

K-NEAREST NEIGHBOR BERBASIS SELEKSI ATRIBUT CHI SQUARE UNTUK KLASIFIKASI PENERIMA BEASISWA KURANG MAMPU

Hamsir Saleh

Fakultas Ilmu Komputer, Program Studi Teknik Informatika
Universitas Ichsan Gorontalo

Email: hamsir.saleh@gmail.com

Hamria

Fakultas Ilmu Komputer, Program Studi Teknik Informatika
Universitas Ichsan Gorontalo

Email: rhiyfatmawatihamka@gmail.com

ABSTRAK

Beasiswa merupakan bantuan dari pemerintah berupa sejumlah uang yang diberikan kepada siswa yang sedang atau yang akan menjalani pendidikan di sekolah. Pemberian beasiswa diharapkan dapat meningkatkan dan menumbuhkan semangat mahasiswa dalam menyelesaikan pendidikannya. Banyaknya peserta mahasiswa yang mengajukan beasiswa serta melebihi kuota yang diberikan mengakibatkan proses penyeleksian membutuhkan waktu yang lama karena proses penyeleksian harus sesuai dengan kriteria agar penerima beasiswa tepat sasaran. Metode Penelitian yang digunakan dalam penelitian ini yaitu menggunakan metode penelitian eksperimen yang berdasarkan kerangka kerja. Hasil penelitian ini yaitu berupa hasil pengujian yang menunjukkan nilai akurasi *K-Nearest Neighbor* dengan jumlah parameter *K* yang digunakan sebanyak 6 yaitu 1,3,5,7,9,10 dengan jumlah data sebanyak 254 data dengan 14 atribut didapat nilai *K* sebesar 85,78%, Metode Chi Square sebagai seleksi atribut membantu meningkatkan hasil akurasi klasifikasi *K-NN*. *K-Nearest Neighbor* berbasis seleksi atribut Chi Square lebih akurat dan efektif dalam klasifikasi penerima beasiswa dari data yang digunakan didapat hasil akurasi sebesar 88,53% dan nilai AUC 0,801 termasuk dalam kategori "ROC Curve Good".

Kata kunci: Beasiswa, Klasifikasi, Mahasiswa, *K-Nearest Neighbor*, *Chi-Square*

ABSTRACT

*Scholarships are assistance from the government in the form of an amount of money given to students who are currently or who will be studying at school. Scholarships are expected to increase and foster student enthusiasm in completing their education. The large number of student participants applying for scholarships and exceeding the quota given resulted in the selection process taking a long time because the selection process must comply with the criteria so that scholarship recipients are right on target. The research method used in this research is using an experimental research method based on a framework. The results of this study are in the form of test results which show the accuracy value of *K-Nearest Neighbor* with the number of *K* parameters used as many as 6, namely 1,3,5,7,9,10 with a total of 254 data with 14 attributes, a *K* value of 85 is obtained. 78%, the Chi Square method as attribute selection helps improve the results of the *K-NN* classification accuracy. *K-Nearest Neighbor* based on Chi Square attribute selection is more accurate and effective in classifying scholarship recipients. From the data used, an accuracy of 88.53% is obtained and an AUC value of 0.801 is included in the "ROC Curve Good" category.*

Keywords: *Scholarship, Classification, Student, K-Nearest Neighbor, Chi-Square*

1. PENDAHULUAN

Beasiswa adalah pemberian berupa bantuan keuangan yang diberikan kepada individu, pelajar atau mahasiswa yang dipergunakan untuk keberlangsungan pendidikan yang ditempuh. Sesuai dengan ketentuan pasal 4 ayat (1) UU PPh/2000 yang menyebutkan bahwa penghasilan adalah tambahan kemampuan ekonomis dengan nama dan dalam bentuk apapun yang diterima atau diperoleh dari sumber Indonesia atau luar Indonesia yang dapat digunakan untuk konsumsi atau menambah kekayaan wajib pajak, karena beasiswa diartikan menambah kemampuan ekonomis bagi penerimanya, berarti beasiswa merupakan penghasilan. Beasiswa dapat diberikan oleh lembaga pemerintah, yayasan ataupun perusahaan yang dapat dikategorikan pada pemberian cuma-cuma atau pemberian dengan ikatan kerja setelah selesai pendidikannya [1].

