



Klasifikasi Penerima Bantuan Beasiswa Menggunakan Algoritma *K-Nearest Neighbour* Dengan Seleksi Fitur *Backward Elimination*

Ikrimatul Wilda Lorenza^{1*}, Reni Umilasari², Moh Dasuki³

^{1,2,3}Program Studi Teknik Informatika, Universitas Muhammadiyah Jember^{1,2,3}

Email: krimatul124@gmail.com¹, reni.umilasari@unmuhjember.ac.id², moh.dasuki22@unmuhjember.ac.id³

ABSTRAK

Penelitian ini menjelaskan penggunaan algoritma *K-Nearest Neighbor* terhadap klasifikasi data bantuan beasiswa. Dataset yang digunakan dalam penelitian ini berasal dari SMK Darus Sholihin Puger dengan jumlah variabel 11, dengan output iya dan tidak yang diperoleh dari data bantuan beasiswa sebanyak 293 siswa, metode yang digunakan adalah *K-Nearest Neighbor*. Pada metode *K-Nearest Neighbor* dengan nilai $K=3$ mendapatkan akurasi sebesar 82,91%, presisi 84,52%, dan recall 75,45%. Sedang *K-Nearest Neighbor* menggunakan *Backward Elimination* dengan nilai $K=3$ mendapatkan akurasi sebesar 83,76%, presisi 85,23%, recall 76,78%. Penentuan kriteria terhadap output data hasil menggunakan algoritma *K-Nearest Neighbor* terhadap data SMK Darus Sholihin Puger diperoleh hasil peningkatan terbaik dengan nilai akurasi terbesar pada $K=7$ yaitu 2,56%, presisi terbesar didapatkan pada $K=7$ yaitu 11,16%, sedangkan recall terbesar didapatkan pada $K=9$ yaitu 5,09%.

Kata Kunci: Klasifikasi, *K-Nearest Neighbor*, *Backward Elimination*, Bantuan Beasiswa

ABSTRACT

This study describes the use of the K-Nearest Neighbor algorithm for the classification of scholarship aid data. The dataset used in this study comes from Darus Sholihin Puger Vocational School with 11 variables with yes and no outputs obtained from scholarship aid data of 293 students. In the K-Nearest Neighbor method with a value of $K=3$, the accuracy is 82.91%, 84.52% precision, and 75.45% recall. Meanwhile, K-Nearest Neighbor using Backward Elimination with a value of $K=3$ gets 83.76% accuracy, 85.23% precision, 76.78% recall. Determination of the criteria for the output data results using the K-Nearest Neighbor algorithm on the data of SMK Darus Sholihin Puger, the best improvement results are obtained with the largest accuracy value at $K=7$ which is 2.56%, the largest precision is obtained at $K=7$ which is 11.16%, while the largest recall was obtained at $K=9$, namely 5.09%.

Keywords: Classification, *K-Nearest Neighbor*, backward elimination, scholarship assistance

1. PENDAHULUAN

Beasiswa adalah bantuan berupa keuangan yang diberikan kepada siswa-siswi kurang mampu yang bertujuan untuk berlangsungnya pendidikan yang ditempuh selama masa studi. Di setiap lembaga pendidikan khususnya sekolah banyak sekali beasiswa yang ditujukan kepada siswa-siswa, baik yang kurang mampu maupun yang berprestasi, bantuan beasiswa bertujuan untuk membantu meringankan beban biaya pendidikan sekolah. Saat ini bantuan berupa beasiswa sangatlah diperlukan banyak siswa untuk berlangsungnya pendidikan yang akan ditempuh. Banyaknya jumlah siswa yang mengajukan permohonan untuk menerima beasiswa dan kriteria penilaian yang banyak pula maka tidak semua siswa yang mengajukan permohonan beasiswa dapat diterima.

Salah satu cara untuk mengetahui seleksi bantuan beasiswa ini yaitu menggunakan klasifikasi. Data bantuan beasiswa ini di perlu klasifika agar dapat mengetahui hasil dengan akurat. Hasil klasifikasi yang akurat membantun para guru dalam mengambil keputusan dengan tepat. *K-Nearest Neighbor* merupakan teknik yang sederhana, efisien dan efektif dalam bidang pengenalan pola, kategori teks, pengolahan objek dan lain-lain, karena sederhana dalam pengolahannya dan dapat melakukan training data dalam jumlah yang besar. Penerapan algoritma *K-Nearest Neighbor* sebagai proses klasifikasi yang dirancang untuk mengetahui hasil data bantuan beasiswa, dengan mengklasifikasi data bantuan beasiswa dapat memperoleh nilai akurasi tertinggi. Maka dari itu penulis memutuskan menggunakan *K-Nearest Neighbor* pada proses klasifikasi dalam menentukan nilai akurasi.

