KLASIFIKASI KERINGANAN UKT MAHASISWA UHO MENGGUNAKAN K-NEAREST NEIGHBOR (KNN)

Rina, Puspita Maharani Hasan, Nurfatzma Ayu, Rizal Adi Saputra
Teknik Informatika, Universitas Halu Oleo
Kendari, Sulawesi Tenggara
waoderina75@gmail.com

ABSTRAK

Biaya kuliah sering kali menjadi tantangan bagi mahasiswa, terutama mereka yang berasal dari keluarga dengan keterbatasan ekonomi. Proses seleksi manual untuk mendapatkan keringanan Uang Kuliah Tunggal (UKT) membutuhkan waktu dan sumber daya yang signifikan, terutama ketika jumlah data yang diolah semakin besar. Dalam konteks ini, diperlukan sistem pendukung keputusan yang dapat mempercepat dan memperbaiki keakuratan seleksi. Sebagai solusi untuk masalah tersebut, penelitian ini bertujuan untuk mengembangkan sistem klasifikasi menggunakan algoritma K-Nearest Neighbor (KNN) untuk menentukan kelayakan pemberian keringanan UKT kepada mahasiswa Universitas Halu Oleo (UHO). Keringanan UKT adalah bentuk bantuan keuangan yang diberikan kepada mahasiswa berdasarkan beberapa faktor, seperti kondisi ekonomi, prestasi akademik, dan latar belakang sosial. Algoritma KNN dipilih karena kemampuannya dalam melakukan klasifikasi berdasarkan kemiripan dengan data pelatihan. KNN mencari K tetangga terdekat dari data uji dan mengklasifikasikannya berdasarkan mayoritas kelas tetangga terdekat tersebut. Dataset yang digunakan terdiri dari atribut-atribut yang relevan dengan kelayakan pemberian keringanan UKT, seperti pendapatan keluarga, indeks prestasi mahasiswa, dan kategori sosial ekonomi. Dataset telah diberi label kelayakan pemberian keringanan UKT (misalnya "Layak" dan "Tidak Layak") berdasarkan kebijakan yang ada. Hasil perhitungan menggunakan tetangga terdekat k = 2 diperoleh nilai akurasi tertinggi yaitu 90% dan nilai Error rate terendah yaitu 0.1.

Kata kunci : *kecerdasan buatan, pembelajaran mesin, k-nearest neighbor, confussion matrix, klasifikasi*

1. PENDAHULUAN

Biaya kuliah menjadi faktor penting dalam pendidikan tinggi. Untuk membantu mahasiswa yang kesulitan secara finansial, banyak lembaga pendidikan menawarkan program keringanan biaya kuliah, seperti Uang Kuliah Tunggal (UKT). Uang Kuliah Tunggal (UKT) merupakan sistem pembayaran perkuliahan dimana biaya ini ditanggung oleh setiap mahasiswa sesuai kemampuan ekonomi dari masing-masing mahasiswa. Saat ini proses pendataan mahasiswa yang mengajukan keringanan UKT masih dilakukan secara manual. Proses ini tidak akan menjadi masalah besar apabila data yang diolah dalam jumlah yang sedikit, namun bila data yang diolah dalam jumlah yang banyak, maka akan diperlukan waktu dan tenaga yang cukup lama dalam melakukan seleksi berkas. Pentingnya proses seleksi tidak dapat dipisahkan dari keputusan yang akan diambil, terutama jika mahasiswa yang mengajukan permohonan telah memenuhi persyaratan yang ditetapkan oleh pihak kampus. Oleh karena itu, diperlukan sistem pendukung keputusan yang dapat membantu menentukan mahasiswa yang berhak menerima bantuan pengurangan Uang Kuliah Tunggal (UKT) [1].

Penentuan kelayakan pemberian keringanan UKT seringkali melibatkan pertimbangan subjektif dan memakan waktu. Penentuan besaran UKT untuk setiap kelompok diterapkan secara merata bagi semua mahasiswa, terlepas dari jalur penerimaan mereka. Penetapan kelompok UKT dilakukan dengan mempertimbangkan kondisi ekonomi: a) mahasiswa; b) orang tua mahasiswa; atau c) pihak lain yang membiayai mahasiswa [2].