Pendidikan formal merupakan salah satu faktor pendukung sumber daya manusia yang berkualitas. Dengan melalui dunia pendidikan diharapkan kecerdasan seseorang dan kemampuan berpikir dapat dikembangkan agar mampu mengolah dan memanfaatkan sumber daya yang tersedia. Dalam rangka mendorong peningkatan kualitas sumber daya manusia yang cerdas, terampil, sehat dan berdaya saing serta memperluas kesempatan bagi warga masyarakat yang tidak mampu dan berprestasi akademik dalam bidang pendidikan maka perlu untuk difasilitasi dengan memberikan bantuan berupa program bantuan studi atau beasiswa. Pemerintah daerah khususnya di kabupaten Boalemo melalui Dinas Pendidikan Kabupaten Boalemo memberikan bantuan beasiswa kepada masyarakat Boalemo yang sedang menempuh pendidikan pada jenjang pendidikan tinggi baik didalam dan Luar Negeri yang berprestasi dan tidak mampu. Beasiswa merupakan pemberian bantuan berupa uang yang diberikan kepada perorangan dengan tujuan untuk keberlangsungan pendidikan yang ditempuh. Dalam penelitian ini beasiswa yang digunakan yaitu beasiswa bantuan pendidikan program S1 bagi mahasiswa kurang mampu di Kabupaten Boalemo.

Dalam penelitian ini, proses seleksi penentuan penerima beasiswa menggunakan enam faktor penentu yaitu semester, IPK, penghasilan orang tua, tanggungan orang tua, penerima beasiswa, kelengkapan berkas, penerimaan. Data pemohon beasiswa dibuat dalam format Ms-Excel yang kemudian dibahas pada rapat seleksi untuk penentuan pemohon layak atau tidak layak menerima beasiswa. Pengambilan keputusan ini masih membutuhkan waktu yang cukup lama disebabkan banyaknya peserta mahasiswa yang mengajukan beasiswa serta melebihi kuota yang diberikan, dan bisa beresiko tidak konsisten akibat adanya unsur subjektifitas. Algoritma klasifikasi data mining dapat membantu mengetahui pola pemetaan karakteristik pemohon ke keputusan yang diambil sehingga proses seleksi menjadi lebih mudah, tidak memerlukan waktu yang lama dan unsur subjektifitas dari pengambilan keputusan dapat dihindari.

Klasifikasi merupakan teknik data mining yang menggunakan proses menemukan model yang menggambarkan dan membedakan kelas suatu data untuk memprediksi kelas dari suatu obyek yang belum diketahui label kelasnya[2]. Klasifikasi adalah salah satu metode data mining yang digunakan dalam menganalisis sekumpulan data penting. Banyak metode klasifikasi yang dapat digunakan secara otomatis memprediksi kelas dari data lain yang belum diklasifikasikan. Klasifikasi dan prediksi merupakan dua masalah utama penelitian yang berkaitan dengan hasil klasifikasi. Kedua hal ini merupakan bentuk analisis data yang dapat digunakan untuk menggambarkan ekstrak model yang penting pada kelas data atau untuk memprediksi tren data masa depan[3][4].

Algoritma K-NN merupakan metode klasifikasi klasik yang paling sederhana. Algoritma K-NN melakukan klasifikasi berdasarkan data pembelajaran yang jaraknya paling dekat (*nearest*) dan menggunakan prinsip ketetanggaan (*neighbor*) untuk memprediksi *class* yang baru. Jumlah tetangga yang dipakai adalah sebanyak K tetangga [5]. Algoritma *K-Nearest Neighbor* dengan seleksi atribut *Chi-Square* adalah salah satu teknik klasifikasi yang dapat digunakan untuk menghasilkan keputusan yang akurat. Metode klasifikasi algoritma *K-Nearest Neighbor* adalah salah satu metode pengklasifikasin data yang memiliki konsistensi yang efektif dan kuat dalam melakukan *training* terhadap data yang besar. Seleksi fitur *Chi-Square* merupakan salah satu metode *feature selection* yang dapat mereduksi fitur-fitur yang tidak relevan dan dapat meningkatkan akurasi[6].

