

Penerapan Algoritma *K-Nearest Neighbour* dengan *Euclidean Distance* untuk Menentukan Kelompok Uang Kuliah Tunggal Mahasiswa

Fenny Purwani^{1,*}, Ragil Tri Wahyudi¹, Irfan Dwi Jaya¹

¹ Program Studi Sistem Informasi, UIN Raden Fatah Palembang, Indonesia

* Correspondence: fennypurwani_uin@radenfatah.ac.id

Copyright: ©2022 by the authors

Received: 19 September 2022 | Revised: 9 Oktober 2022 | Accepted: 17 November 2022 | Published: 20 Desember 2022

Abstrak

Uang kuliah tunggal (UKT) merupakan besaran biaya pendidikan yang ditentukan berdasarkan kemampuan ekonomi mahasiswa. Pada penerapannya masih banyak mahasiswa yang keberatan dengan kelompok UKT yang didapatkan. Oleh karena itu pihak universitas harus menerapkan metode yang tepat dan akurat dalam menentukan kelompok UKT. Penelitian ini bertujuan untuk mendapatkan hasil klasifikasi kelompok UKT mahasiswa menggunakan algoritma K-Nearest Neighbor (KNN) dengan perhitungan jarak Euclidean Distance dan mengetahui akurasi algoritma tersebut dengan nilai k optimal. Penelitian ini menggunakan metode kuantitatif dengan pendekatan deskriptif. Adapun teknik pengumpulan data yang digunakan adalah wawancara, studi pustaka, dan dokumentasi. Data yang telah dikumpulkan yaitu sebanyak 1.650 data verifikasi UKT mahasiswa tahun 2019-2021 diolah dengan data mining menggunakan bantuan *software* Rstudio. Hasil penelitian menunjukkan bahwa klasifikasi dengan KNN dapat diterapkan dalam penentuan UKT mahasiswa. Dengan data testing sebanyak 320 mahasiswa terdapat 23 mahasiswa ditentukan mendapat UKT I, 149 UKT II, 129 UKT III, 32 UKT IV, dan 2 mahasiswa mendapat UKT V. Adapun akurasi dari algoritma tersebut yaitu sebesar 87,58% dalam kategori Good Classification dengan nilai k optimal pada KNN yang didapatkan dengan K-Fold Cross Validation adalah k=1.

Kata kunci: data mining; klasifikasi; k-nearest neighbor; euclidean distance

Abstract

Single tuition fee or called UKT is the amount of tuition fee determined based on the student's economic ability. In its application, there are still many students who object to the UKT group that is obtained. Therefore, the university must apply the right and accurate method in determined the UKT group. This study aims to obtain the result of student's UKT group classification using the K-Nearest Neighbor (KNN) algorithm with Euclidean Distance calculation and determine the accuracy of the algorithm with the optimal k value. This study used a quantitative method with a descriptive approach. The data collection techniques used are interviews, literature study, and documentation. The data that has been collected is 1,650 student's UKT verification data for 2019-2021 which be processed with data mining using the RStudio software. The results showed that the classification with KNN can be applied in determined student's UKT. With data testing many as 320 students, 23 students were determined to get UKT I, 149 UKT II, 129 UKT III, 32 UKT IV, and 2 students got UKT V. The accuracy of the algorithm is 87.58% in the Good Classification category. The optimal k for KNN obtained with K-Fold Cross Validation is k=1.

Keywords: data mining; classification; k-nearest neighbor; euclidean distance

PENDAHULUAN

Uang kuliah tunggal merupakan besaran biaya pendidikan yang dibebankan kepada mahasiswa dan ditentukan berdasarkan kemampuan ekonomi keluarganya (Effendi & Latifah,



2021; Purnomo & Saifullah, 2022). Penerapan kebijakan uang kuliah bertujuan untuk memberikan efisiensi, keadilan, dan kepastian besaran biaya pendidikan, sehingga mahasiswa dapat membayar biaya pendidikan sesuai dengan kemampuan ekonominya. Penerapan kebijakan uang kuliah sudah banyak diterapkan oleh berbagai universitas di Indonesia. Misalnya UIN Maulana Malik Ibrahim, UIN Suska Riau, dan UIN Sumatera Utara yang menerapkan kebijakan ini sejak 2014.

Pada implementasi kebijakan uang kuliah ini, permasalahan utama yang sering terjadi adalah adanya ketidaktepatan dalam penentuan kelompok uang kuliah yang kurang sesuai dengan kondisi ekonomi mahasiswa. Banyak mahasiswa yang memiliki taraf ekonomi kurang mampu mendapatkan kelompok uang kuliah yang kurang sesuai hingga ada yang mendapat tarif tertinggi (Retnoningsih & Marom, 2017). Ketidaktepatan penentuan kelompok uang kuliah tersebut menyebabkan banyak mahasiswa keberatan dengan kelompok uang kuliah yang didapatkan. Penentuan kelompok uang kuliah yang tidak sesuai dengan kemampuan mahasiswa mengindikasikan adanya ketidakberadilan dalam penerapan kebijakan tersebut.