2. KAJIAN PUSTAKA

A. Bantuan Beasiswa

Beasiswa yayasan adalah bantuan yang di berikan oleh pihak perguruan tinggi untuk diperuntukan kepada siswa yang berasal dari kalangan keluarga yang kurang mampu, berprestasi maupun aktif agar siswa tersebut tetap bisa melanjutkan ke perguruan tinggi. Beasiswa ini di salurkan melalui perguruan tinggi dan berada di bawah donator pihak yayasan maupun perguruan tinggi itu sendiri. Untuk penyaluran beasiswa yayasan ini kepada siswa baru, harus memperhatikan syarat-syarat tertentu sebelum beasiswa itu di berikan kepada siswa yang bersangkutan (Assrani et al., 2018). Untuk syarat-syarat yang harus dipenuhi seperti Layak PIP, Jenis Tinggal, Alat Transportasi, Penerima KPS, Pendidikan Ayah, Pendidikan Ibu, Penghasilan Ayah, Penghasilan Ibu, Pekerjaan Ayah, Pekerjaan Ibu, Nerima KIP.

B. Data Mining

Data Mining merupakan analisa terhadap informasi(umumnya informasi yang berdimensi besar) untuk menciptakan ikatan yang jelas dan merumuskan yang belum diketahui tadinya dengan metode terbaru dimengerti dan bermanfaat untuk *owner* informasi tersebut (Larose, 2006). Data mining mempunyai sebagian metode yaitu klasifikasi, clustering, regresi, seleksi variabel, dan *market basket analyst* (Santosa, 2007).

C. K-Nearest Neighbor

K-Nearest Neighbor merupakan salah satu model yang dapat diterapkan dalam melakukan klasifikasi terhadap sejumlah data, dengan mencari data yang mempunyai jarak terdekat dengan suatu objek penelitian, sesuai dengan jumlah tetangga terdekatnya yang diinisialisasikan dengan K. Pencarian jarak terdekat biasanya dihitung menggunakan jarak *Euclidean*. Jarak *Euclidean* memiliki persamaan sebagai berikut (Amin, 2016):

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

Keterangan:

$d(x, y)$ = Jarak *Euclidean*

x_i = Data *Training* ke- i

y_i = Data *Testing* ke- i

D. Backward Elimination

Backward Elimination merupakan salah satu tahap yang dapat dilakukan dalam melakukan seleksi atribut pada tahap *preprocessing*. Algoritma *Backward Elimination* dipilih karena kemampuan algoritma ini yang memungkinkan untuk mendapatkan beberapa atribut yang awalnya memiliki kemampuan klasifikasi rendah secara individu namun jika digabungkan dengan atribut lainnya akan memiliki akurasi yang tinggi (Dallal, 2005).

Untuk mencari nilai prediksi dari masing-masing atribut (Suyono, 2015).

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + \varepsilon_i$$

Keterangan:

Y : Atribut terkait

X : Atribut bebas

β : Parameter regresi

ε : Nilai kesalahan/standar error estimasi

Mencari masing-masing nilai MSE (*Mean Square Error*) tiap atribut dengan rumus (Suyono, 2015):

$$MSE = \frac{\sum (Y_t - \hat{Y}_t)^2}{n - p}$$

Keterangan:

\hat{y}_t : Nilai prediksi

\bar{y} : Nilai rata-rata prediksi

p : Jumlah atribut

\hat{y}_t : Nilai aktual
 n : Jumlah *record* data
Untuk mencari nilai t hitung.

$$t \text{ hitung} = \frac{(bi)}{Se_{(bi)}}$$

Keterangan:

bi : Koefisien Regresi
 Se_{bi} : Standar Error
Untuk mencari nilai SSE

$$SSE = \sum(Y_t - \hat{Y}_t)^2$$

Keterangan:

SSE : *Sum Square Error*
 Y_t : Nilai Aktual
 \hat{Y}_t : Nilai Prediksi
Untuk mencari *Standar Error*

$$Se_{(bi)} = \sqrt{MSE \times C_{ii}}$$

Keterangan:

$Se_{(bi)}$: *Standar Error*
 MSE : *Mean Square Error*
 C_{ii} : Nilai Invers

E. *Confusion Matrix*

Confusion Matrix merupakan metode yang dapat digunakan untuk pengevaluasian suatu model klasifikasi. *Confusion matrix* berisi nilai dari kelas prediksi yang didapatkan dari model yang dibandingkan dengan kelas yang asli dari dataset dengan kata lain berisi informasi jumlah dari kelas aktual dan kelas prediksi pada klasifikasi (Abdillah, 2015). Untuk menghitung nilai akurasi, presisi, recall menggunakan persamaan sebagai berikut:

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN} \times 100\%$$

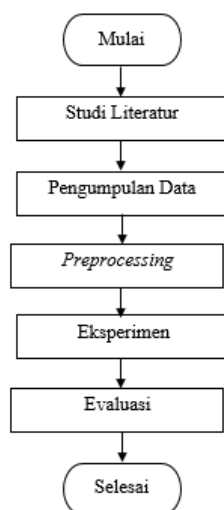
$$Presisi = \frac{TP}{TP + FP} \times 100\%$$

$$Recall = \frac{TP}{TP + FN} \times 100\%$$

F. *RapidMiner*

RapidMiner (Yale) adalah perangkat lunak *open source* untuk *knowledge discovery* dan data mining. *Rapid Miner* memiliki kurang lebih 400 prosedur (operator) data mining, termasuk operator untuk masukan, *output*, data *preprocessing* dan visualisasi. Aplikasi data mining yang telah dikembangkan menggunakan *Rapid Miner* banyak digunakan di dunia bisnis maupun penelitian. Beberapa fitur dari *Rapid Miner*, antara lain: Berlisensi gratis (*open source*), multi-platform karena diprogram dalam bahasa Java dan internal data berbasis XML sehingga memudahkan pertukaran data eksperimen.

3. METODE PENELITIAN



Gambar 1. Alur Penelitian

A. Studi Literatur

Studi literatur yaitu untuk memperoleh landasan-landasan yang digunakan untuk penelitian ini, bisa didapatkan melalui jurnal, buku dan internet yang berhubungan dengan penelitian.

B. Pengumpulan Data

Data yang digunakan diambil dari SMK Darus Sholihin Puger. Data yang digunakan yaitu 293 data, melakukan pengujian hasil prediksi meliputi 2 kelas klasifikasi yaitu iya dan tidak. Selanjutnya dilakukan transformasi data dengan menyamakan format pada setiap atribut dari dataset dengan mengubah kategori menjadi inisial angka.

C. Preprocessing Data

Setelah melakukan pengumpulan data, tahap selanjutnya yaitu preprocessing data sebelum data tersebut pada proses pengujian. Dalam proses ini yaitu dilakukan pembersihan nilai kosong. Tidak semua atribut digunakan, hanya menggunakan atribut yang dianggap relevan atau atribut yang berpengaruh saja yang akan digunakan pada penelitian ini.

4. HASIL DAN PEMBAHASAN

Pengujian dilakukan menggunakan algoritma *K-Nearest Neighbor* tanpa *Backward Elimination* dan menggunakan *Backward Elimination* untuk mengetahui nilai akurasi, presisi dan *recall*, dengan mencari nilai $K = 3, 5, 7, 9$.

Tabel 1. Hasil Keseluruhan Pada KNN Tanpa *Backward Elimination*.

Algoritma	Nilai K KNN	Akurasi	Presisi	Recall
KNN	3	82,91%	84,52%	75,45%
	5	75,21%	71,83%	68,02%
	7	72,65%	63,67%	68,55%
	9	72,65%	62,97%	68,83%

Hasil akurasi, presisi dan recall secara keseluruhan pada algoritma *K-Nearest Neighbor* (KNN) dengan seleksi fitur *backward elimination* ditampilkan pada Tabel 2.