Chi-square merupakan *feature selection* yang digunakan untuk mereduksi fitur-fitur yang tidak relevan dalam proses klasifikasi oleh *K-Nearest Neighbor*. Seleksi fitur *chi-square* digunakan untuk

uji independensi dan estimasi terhadap sebuah *term* dengan kategorinya untuk mengetahui ketergantungan suatu *class* pada suatu fitur. *Chi-square* di sebut juga dengan Kai Kuadrat. *Chi-square* adalah salah satu jenis uji komparatif non parametris yang dilakukan pada dua variabel dimana skala data kedua variabel adalah nominal. Tujuan dari penggunaan seleksi fitur yaitu untuk menghilangkan fitur pengganggu dalam klasifikasi. Berdasarkan teori statistika, seleksi fitur *chi-square* memiliki dua peristiwa diantaranya kemunculan dari fitur dan kemunculan dari kategori yang selanjutnya setiap nilai *term* diurutkan dari yang tertinggi[2][7][8].

Tujuan penelitian ini yaitu untuk mengklasifikasikan peserta mahasiswa yang mendaftarkan beasiswa dengan mengelompokkan menjadi diterima atau tidak diterima menggunakan algoritma *K-Nearest Neighbor* berbasis seleksi atribut *Chi-Square*. Penelitian ini menggunakan metode penelitian eksperimen yang berdasarkan kerangka kerja yang terdiri atas lima tahapan yaitu, pengumpulan data set, *pre-processing*, metode yang digunakan, pengujian model, evaluasi dan validasi.

Berdasarkan permasalahan yang ada bahwa banyaknya waktu yang dibutuhkan untuk penentuan hasil seleksi yang disebabkan banyaknya peserta pemohon beasiswa serta resiko tidak konsisten karena adanya unsur subyektifitas. Sehingga dalam penelitian ini dibutuhkan sistem yang dapat membantu dalam menganalisa pengambilan keputusan hasil seleksi penentuan penerima beasiswa dengan menggunakan algoritma *K-Nearest Neighbor* dengan seleksi fitur *Chi Square*.

2. METODOLOGI PENELITIAN

2.1 *Beasiswa*

Beasiswa adalah pemberian berupa bantuan keuangan yang diberikan kepada individu, pelajar atau mahasiswa yang dipergunakan untuk keberlangsungan pendidikan yang ditempuh. Sesuai dengan ketentuan pasal 4 ayat (1) UU PPh/2000 yang menyebutkan bahwa penghasilan adalah tambahan kemampuan ekonomis dengan nama dan dalam bentuk apapun yang diterima atau diperoleh dari sumber Indonesia atau luar Indonesia yang dapat digunakan untuk konsumsi atau menambah kekayaan wajib pajak, karena beasiswa diartikan menambah kemampuan ekonomis bagi penerimanya, berarti beasiswa merupakan penghasilan. Beasiswa dapat diberikan oleh lembaga pemerintah, yayasan ataupun perusahaan yang dapat dikategorikan pada pemberian cuma-cuma atau pemberian dengan ikatan kerja setelah selesai pendidikannya [1]. Beasiswa yang digunakan dalam penelitian ini yaitu beasiswa kurang mampu.