UIN Raden Fatah Palembang merupakan perguruan tinggi islam negeri di bawah naungan Kementerian Agama Republik Indonesia yang telah menerapkan kebijakan uang kuliah sejak tahun 2014. Pada awal kebijakan ini diterapkan, penentuan kelompok uang kuliah dilakukan dengan wawancara terhadap masing-masing wali mahasiswa. Pada tahun 2016 UIN Raden Fatah Palembang mulai membangun sistem penentuan uang kuliah yang dapat mendukung proses penentuan kelompok uang kuliah tunggal mahasiswa. Namun dalam penerapannya masih terdapat permasalahan yang terjadi.

Banyaknya mahasiswa yang keberatan dengan kelompok uang kuliah yang didapatkan menjadi indikasi permasalahan dalam penerapan kebijakan uang kuliah di UIN Raden Fatah Palembang. Kondisi ideal yang diharapkan adalah seluruh mahasiswa baru puas atau tidak keberatan dengan kelompok UKT yang didapatkan. Selain itu terdapat 12 atribut yang digunakan dalam penentuan kelompok uang kuliah di universitas ini. Pihak universitas harus menerapkan metode yang tepat dan akurat untuk mempertimbangkan nilai setiap atribut sehingga dapat memenuhi rasa keadilan pada kelompok uang kuliah yang didapatkan setiap mahasiswa. Di sisi lain, terdapat data verifikasi uang kuliah mahasiswa di tahun-tahun sebelumnya yang belum dimanfaatkan secara optimal. Belum pernah dilakukan penggalan pengetahuan pada kumpulan data verifikasi tersebut.

Alternatif solusi yang dilakukan dalam penelitian ini adalah dengan memanfaatkan data verifikasi uang kuliah mahasiswa di tahun sebelumnya yang tersimpan dalam gudang data, untuk digali informasi atau pengetahuan baru di dalamnya. Sehingga dapat dimanfaatkan untuk membangun model baru dalam penentuan kelompok uang kuliah mahasiswa. Proses penggalan informasi baru dari sekumpulan data dikenal dengan data mining (Witten et al., 2011). Salah satu teknik dalam data mining adalah klasifikasi yaitu teknik dalam data mining dengan mengklasifikasikan kelas target secara akurat (Muktamar et al., 2015). Terdapat berbagai algoritma yang dapat digunakan pada klasifikasi seperti Decision Tree, Naïve Bayes, K-Nearest Neighbor, dan lainnya. Penelitian ini menggunakan K-Nearest Neighbor (KNN) untuk mengklasifikasikan kelompok uang kuliah mahasiswa. Penelitian menunjukkan bahwa KNN memiliki performa lebih baik dibandingkan algoritma lain yaitu Naive Bayes (Nurmalasari et al., 2021). KNN merupakan algoritma yang bekerja dengan mengklasifikasi data uji berdasarkan kedekatan jarak terhadap k data latih (Setianto et al., 2019). Penentuan jarak terdekat dalam algoritma KNN dapat menggunakan beragam rumus seperti Euclidean, Manhattan, dan lainnya. Penelitian ini menggunakan penghitungan jarak Euclidean karena Euclidean memberikan performa akurasi paling baik dibandingkan Minkowski dan Manhattan yaitu sebesar 84,47% (Nishom, 2019).

Terdapat penelitian terdahulu yang berkaitan dengan penerapan algoritma KNN dan penentuan UKT menggunakan data mining. Penelitian untuk mengklasifikasikan kelompok

uang kuliah tunggal mahasiswa menggunakan KNN, dengan perbandingan data testing dan data training sebesar 80%:20% menunjukkan hasil performa akurasi algoritma tersebut adalah sebesar 84,21%. Adapun nilai k yang digunakan dalam penelitian tersebut adalah $k=13$ (Sukamto et al., 2020). Pada penelitian tersebut penentuan nilai k pada KNN dilakukan hanya dengan menguji coba satu nilai k saja yaitu 13, tidak dengan nilai k yang lain. Hal tersebut belum dapat membuktikan apakah nilai k tersebut merupakan nilai k optimal pada algoritma KNN untuk menentukan kelompok UKT mahasiswa. Untuk menentukan nilai k optimal pada algoritma K-Nearest Neighbor dapat dilakukan dengan optimasi parameter menggunakan K-Fold Cross Validation (Banjarsari et al., 2015). Oleh karena itu pada penelitian ini mencoba melakukan pencarian nilai k optimal dengan skenario nilai k pada rentang 1-13 dengan menggunakan K-Fold Cross Validation. Sehingga nilai k yang digunakan nantinya merupakan nilai yang benar-benar memberikan performa akurasi terbaik.