Tabel 2. Hasil Keseluruhan Pada KNN Menggunakan *Backward Elimination*

Algoritma	Nilai K KNN	Akurasi	Presisi	Recall
KNN	3	83,76%	85,23%	76,78%
	5	77,78%	76,54%	70,06%
	7	75,21%	74,83%	64,50%
	9	74,79%	63,40%	73,92%

Tabel 3. Nilai Peningkatan Akurasi, Presisi dan Recall

Algoritma	Nilai K KNN	Akurasi	Presisi	Recall
KNN	3	0,85%	0,29%	1,33%
	5	2,52%	4,71%	2,04%
	7	2,56%	11,16%	4,05%
	9	2,14%	0,43%	5,09%

Peningkatan pada K = 3 yaitu akurasinya sebesar 0,85%, presisi sebesar 0,29%, recall sebesar 1,33%. Peningkatan pada K = 5 yaitu akurasinya sebesar 2,52%, presisi sebesar 4,71%, recall sebesar 2,04%, untuk peningkatan pada K = 7 yaitu akurasinya sebesar 2,56%, presisi sebesar 11,16%, recall sebesar 4,05%. Dan peningkatan pada K = 9 yaitu akurasinya sebesar 2,14%, presisi sebesar 0,43%, recall sebesar 5,09%.

5. KESIMPULAN

A. Kesimpulan

Berdasarkan penelitian ini yang menggunakan klasifikasi algoritma *K-Nearest Neighbour* tanpa *Backward Elimination* dan menggunakan *Backward Elimination* maka dapat disimpulkan:

1. Hasil dari akurasi tertinggi dari klasifikasi penerima bantuan beasiswa dengan menggunakan algoritma *K-Nearest Neighbour* tanpa menggunakan backward elimination yaitu sebesar 82,91% pada K = 3, sedangkan nilai akurasi menggunakan backward elimination yaitu sebesar 83,76% pada K = 3.
2. Hasil dari presisi pada klasifikasi penerima bantuan beasiswa dengan menggunakan algoritma *K-Nearest Neighbour* tanpa menggunakan backward elimination yaitu sebesar 84,52% pada K = 3, sedangkan nilai presisi menggunakan backward elimination yaitu sebesar 85,23% pada K = 3.
3. Hasil dari recall pada klasifikasi penerima bantuan beasiswa dengan menggunakan algoritma *K-Nearest Neighbour* tanpa menggunakan backward elimination yaitu sebesar 75,45% pada K = 3, sedangkan nilai recall menggunakan backward elimination yaitu sebesar 76,78% pada K = 3.
4. Didapatkan hasil peningkatan nilai akurasi, presisi dan recall tertinggi yaitu dengan menggunakan seleksi fitur backward elimination dengan akurasi sebesar 2,56% presisi sebesar 11,16%, dan recall sebesar 5,09%.

B. Saran

Berdasarkan penelitian yang dilakukan, adapun saran yang diberikan untuk penelitian selanjutnya adalah sebagai berikut:

1. Dapat menggunakan metode yang berbeda dengan konsep yang sama, sehingga dapat membandingkan nilai akurasi dengan lebih banyak algoritma.
2. Menambahkan seleksi fitur untuk mengetahui atribut yang berpengaruh untuk penerima bantuan beasiswa.
3. Dari penelitian yang penulis lakukan maka diharapkan untuk penelitian selanjutnya dilakukan perbandingan nilai k pada algoritma KNN
4. Mengetahui perbedaan nilai akurasi, serta melakukan pembobotan dari setiap atribut dengan seleksi fitur yang lain.

6. DAFTAR PUSTAKA

- Abdillah, S. (2015). Penerapan Algoritma Decision Tree C4.5 Untuk Diagnosa Penyakit Stroke Dengan Klasifikasi Data Mining Pada Rumah Sakit Santa Maria Pemasang. *Jurnal Teknik Informatika*, 3(4), 1–12.

- Amin, M. (2016). Optimasi Klasifikasi Penilaian Akreditasi Lembaga Kursus Menggunakan Metode K-NN dan Naive Bayes. *Technologia*, 7(3), 148–154.
- Assrani, D., Huda, N., Sidabutar, R., Saputra, I., & Sulaiman, O. K. (2018). Penentuan Penerima Bantuan Siswa Miskin Menerapkan Metode Multi Objective Optimization on The Basis of Ratio Analysis (MOORA). In *JURIKOM* (Vol. 5, Issue 1). <http://ejurnal.stmik-budidarma.ac.id/index.php/jurikom>.
- Dallal, G. E. (2005). *Simplifying a Multiple Regression Equation. The Little Handbook of Statistical Practice*. Tufts University.
- Larose, D. T. (2006). *Data Mining Methods and Models*. John Wiley & Sons.
- Santosa, B. (2007). *Teknik Pemanfaatan Data untuk Keperluan Bisnis*. Graha Ilmu.
- Suyono. (2015). *Analisis Regresi untuk Penelitian*. Deepublish.