Beasiswa berfungsi sebagai bantuan dana bagi mahasiswa berprestasi maupun yang kurang mampu untuk memperoleh pendidikan yang layak yang diberikan oleh suatu lembaga swasta maupun pemerintah. Ada beberapa tujuan dari pemberian beasiswa kepada mahasiswa diantaranya [9]:

1. Untuk menghasilkan sumber daya manusia yang berpotensi untuk berperan dalam mempercepat pembangunan bangsa menuju pada kemandirian ditengah-tengah percaturan global.
2. Menwujudkan keadilan dan deokratisasi dalam bidang pendidikan dengan memberikan beasiswa kepada mahasiswa yang berprestasi.
3. Memberikan bantuan dana kepada mahasiswa yang mengalami kendala secara ekonomis ataupun geografis.

2.2 *Klasifikasi*

Klasifikasi dalam penelitian ini merupakan salah satu metode data mining yang digunakan dalam menganalisis sekumpulan data penting. Banyak metode klasifikasi yang dapat digunakan secara otomatis memprediksi kelas dari data lain yang belum diklasifikasikan. Klasifikasi dan prediksi merupakan dua masalah utama penelitian yang berkaitan dengan hasil klasifikasi. Kedua hal ini merupakan bentuk analisis data yang dapat digunakan untuk menggambarkan ekstrak model

yang penting pada kelas data atau untuk memprediksi tren data masa depan [3][4]. Proses klasifikasi digambarkan pada empat komponen, yaitu [2];

1. Kelas, yaitu variabel dependen yang berupa kategorikal dengan merepresentasikan label pada objek
2. Prediktor, yaitu variabel dependen yang direpresentasikan oleh atribut atau karakteristik data.
3. Dataset training, yaitu data yang berisi nilai dari variable dependen dan independen yang digunakan untuk menentukan kelas yang sesuai berdasarkan predictor.
4. Dataset testing, yaitu data yang diklasifikasikan oleh model yang telah dibuat dan akurasi klasifikasi yang telah dievaluasi.

2.3 Metode Penelitian

Penelitian ini menggunakan metode penelitian eksperimen yang berdasarkan kerangka kerja yang terdiri atas lima tahapan sebagai berikut:

2.3.1 Pengumpulan Data

Data yang digunakan dalam penelitian ini adalah data mahasiswa pengusul beasiswa pada Dinas Pendidikan Kabupaten Boalemo dengan jumlah 406 data set.

2.3.2 Pre-processing

Tahap *pre-processing* dilakukan untuk penyeleksian data, yang mana data dibersihkan dan ditransformasikan ke bentuk yang diinginkan sehingga dapat dilakukan persiapan dalam pembuatan model. Tahap pengolahan awal data dilakukan untuk mempersiapkan data yang benar-benar valid sebelum diproses pada tahap berikutnya. Pada tahap ini dilakukan *cleansing*, transformasi, reduksi dan seleksi fitur. Data yang dapat diolah untuk mendapatkan atribut yang relevan dan sesuai.

2.3.3 Metode yang digunakan

Pada tahap ini data dianalisis, dikelompokkan variabel mana yang berhubungan dengan satu sama lainnya. Setelah data dianalisis lalu diterapkan model-model yang sesuai dengan jenis data. Pembagian data ke dalam data latihan (*data training*) dan data uji (*data testing*) yang diperlukan untuk pembuatan model. Model algoritma yang digunakan dalam penelitian ini yaitu algoritman *k-nearest neighbor* dengan seleksi fitur *chi-square*.

2.3.4 Pengujian model

Tahap pengujian model dilakukan untuk klasifikasi data penerima beasiswa. Data set ini akan dilakukan pengujian dengan dua model. Yang pertama adalah penerapan algoritma *k-nearest neighbor* tanpa *chi-square*. Yang kedua adalah penerapan algoritma *k-nearest neighbor* dengan *chi-square*. Eksperimen yang dilakukan dengan menggunakan metode *K-Nearest Neighbor* berbasis seleksi fitur *chi-square* untuk mendapatkan akurasi pada klasifikasi penerima beasiswa. Diperlukan pemilihan variabel terbaik dan pengaturan parameter-parameter nilai *k* untuk mendapatkan nilai akurasi yang tepat. Berikut adalah langkah dalam pemilihan arsitektur yang diusulkan:

1. Data yang sudah diolah divalidasi menggunakan *k-fold validation*
2. Mengubah parameter *K* yaitu jumlah tetangga terdekat dalam melakukan proses prediksi menggunakan *k-nearest neighbor*. Nilai parameter *K* yang digunakan adalah 1-9

2.3.5 Evaluasi dan Validasi hasil

Tahap ini dilakukan untuk mengetahui tingkat keberhasilan dari metode yang diusulkan pada data set penerima beasiswa menggunakan metode *k-nearest neighbor* berbasis seleksi fitur *chi-square*. Maka pada penelitian ini akan digunakan *confusion matrix* untuk memudahkan peneliti mendapatkan tingkat akurasi dari klasifikasi data penerima beasiswa. Setelah data diuji dimasukkan kedalam *confusion matrix* yang selanjutnya dihitung nilai *accuracy*, *precision* dan *recall*.

Setelah dilakukan evaluasi, selanjutnya *Receiver Operating Characteristics* (ROC) digunakan untuk menggambarkan performa dari setiap klasifikasi yang diujikan secara dua dimensi, yang mana nilai *false positif* merupakan garis *horizontal* sedangkan nilai *true positif* merupakan garis *vertical*. Area dibawah grafik ROC merupakan nilai dari *Area Under Curve* (AUC). Nilai AUC diukur dengan cara memperkirakan probabilitas nilai output dari sampel yang dipilih acak dari populasi positif ataupun negative, semakin besar nilai dari AUC maka semakin kuat klasifikasinya. Kategori hasil AUC dapat dilihat berdasarkan nilai kualitas suatu klasifikasi dapat dilihat pada tabel berikut ini [10][11].

Tabel 1. Kriteria AUC

<i>Nilai AUC</i>	<i>Penjelasan</i>
90 % - 100 %	<i>Baik Sekali</i>
80% - 90%	<i>Baik</i>
70% - 80%	<i>Cukup</i>
60% - 70%	<i>Kurang</i>
<60%	<i>Kurang Sekali</i>

3. HASIL DAN PEMBAHASAN

3.1 *Pengumpulan Data*

Data yang digunakan dalam penelitian ini adalah data yang diperoleh dari studi lapangan. Data ini berupa data calon penerima beasiswa yang digunakan sebagai data latih untuk mengklasifikasi penerima beasiswa. Data yang dikumpulkan berupa data calon penerima beasiswa yang telah diketahui kelasnya yaitu “Ya” menerima beasiswa, “No” tidak menerima beasiswa. Jumlah data yang digunakan sebanyak 254 record dengan 14 atribut.

3.2 *Preprocessing*

Tahap preprocessing dilakukan untuk penyeleksian data, data dibersihkan dan ditransformasikan kebentuk yang diinginkan sehingga dapat dilakukan persiapan dalam pembuatan model. Tahap awal pengolah data dilakukan untuk mempersiapkan agar data-data ini benar-benar valid sebelum diproses pada tahap berikutnya. Karena data yang digunakan merupakan data lengkap maka tidak dilakukan proses penanganan *missing value*. Dalam data ini juga tidak terdapat duplikasi data dan jumlah data yang akan diolah sudah sesuai dengan kebutuhan saat proses klasifikasi.

3.3 *Hasil Eksperimen*

3.3.1 *Evaluasi Menggunakan Algoritma K-NN*

Penerapan algoritma dalam penelitian ini menggunakan Tools Rapidminer. Hasil eksperimen yang didapatkan dengan menggunakan nilai K yaitu 1, 3, 5, 7, 9 dengan 10 *K-fold cross validation* dengan hasil pengujian model pada *confusion matrix* untuk mendapatkan hasil akurasi terbaik. Berikut hasil eksperimen yang didapatkan:

Tabel 2. Pengujian K-NN Dengan 10 K-Fold Cross Validation

Nilai K	K-Fold Cross Validation	Metode K-NN	
		Accuracy	AUC
1	10	81.39%	0.500
3	10	83.38%	0,753
5	10	85.78%	0.794
7	10	85.35%	0.742
9	10	85.35%	0.776

Dari hasil eksperimen yang telah dilakukan, didapat nilai akurasi terbaik pada K = 5 dengan nilai akurasi 85.78% dan nilai AUC 0.794. Untuk pengujian dengan *confusion matrix* menggunakan K-NN dengan nilai K = 5 dapat dilihat pada tabel berikut ini.