Penelitian ini bertujuan untuk mendapatkan hasil klasifikasi kelompok UKT mahasiswa UIN Raden Fatah Palembang menggunakan algoritma K-Nearest Neighbor (KNN) dengan perhitungan jarak Euclidean Distance. Hasil klasifikasi yang dimaksud adalah berapa mahasiswa yang ditentukan mendapat UKT I, II, III, IV, dan V. Tujuan selanjutnya dari penelitian ini adalah untuk mengetahui dari akurasi algoritma KNN dengan nilai k optimal.

METODE

Penelitian ini merupakan penelitian deskriptif kuantitatif dengan sampel penelitian sebesar 1.650 data mahasiswa. Pengumpulan data dilakukan dengan teknik wawancara, studi pustaka, dan dokumentasi. Data yang telah dikumpulkan akan diolah dengan data mining menggunakan algoritma K-Nearest Neighbor dengan perhitungan jarak Euclidean Distance. K-Nearest Neighbor (KNN) merupakan algoritma yang bekerja dengan mengklasifikasikan suatu objek berdasarkan data latih dengan melihat jarak terdekat dari objek yang diuji berdasarkan nilai k (Setianto et al., 2019). Adapun *Euclidean Distance* merupakan metode untuk mencari kedekatan dan kemiripan antar dua titik variabel, dimana semakin dekat dan mirip kedua titik tersebut maka semakin kecil jaraknya (Hidayat et al., 2021; Nuranti et al., 2021; Wahyudi et al., 2022; Yudhana et al., 2020).

Pada penelitian ini tahapan penelitian dibagi menjadi empat tahap agar dapat merincikan setiap langkah dari penelitian. Keempat tahap tersebut adalah analisis pendahuluan, pengumpulan data, pengolahan data, hingga pembahasan dan kesimpulan. Analisis pendahuluan dilakukan untuk mendapatkan informasi terkait indikasi permasalahan pada objek yang diteliti sehingga dapat diketahui permasalahan dan tujuan dari penelitian yang akan dilakukan. Selanjutnya adalah dilakukan pengumpulan data yang dilakukan dengan teknik wawancara, studi pustaka dan dokumentasi. Data yang telah dikumpulkan selanjutnya diolah dengan data mining berdasarkan tahapan CRISP-DM atau *Cross Industry Standard Process for Data Mining*. Dimulai dengan tahapan *research understanding*, *data understanding*, *data preparation*, *modeling*, *evaluation*, dan hingga tahap *deployment*. Setelah tahap pengolahan data selanjutnya masuk pada tahap pembahasan hasil dari serangkaian proses data mining yang telah dilakukan. Dari nilai k optimal yang didapatkan, hasil penentuan uang kuliah menggunakan KNN, hingga performa algoritma KNN. Tahap terakhir dalam penelitian ini penarikan kesimpulan dari serangkaian proses penelitian dan hasil yang didapatkan.

HASIL DAN PEMBAHASAN

Hasil

Pengolahan data mining merupakan hasil utama dalam penelitian ini yang temuan-temuannya akan dibahas dibagian pembahasan. Proses pengolahan data mining dilakukan dengan tahapan sesuai CRISP-DM (*Cross Industry Standard Porcess for Data Mining*). Berikut merupakan uraian setiap tahapan pengolahan data mining yang dilakukan. Proses data

mining diawali dengan melakukan pemahaman penelitian (*research understanding*) yang bertujuan untuk memahami permasalahan dan menetapkan tujuan dilakukannya data mining (Larose, 2014). Sebagaimana permasalahan yang telah diuraikan dalam latar belakang, maka penerapan data mining pada penelitian ini bertujuan untuk mendapatkan hasil klasifikasi kelompok UKT mahasiswa menggunakan algoritma KNN menggunakan perhitungan Euclidean dan untuk mengetahui akurasi dari algoritma tersebut.

Tahap setelah pemahaman penelitian adalah pemahaman data (*data understanding*). Tahap ini bertujuan untuk memahami karakteristik data yang akan diolah (Zong, Xia & Zhang, 2021). Adapun data yang akan digunakan dalam proses data mining adalah sampel data verifikasi UKT mahasiswa baru UIN Raden Fatah Palembang tahun 2019-2021 sebanyak 1.650 data.

