik sebesar 77% dengan nilai AUC (*Area Under Curve*) 0,90 dari jumlah data training 115, data *testing* 555 dan nilai K adalah 4. Karena nilai AUC (*Area Under Curve*) berada diantar rentang 0.80–0.90, maka metode tersebut termasuk dalam kategori *good classification* (sangat baik).

DAFTAR PUSTAKA

- [1] Y. S. Nugroho, "Klasifikasi dan Klastering Mahasiswa Informatika Universitas Muhammadiyah Surakarta," pp. 89–98, 2015.
- [2] E. Turban, J. E. Aronson, T.-P. Lian, and D. Prabatini, *DECISION support systems and intelligent systems (sistem pendukung keputusan dan sistem cerdas)*. Andi Offset, 2005.
- [3] Sumarlin, "Implementasi Algoritma K-Nearest Neighbor Sebagai Pendukung Keputusan Klasifikasi Penerima Beasiswa PPA dan BBM," *J. Sist. Inf. Bisnis*, vol. 5, no. 1, pp. 52–62, 2015.
- [4] Merluarini, Bisri, Safitri, Diah, and H. Abdul, "Perbandingan Analisis Klasifikasi Menggunakan Metode K-Nearest Neighbour (K-NN) dan Multivariate Adaptive Regression Spline (MARS) pada Data Akreditasi Sekolah Dasar Negeri di Kota Semarang," *Gaussian*, vol. 3, no. 3, 2014.
- [5] D. N. Triwibowo, A. Kurniadi, and S. Hartinah, "Sistem Pendukung Keputusan Pemilihan Mahasiswa Penerima Beasiswa dengan K-NN dan ELECTRE," *Inspir. J. Teknol. Inf. dan Komun.*, vol. 9, no. 2, p. 89, 2019, doi: 10.35585/inspir.v9i2.2531.
- [6] S. Soegiarto and B. Bahar, "Model Aplikasi Penentuan Jenis Beasiswa Berbasis Algoritma K-NN Termodifikasi," *J. Ilm. Tek. Inform.*, vol. 5, no. 3, pp. 1181–1186, 2017, [Online]. Available: <http://ojs.stmik-banjarbaru.ac.id/index.php/jutisi/article/view/112>

-
- [7] R. Rosyidi, "Perbandingan Algoritma K-Nn Dan Cart Pada Data Mining Penerimaan Beasiswa," *CESS (Journal Comput. Eng. Syst. Sci.)*, vol. 4, no. 2, pp. 169–177, 2019.
- [8] H. Risman, D. Nugroho, and Y. R. WU, "Penerapan Metode K-Nearest Neighbor Pada Aplikasi," *Jural TIKomSiN*, vol. 3, no. 2, pp. 19–25, 2013.
- [9] O. B. Tamonob, K. Letelay, and S. Mola, "Penerapan Metode Smart (Simple Multi Attribute Rating Technique) Dan Algoritma K-NN (K-Nearest Neighbor) Dalam Penentuan Status Kesehatan Bayi Baru Lahir Di Rumah Sakit Bhayangkara Kupang," *Pros. Semmau*, no. September, pp. 505–511, 2017.
- [10] F. Gorunescu, *Data mining: concepts and techniques*. Berlin: Springer, 2011. doi: 10.1007/978-3-642-19721-5.
- [11] J. P. Jiawei Han, Micheline Kamber, *Data Mining: Concept and Techniques Thrid Edition*, Thrid Edit. USA: Elsevier, 2012.
- [12] E. Prasetyo, *Data Mining*. Yogyakarta: Andi Offset, 2012.
- [13] T. Mitsa, *Temporal Data Mining*. USA, 2010.