Tabel 3. Hasil Confusion Matrix Menggunakan K-NN

	<i>True Ya</i>	<i>True No</i>	<i>Class Precision</i>
<i>Pred. Ya</i>	194	21	90.23%
<i>Pred. No</i>	15	23	60.53%
<i>Class Recall</i>	92.82%	52.27%	

Eksperimen ini menggunakan data sebanyak 253 record. Berdasarkan *confusion matrix* terlihat bahwa 194 record diprediksi “Ya” pada kelompok data “Ya” dan sebanyak 21 record diprediksi “No” pada kelompok data “Ya”. Selanjutnya sebanyak 15 record diprediksi “Ya” pada kelompok data “No” dan 23 record diprediksi “No” pada kelompok data “No”.

Precision merupakan proporsi data yang benar-benar kelas No diantara data yang diklasifikasikan sebagai kelas No. Dalam matrix, elemen diagonal dibagi dengan jumlah kolom yang relevan, yaitu:

$$Precision = \frac{TP}{TP+FP} = \frac{194}{194+21} \times 100\% = 90.23\% \quad (1)$$

Recall (r) merupakan jumlah sampel yang diklasifikasikan positif sebagai total sampel dalam dataset berkategori positif, dalam matrix elemen menurun dibagi dengan jumlah kolom yang relevan yaitu,

$$Recall = \frac{TP}{TP+FN} = \frac{194}{194+15} \times 100\% = 92.82\% \quad (2)$$

Dari hasil ini dapat dihitung nilai akurasi, yaitu:

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} = \frac{194+23}{194+21+15+23} \times 100\% = 85.78\% \quad (3)$$

Dari hasil perhitungan yang dilakukan dapat disimpulkan bahwa persentase tingkat akurasi pada *confusion matrix* dengan menggunakan algoritma K-NN mencapai nilai persentase sebesar 85.78%.

3.3.2 Evaluasi Menggunakan Algoritma K-NN dengan Seleksi Atribut Chi Square

Dalam penelitian ini dilakukan juga eksperimen dengan menggunakan seleksi atribut chi square. Seleksi atribut ini dilakukan untuk menentukan atribut mana yang berpengaruh terhadap data. Namun sebelum pengujian dilakukan terlebih dahulu mencari parameter yang sesuai dengan atribut masukan untuk mencari hasil akurasi terbaik. adapun parameter yang digunakan dalam pengujian ini yaitu *number of bins* dengan *k-fold cross validation*.

Tabel 4. Pengujian K-NN Chi Square Dengan 10 K-Fold Cross Validation

Nilai K	K-Fold Cross Validation	Number Of Bins	Metode K-NN	
			Accuracy	AUC
1	10	10	87.73%	0.500
3	10	10	88.18%	0.809
5	10	10	88.53%	0.801
7	10	10	87.43%	0.746
9	10	10	88.24%	0.843

Hasil eksperimen algoritma K-NN dengan seleksi fitur Chi square didapatkan nilai akurasi sebesar 88.49% dan nilai AUC = 0.801 pada nilai K = 5 dengan *k-fold cross validation* = 10 dan *number of bins* = 10.

Pengujian dengan *confusion matrix* menggunakan K-NN dengan seleksi fitur Chi Square pada nilai K = 5 dapat dilihat pada tabel berikut ini.