Tabel 1. Dataset awal

No	Jurusan	Pekerjaan Ayah	...	KIP	UKT
1	Pendidikan Matematika	Buruh harian/petani yang menggarap lahan orang lain	...	Ada	2
2	Pendidikan Matematika	Buruh harian/petani yang menggarap lahan orang lain	...	Ada	2
...
1.650	Pendidikan Biologi	Buruh harian/petani yang menggarap lahan orang lain	...	Tidak Ada	1

Tabel 1 menunjukkan dataset awal yang didapatkan. Dimana terdapat 15 atribut yaitu nomor, jurusan, pekerjaan ayah, pekerjaan ibu, tanggungan, total pendapatan, status ayah, status ibu, tempat tinggal, luas tempat tinggal, kondisi tempat tinggal, sumber air, penerangan, KIP, dan UKT. Dataset awal tersebut akan diolah ditahap selanjutnya yaitu preparation data untuk menyiapkan data sehingga dapat diolah dengan data mining.

Tabel 2. Transformasi atribut pekerjaan ayah dan ibu

Atribut	Sub Atribut	Konversi
Pekerjaan Ayah	Buruh harian/petani yang menggarap lahan orang lain	1
	Buruh kantor/perusahaan/petani yang mengelola lahan sendiri	2
	Pensiun (Purna Bakti)	3
	PNS/BUMN/TNI/POLRI	4
Pekerjaan Ibu	Ibu Rumah Tangga (IRT)	1
	Buruh harian/petani yang menggarap lahan orang lain	2
	Buruh kantor/perusahaan/petani yang mengelola lahan sendiri	3
	Pensiun (Purna Bakti)	4
	PNS/BUMN/TNI/POLRI	5

Pada tahap *data preparation* dilakukan proses meliputi seleksi dan transformasi pada dataset awal. Seleksi dilakukan untuk menghapus dan jurusan karena tidak mempengaruhi hasil

kelompok UKT. Selanjutnya transformasi dilakukan untuk mengkonversi nilai atribut dari tipe nominal menjadi numerik. Tabel 2 merupakan ketentuan pada proses transformasi atribut pekerjaan ayah dan ibu. Setiap nilai atribut awal yang bertipe nominal akan dikonversi menjadi tipe numerik.

Tabel 3. Transformasi atribut

Atribut	Sub Atribut	Konversi
Tanggung	Di atas 6 orang	1
	4 Sampai 6 Orang	2
	Kurang dari 4	3
Total Pendapatan	Dibawah 1 Juta (< 1 Jt)	1
	1 S/d 3 Juta	2
	3 S/d 5 jt	3
	Di atas 5 Juta	4
Status Ayah	Meninggal	1
	Hidup	2
Status Ibu	Meninggal	1
	Hidup	2
Tempat Tinggal	Menumpang	1
	Sewa/Kontrak	2
	Rumah Sendiri	3
Luas Tempat Tinggal	Lebih Kecil Sama dengan 36 M ²	1
	45 M ²	2
	90 M ²	3
	120 M ²	4
	Lebih Besar dari 120M ²	5
Kondisi Tempat Tinggal	Lantai Tanah/Semen tanpa Kamar mandi dalam rumah	1
	Lantai tanah/semen dengan kamar mandi dalam rumah	2
	Dua Lantai dengan tiang	3
	Lantai Permanen /Keramik	4
Sumber Air	Air Sungai / Tadah Hujan	1
	Sumur / Sumur Bor	2
	PDAM	3
	PDAM Swasta	4
Penerangan	Tidak Ada Listrik PLN	1
	PLN Tanpa Meteran (menumpang)	2
	PLN dengan meteran daya 450/900 V	3
	PLN dengan meteran daya 1300 V	4
	PLN dengan daya diatas 1300 V	5
KIP (Kartu Indonesia Pintar)	Ada	1
	Tidak Ada	2

Tabel 3 menunjukkan proses transformasi pada atribut tanggungan, total pendapatan, status ayah dan status ibu, tempat tinggal, luas tempat tinggal, kondisi tempat tinggal, sumber air, penerangan, dan kepemilikan KIP. Semua atribut tersebut akan ditransformasi dari tipe nominal menjadi numerik. Pada tabel 4 akan ditampilkan dataset yang telah melalui transformasi dan siap diolah ditahap pemodelan.

Tabel 4. Dataset hasil preparation data

No	Pekerjaan Ayah	Pekerjaan Ibu	Tanggungan	Total Pendapatan	Status Ayah	Status Ibu	Tempat Tinggal	Luas Tempat Tinggal	Kondisi Tempat Tinggal	Sumber Air	Penerangan	Kepemilikan KIP	Kelompok UKT
1	1	1	2	2	2	2	3	2	2	2	3	1	II
2	1	1	3	2	2	2	1	1	4	3	4	1	II
...
1.650	1	2	2	1	2	2	1	1	1	1	1	2	I

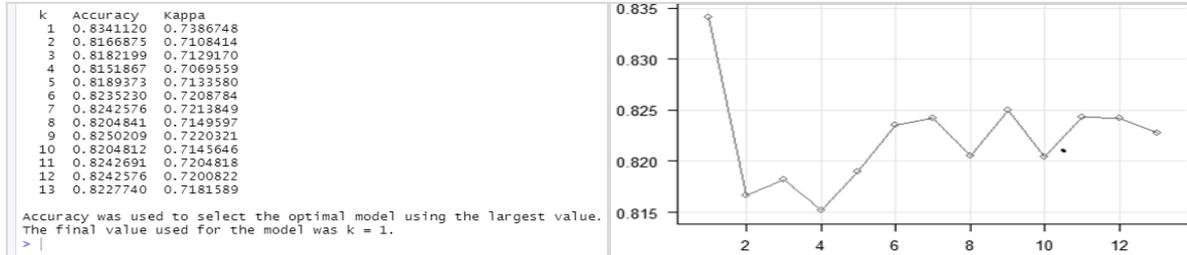
Tabel 4 menunjukkan dataset hasil dari proses transformasi. Setelah proses transformasi dilakukan, dataset akan dibagi menjadi data latih (*training*) dan data uji (*testing*). Proses pembagian data dilakukan dengan Split Train Test yang membagi data menjadi data *training* dan data *testing*. Adapun persentase pembagian data yaitu sebesar 80% untuk data *training* dan 20% untuk data *testing*. Pembagian dataset dengan perbandingan 80:20 dapat memberikan performa model yang terbaik (Primajaya et al., 2020). Sehingga pada penelitian ini dengan dataset sejumlah 1.650 data, terdapat 1.320 data *training* dan 330 data *testing*.

Tahap pemodelan merupakan tahapan implementasi dari teknik dan algoritma yang dipilih (Schuh et al., 2019). Pada penelitian ini akan diterapkan klasifikasi dengan algoritma yang telah dipilih yaitu K-Nearest Neighbor (KNN). Dalam pemodelan ini, akan dicari terlebih dahulu berapakah nilai k yang memberikan performa paling optimal dalam algoritma KNN. Adapun pemodelan akan dilakukan dengan menggunakan software Rstudio sebagaimana tersaji dalam gambar 2.



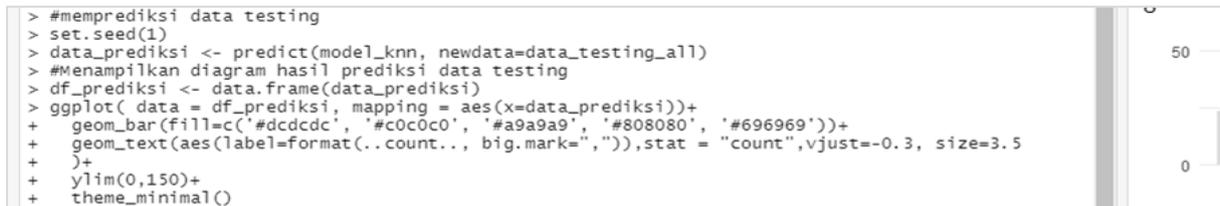
Gambar 1. Pemodelan di RStudio

Nilai k yang diuji coba adalah pada rentang k=1-13. Dengan menjalankan perintah pada gambar 2, yang dimulai dengan memanggil library yang dibutuhkan, mengimport dataset, hingga membuat model dengan algoritma KNN akan didapatkan hasil nilai k yang paling optimal. Gambar 3 sebelah kiri merupakan hasil akurasi setiap nilai k dan sebelah kanan merupakan grafik akurasi dari setiap nilai k.