Tabel 5. Hasil Confusion Matrix Menggunakan K-NN Dengan Seleksi Atribut Chi Square

	True Ya	True No	Class Precision
Pred. Ya	199	19	91.28%
Pred. No	10	25	71.43%
Class Recall	95.22%	56.82%	

Eksperimen ini menggunakan data sebanyak 254 record. Berdasarkan *confusion matrix* terlihat bahwa 199 record diprediksi “Ya” pada kelompok data “Ya” dan sebanyak 19 record diprediksi “No” pada kelompok data “Ya”. Selanjutnya sebanyak 10 record diprediksi “Ya” pada kelompok data “No” dan 25 record diprediksi “No” pada kelompok data “No”.

Precision merupakan proporsi data yang benar-benar kelas No diantara data yang diklasifikasikan sebagai kelas No. Dalam matrix, elemen diagonal dibagi dengan jumlah kolom yang relevan, yaitu:

$$Precision = \frac{TP}{TP+FP} = \frac{199}{199+19} \times 100\% = 91.28\% \quad (4)$$

Recall (r) merupakan jumlah sampel yang diklasifikasikan positif sebagai total sampel dalam dataset berkategori positif, dalam matrix elemen menurun dibagi dengan jumlah kolom yang relevan yaitu,

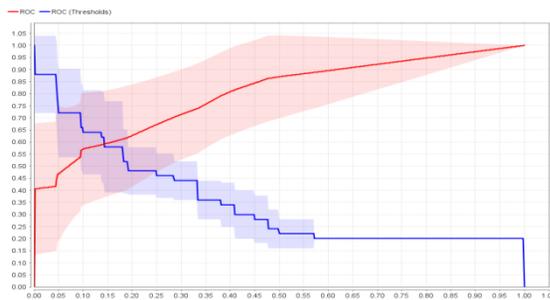
$$Recall = \frac{TP}{TP+FN} = \frac{199}{199+10} \times 100\% = 95.22\% \quad (5)$$

Dari hasil ini dapat dihitung nilai akurasi, yaitu:

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} = \frac{199+25}{199+19+10+25} \times 100\% = 88.53\% \quad (6)$$

Dari hasil perhitungan yang dilakukan dapat disimpulkan bahwa persentase tingkat akurasi pada *confusion matrix* dengan menggunakan algoritma K-NN mencapai nilai persentase sebesar 88.53%.

Dari hasil pengujian yang dilakukan menghasilkan nilai *accuracy*, *precision* dan *recall* serta *ROC (Receiver Operating Characteristic) Curve*. *AUC (Area Under the ROC Curve)* merupakan grafik yang dapat digunakan untuk menilai model. Karena AUC merupakan bagian dari daerah unit persegi nilainya akan selalu antara 0,0 dan 1,0. Pada gambar 1 menunjukkan bahwa garis berwarna merah merupakan kurva *ROC* dengan nilai sebesar 0.801 termasuk klasifikasi baik sedangkan garis berwarna biru merupakan kurva ambang (thresholds).



Gambar 1. Kurva ROC K-NN Berbasis Seleksi Fitur Chi Square

Berdasarkan hasil dengan *confusion matrix* dan kurva ROC menunjukkan rule hasil klasifikasi untuk mengklasifikasi penerima beasiswa termasuk *good clasification* sehingga dapat digunakan untuk mengklasifikasi penerima beasiswa.

3.3.3 Perbandingan Hasil Metode K-NN dan K-NN Chi Square

Dari hasil analisis yang telah dilakukan dengan menggunakan algoritma K-NN, K-NN Chi Square. Hasil akurasi yang didapatkan dapat pada gambar 2 bahwa dimana terjadi kenaikan akurasi pada setiap metode algoritma yang digunakan. Hal ini membuktikan bahwa metode algoritma K-NN berbasis seleksi atribut Chi Square dapat digunakan dalam penelitian ini.



Gambar 2. Perbandingan Metode Algoritma K-NN dan K-NN Chi Square

4. KESIMPULAN

Metode algoritma K-NN berbasis seleksi atribut Chi Square yang digunakan dalam penelitian ini terbukti mampu meningkatkan akurasi dalam klasifikasi penerima beasiswa. Dari akurasi dengan K-NN didapat nilai akurasi sebesar 85,78%. Metode Chi Square sebagai seleksi atribut dapat membantu meningkatkan hasil akurasi klasifikasi K-NN. *K-Nearest Neighbor* berbasis seleksi atribut Chi Square lebih akurat dan efektif dalam klasifikasi penerima beasiswa dari data yang digunakan didapat hasil akurasi sebesar 88,53% dan nilai AUC 0,801 termasuk dalam kategori “*ROC Curve Good*”.