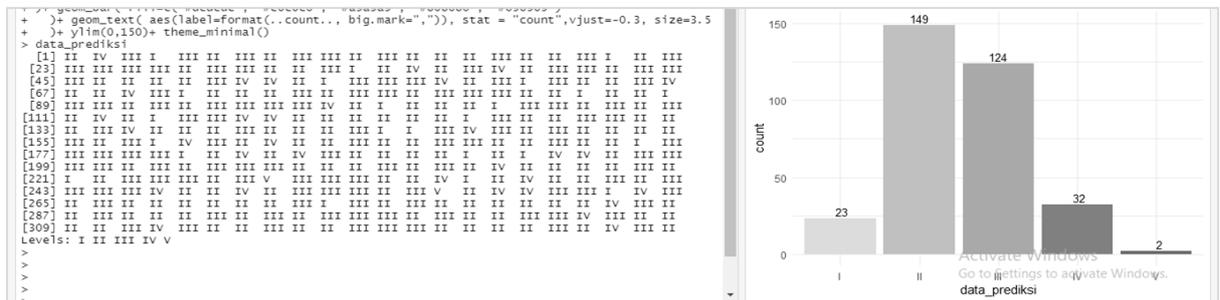


Gambar 2. Hasil akurasi setiap nilai k

Gambar 3 menunjukkan akurasi pada setiap nilai k 1-13. Untuk lebih memudahkan melihat nilai akurasi setiap nilai k maka dibuatkan visualisasi berupa diagram garis untuk melihat nilai k yang paling optimal. Dari hasil di atas dapat diketahui bahwa k=1 merupakan nilai k yang paling optimal. Oleh karena itu nilai k=1 akan digunakan sebagai nilai parameter k pada KNN. Langkah selanjutnya adalah memprediksi atau menentukan label data testing. Gambar 4 merupakan perintah untuk menampilkan hasil klasifikasi pada data testing dan untuk menampilkan grafiknya. Dengan menjalankan perintah pada gambar 4 dimana terdapat fungsi predict yang didalamnya terdapat algoritma KNN dan data yang akan diuji, maka akan dapat menghasilkan hasil klasifikasi setiap data testing dan dapat pula ditampilkan grafik jumlah mahasiswa pada setiap kelompok UKT



Gambar 3. Perintah menampilkan hasil klasifikasi dan grafiknya



Gambar 4. Hasil klasifikasi dan grafik setiap kelompok UKT

Gambar 5 menunjukkan label setiap data testing yang mana dengan melihat grafik disampingnya dapat dilihat jumlah mahasiswa pada masing-masing kelompok UKT. Mahasiswa yang ditentukan mendapat kelompok UKT 1 sebanyak 23 mahasiswa, UKT II sebanyak 149 mahasiswa, UKT III sebanyak 124 mahasiswa, UKT IV sebanyak 32 mahasiswa, dan 2 mahasiswa mendapat UKT V.

Tahap evaluasi bertujuan untuk mengetahui performa akurasi dari model yang telah dibangun (Larose, 2014). Evaluasi dilakukan pada RStudio dengan menambahkan fungsi `confusionmatrix`. Gambar 6 merupakan hasil evaluasi yang menunjukkan bahwa akurasi dari algoritma KNN sebesar 0,8758 atau 87,58%. Artinya akurasi algoritma KNN dalam menentukan atau mengklasifikasikan kelompok UKT mahasiswa pada kategori Good Classification (Maulidah et al., 2020).

```
> #Menampilkan tabel konfusi matriks performa model
> confusionMatrix(table(data_prediksi, data_testing_all$UKT), mode="everything")
Confusion Matrix and Statistics

          Accuracy : 0.8758
          95% CI   : (0.8352, 0.9093)
 No Information Rate : 0.4121
 P-value [Acc > NIR] : < 2.2e-16

          Kappa   : 0.8079
```

Gambar 5 Evaluasi akurasi algoritma KNN

Pembahasan

Berdasarkan hasil yang telah didapatkan, akurais algoritma KNN dalam menentukan kelompok UKT mahasiswa yaitu sebesar 87,58%. Angka tersebut menunjukkan bahwa akurasi algoritma KNN berada pada kategori Good Classification (Maulidah et al., 2020). Dengan temuan akurasi algoritma KNN dalam kategori tersebut, dapat menjadi rekomendasi bagi pihak penentu UKT di UIN Raden Fatah Palembang untuk dapat menerapkan algoritma tersebut dalam penentuan UKT. Sehingga mahasiswa dapat terklasifikasi dalam kelompok UKT yang sesuai dengan kondisi ekonominya masing-masing.

Pada hasil analisa dari klasifikasi yang telah dilakukan, terdapat 289 data yang diprediksi sama dari 320 data yang diuji. Adapun perbedaan hasil kelompok UKT dari penerapan algoritma KNN dengan penentuan yang lama adalah sebanyak 41 data atau sebesar 12,34%. Adanya perbedaan hasil kelompok UKT yang didapatkan, dikarenakan adanya kebijakan tertentu yang diberlakukan oleh pihak terkait. Misalnya adanya verifikasi ulang terhadap kesesuaian data yang diberikan mahasiswa dengan berkas pendukung yang diberikan dan juga penetapan kelompok UKT minimum bagi mahasiswa yang lulus jalur mandiri.

Terkait nilai k yang paling optimal dari percobaan k=1-13 yang telah dilakukan, menunjukkan bahwa k=1 memberikan akurasi tertinggi dibandingkan nilai k lain. Dengan K-Fold Cross Validation serta bantuan *software* Rstudio ditemukan nilai k optimal pada KNN yaitu k=1. Berbeda dengan penelitian sebelumnya dalam menentukan kelompok UKT menggunakan algoritma KNN yang memakai nilai k=13 sebagai parameter nilai k pada KNN. Penggunaan nilai k yang terlalu besar membuat batasan antara setiap klasifikasi menjadi kabur (Banjarsari et al., 2015). Pada penelitian ini ditemukan perbedaan nilai k optimal pada KNN yang digunakan dalam penentuan kelompok UKT mahasiswa yaitu k=1. Terdapat beberapa penelitian terdahulu yang relevan dimana menunjukkan nilai k=1 menjadi nilai k yang memberikan performa terbaik pada model KNN. (Simanjuntak et al., 2017) dalam penelitiannya untuk mengklasifikasikan penyakit tanaman kedelai menggunakan K-Nearest Neighbor mendapatkan hasil bahwa k=1 memberikan akurasi terbaik pada model sebesar 100% atau dengan akurasi rata-rata 5 percobaan sebesar 98,83%. (Laksono et al., 2020) pada penelitiannya untuk mengklasifikasikan spam dan ham email menunjukkan k=1 memberikan akurasi tertinggi pada algoritma KNN yaitu sebesar 91,4%. Penggunaan K-Fold Cross Validation dapat menemukan nilai k optimal pada algoritma KNN. Metode tersebut dapat menampilkan akurasi pada semua nilai k yang diuji sehingga dapat dilihat nilai yang terbaik. Dengan bantuan Rstudio dapat mempermudah proses tersebut yang dapat menampilkan akurasi pada semua rentang nilai k yang digunakan secara bersamaan. Berbeda dengan *software* Rapid Miner yang harus menguji coba satu persatu nilai yang diuji.

Simpulan

Proses data mining dengan teknik klasifikasi menggunakan algoritma K-Nearest Neighbor (KNN) dapat diterapkan dalam penentuan uang kuliah tunggal (UKT) mahasiswa baru dengan akurasi sebesar 87,58% yang berada dalam kategori *Good Classification*. Maka dari hasil yang didapatkan dapat direkomendasikan penerapan algoritma KNN untuk menentukan kelompok UKT mahasiswa di UIN Raden Fatah Palembang. Adapun saran bagi penelitian selanjutnya agar dapat dikembangkan dengan melakukan perbandingan beberapa algoritma klasifikasi dalam penentuan UKT. Sehingga dapat memberikan hasil yang bervariasi dan dapat ditemukan atribut yang memiliki kontribusi terbesar dalam penentuan kelompok UKT serta dapat ditemukan model dengan algoritma mana yang memberikan performa paling baik dalam penentuan UKT mahasiswa.

REFERENSI

- Banjarsari, M. A., Budiman, I., & Farmadi, A. (2015). Penerapan K-Optimal Pada Algoritma Knn untuk Prediksi Kelulusan Tepat Waktu Mahasiswa Program Studi Ilmu Komputer Fmipa Unlam Berdasarkan IP Sampai Dengan Semester 4. *Klik*, 2(2), 50-64.
- Binabar, S. W., & Ivandari. (2017). Optimasi Parameter K pada Algoritma KNN untuk Deteksi Penyakit Kanker Payudara. *IC-Tech*, 13(1), 11–18.
- Effendi, M., & Latifah, N. A. (2021). Penetapan Harga Jasa Pendidikan di Perguruan Tinggi Keagamaan Islam Negeri (PTKIN). *EDUKASIA: Jurnal Pendidikan Dan Pembelajaran*, 2(2), 127–143.
- Hidayat, W., Utami, E., Iskandar, A. F., Hartanto, A. D., & Prasetio, A. B. (2021). Perbandingan Performansi Model pada Algoritma K-NN terhadap Klasifikasi Berita Fakta Hoaks Tentang Covid-19. *Edumatic: Jurnal Pendidikan Informatika*, 5(2), 167–176. <https://doi.org/10.29408/edumatic.v5i2.3664>
- Laksono, E., Basuki, A., & Bachtiar, F. (2020). Optimization of k value in knn algorithm for spam and ham email classification. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 4(2), 377-383. <https://doi.org/10.29207/resti.v4i2.1845>
- Larose, D. T., & Larose, C. D. (2014). *Discovering knowledge in data: an introduction to data mining*. John Wiley & Sons. <https://doi.org/10.1002/9781118874059>
- Maulidah, M., Gata, W., Aulianita, R., & Agustyaningrum, C. I. (2020). Algoritma Klasifikasi Decision Tree Untuk Rekomendasi Buku Berdasarkan Kategori Buku. *E-Bisnis: Jurnal Ilmiah Ekonomi dan Bisnis*, 13(2), 89-96. <https://doi.org/10.51903/e-bisnis.v13i2.251>
- Muktamar, B. A., Setiawan, N. A., & Adji, T. B. (2015). Analisis perbandingan tingkat akurasi algoritma naïve bayes classifier dengan correlated-naïve bayes classifier. *Semnasteknomedia Online*, 3(1), 2-1.
- Nishom, M. (2019). Perbandingan Akurasi Euclidean Distance, Minkowski Distance, dan Manhattan Distance pada Algoritma K-Means Clustering berbasis Chi-Square. *Jurnal Informatika*, 4(01), 20-24. <https://doi.org/10.30591/jpit.v4i1.1253>
- Nuranti, M., Aini, M. N., & Enri, U. (2021). Komparasi Distance Measure Pada K-Medoids Clustering untuk Pengelompokan Penyakit Ispa. *Edumatic: Jurnal Pendidikan Informatika*, 5(1), 99–107. <https://doi.org/10.29408/edumatic.v5i1.3359>
- Nurmalasari, M. D., Kusriani, K., & Sudarmawan, S. (2021). Komparasi Algoritma Naïve Bayes dan K-Nearest Neighbor untuk Membangun Pengetahuan Diagnosa Penyakit Diabetes. *Jurnal Komtika (Komputasi dan Informatika)*, 5(1), 52-59. <https://doi.org/10.31603/komtika.v5i1.5140>
- Primajaya, A., Sari, B. N., & Khusaeri, A. (2020). Prediksi Potensi Kebakaran Hutan dengan Algoritma Klasifikasi C4. 5 Studi Kasus Provinsi Kalimantan Barat. *JEPIN (Jurnal Edukasi dan Penelitian Informatika)*, 6(2), 188-192. <https://doi.org/10.26418/jp.v6i2.37834>