Walaupun model algoritma K-NN yang diterapkan dalam penelitian ini telah berjalan dengan baik, namun terdapat beberapa saran dalam pengembangannya diantaranya yaitu bahwa prosedur ini tidak selalu mengarahkan ke model pemilihan atribut yang terbaik. Untuk lebih meningkatkan akurasi klasifikasi penerima beasiswa dapat dilakukan penggabungan metode klasifikasi lainnya, penggunaan metode fitur seleksi atau menambahkan metode optimasi lainnya.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada DRPM Kemenristek Dikti yang telah memberi dukungan financial pada Penelitian Dosen Pemula pada penelitian ini.

DAFTAR PUSTAKA

- [1] Astuti, Sariyah. 2015. "Sistem Pendukung Keputusan Penentuan Beasiswa Menggunakan Metode Simple Additive Weighting (SAW) Studi Kasus Pada SMP Dharma Bhakti Pubian." *J. TAM (Technol. Accept. Model)* vol. 4. no. 1, pp. 13–18.
- [2] Noviana, Dita. 2019. "Analisis Rekomendasi Penerima Beasiswa Menggunakan Algoritma K-Nearest Neighbor (K-NN) dan Algoritma C4.5," Semin. Nas. Penelit. Pendidik. Mat. 2019 UMT, pp. 79–87.
- [3] Sulistiani, Heni. 2018. "Penerapan Algoritma Klasifikasi Sebagai Pendukung Keputusan Pemberian Beasiswa Mahasiswa," SNTI.
- [4] Pratama, Aria., et al. 2021. "Klasifikasi Penerima Beasiswa Dengan Menggunakan Algoritma K Nearest Neighbor," *J. Data Sci. Inform.*, vol. 1, no. 1, pp. 11–15.
- [5] Siringoringo, Rimbun. "Integrasi Metode Resampling dan K-Nearest Neighbor pada Prediksi Cacat Softwar Aplikasi Android," *Jurnal ISD*, vol.2, no.1, pp. 47-58
- [6] Listiowarni, Indah. 2018, "Feature Selection Chi-Square dan K-NN pada Pengkategorian Soal Ujian Berdasarkan Cognitive Domain Taksonomi Bloom," *J. Komput. Terap.*, vol. 4, no. 1, pp. 21–30.
- [7] Anisah, Siti. 2016. "Klasifikasi Teks Menggunakan Chi Square Feature Selection Untuk Menentukan Komik Berdasarkan Periode, Materi Dan Fisikdengan Algoritma Naivebayes," *Compiler*, vol. 5, no. 2, pp. 59–66.
- [8] Irvantoro, Deni. 2019. "Feature Selection Menggunakan Chi-Square Dan N-Gram Dengan Algoritma Naive Bayes Classifier Untuk Analisis Sentimen Review Produk Elektronik," *Repos. Unmuhjember*.
- [9] Ramadhani, Februariyanti, 2019 "Sistem Pendukung Keputusan Penerimaan Beasiswa Mahasiswa Menggunakan Metode SAW (Simple Additive Weighting)," *Proceeding SINTAK*, pp. 1–8.
- [10] Chamidah, Nurul. 2021. "Perbandingan Metode Klasifikasi Naïve Bayes, Decision Tree dan K-Nearest Neighbor pada data Log Firewall," *SENAMIKA*.
- [11] Widartha, Vandha. 2021. "Perbandingan Akurasi Algoritma K-Nearest Neighbor dan Logistic Regression untuk Klasifikasi Penyakit Diabetes," *e-Proceeding of Engineering*, vol.8, no.5, pp. 9757-9764.