- Purnomo, A., & Saifullah, S. (2022). Tinjauan Utilitarianisme Hukum Atas Penerapan Regulasi Uang Kuliah Tunggal (UKT) di Perguruan Tinggi Keagamaan Islam Negeri. *AL-MANHAJ: Jurnal Hukum Dan Pranata Sosial Islam*, 4(2), 229–240.
- Purwaningsih, E., & Nurelasari, E. (2021). Penerapan K-Nearest Neighbor Untuk Klasifikasi Tingkat Kelulusan Pada Siswa. *Syntax: Jurnal Informatika*, 10(01), 46-56. <https://doi.org/10.35706/syji.v10i01.5173>
- Retnoningsih, Y. D., & Marom, A. (2017). Analisis Kebijakan Penyelenggaraan Pendidikan Berbasis Uang Kuliah Tunggal Bagi Perguruan Tinggi Negeri Fakultas Ilmu Sosial Dan Ilmu Politik Universitas Diponegoro Semarang Jawa Tengah. *Journal of Public Policy and Management Review*, 6(2), 482-497.
- Schuh, G., Reinhart, G., Prote, J.-P., Sauermann, F., Horsthofer, J., Oppolzer, F., & Knoll, D. (2019). Data mining definitions and applications for the management of production complexity. *Procedia CIRP*, 81, 874–879. <https://doi.org/10.1016/j.procir.2019.03.217>
- Setianto, Y. A., Kusriani, K., & Henderi, H. (2019). Penerapan Algoritma K-Nearest Neighbour Dalam Menentukan Pembinaan Koperasi Kabupaten Kotawaringin Timur. *Creative Information Technology Journal*, 5(3), 232-241. <https://doi.org/10.24076/citec.2018v5i3.179>
- Simanjuntak, T. H., Mahmudy, W. F., & Sutrisno. (2017). Implementasi Modified K-Nearest Neighbor Dengan Otomatisasi Nilai K Pada Pengklasifikasian Penyakit Tanaman Kedelai. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 1 (2), 75-79.
- Sukanto, S., Adriyani, Y., & Aulia, R. (2020). Prediksi Kelompok UKT Mahasiswa Menggunakan Algoritma K-Nearest Neighbor. *JUITA: Jurnal Informatika*, 8(1), 121-130. <https://doi.org/10.30595/juita.v8i1.6267>
- Wahyudi, M. I., Wibowo, E. W., & Sopiullah, S. (2022). Web-Based Face Recognition using Line Edge Detection and Euclidean Distance Method. *Edumatic: Jurnal Pendidikan Informatika*, 6(1), 135–142. <https://doi.org/10.29408/edumatic.v6i1.5525>
- Witten, I. H., Frank, E., Hall, M. a., & Mark, A. (2011). *Data Mining: Practical Machine Learning Tools And Techniques*. Amsterdam: Elsevier.
- Yustanti, W. (2012). Algoritma K-Nearest Neighbour untuk Memprediksi Harga Jual Tanah. *Jurnal Matematika, Statistika dan Komputasi*, 9(1), 57-68.
- Zong, C., Xia, R., & Zhang, J. (2021). *Text Data Mining*. Singapore: Springer. <https://doi.org/10.1007/978-981-16-0100-2>