Penelitian ini bertujuan mengembangkan sistem klasifikasi menggunakan algoritma *K-Nearest Neighbor* (KNN) untuk menentukan kelayakan pemberian keringanan UKT.

Algoritma KNN adalah metode yang digunakan untuk mengklasifikasikan data berdasarkan jarak terdekat dari suatu objek data ke objek lainnya [3].

Dekat atau tidaknya jarak tetangga pixel itu dihitung dengan jarak *Euclidean*. Algoritma KNN bekerja dengan mencari K tetangga terdekat dari suatu data uji dalam ruang atribut, kemudian mengklasifikasikan data uji berdasarkan mayoritas kelas dari tetangga terdekat tersebut. Nilai k yang digunakan menyatakan jumlah tetangga terdekat yang dilibatkan dalam penentuan prediksi label kelas pada data uji.

Algoritma KNN telah digunakan dalam berbagai aplikasi, termasuk klasifikasi data, pengenalan pola, dan sistem rekomendasi. Dalam konteks penelitian ini, KNN digunakan untuk mengklasifikasikan kelayakan pemberian keringanan UKT kepada mahasiswa berdasarkan atribut-atribut yang relevan. *Dataset* berisi atribut kelayakan pemberian keringanan UKT seperti pendapatan orang tua, jumlah tanggungan orang tua, jumlah kendaraan yang dimiliki, dan kategori sosial ekonomi. Diharapkan sistem ini membantu lembaga pendidikan mengidentifikasi

2. TINJAUAN PUSTAKA

2.1. K-Nearest Neighbor

Metode K-Nearest Neighbor adalah salah satu metode supervised learning yang melakukan klasifikasi berdasarkan jarak ke tetangga terdekat dalam dataset pelatihan [4].

Algoritma ini menerapkan "lazy learning" atau "instant based learning" dan merupakan algoritma non parametrik. Algoritma KNN digunakan untuk klasifikasi dan regresi.

Objek yang diklasifikasikan menggunakan metode ini akan diklasifikasi berdasarkan data latih yangtetangganya paling dekat dengan objek atau selisih dari objek yang terkecil. KNN adalah sebuah metode yang dimonitori dengan hasil instance query yang diklasifikasikan berdasar pada sebagian besar kategori dalam KNN.

K-NN adalah metode klasifikasi objek berdasarkan data latih yang memiliki jarak terdekat atau beberapa kemiripan dengan objek tersebut. Jarak antar tetangga diukur menggunakan jarak *Euclidean* [5].

Adapun rumus dari perhitungan jarak *Euclidean*, sebagai berikut.

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

dengan d adalah jarak *Euclidean*, p adalah dimensi data, i adalah variabel data, x1 adalah data sampel, dan x2 adalah data *training* atau data *testing*.

2.2. *Cocfusion Matrix*

Confusion matrix adalah metode evaluasi yang menggunakan tabel matriks. Matriks ini terdiri dari beberapa sel yang menunjukkan jumlah data uji yang diklasifikasikan dengan benar dan jumlah data uji yang diklasifikasikan secara salah [6]. Metode ini digunakan untuk menghitung keakuratan konsep data mining. *Confusion matrix* dapat dinyatakan seperti tabel berikut.

Tabel 1. *Confusion Matrix*

Variabel	**Klasifikasi**	
	Layak	**Layak**
Layak	*True negative* (TN)	*False positive* (FP)
Tidak Layak	*False negative* (FN)	*True positive* (TP)

Tabel diatas merupakan tebel confusion matriks dimana *true positive* adalah prediksi hasil yang bernilai positif, data sebenarnya yang positif, *true negative* adalah prediksi hasil yang bernilai negatif, data yang sebenarnya negatif, *false positive* adalah perdiksi hasil yang bernilai positif, data yang sebenarnya negatif, dan *false negative* adalah prediksi hasil yang bernilai negatif, data yang sebenarnya positif.

Dengan metode *Confusion matrix* dapat dilakukan perhitungan dan menghasilkan empat nilai evaluasi yang mencakup penggunaan *Confusion matrix* dan menampilkan nilai-nilai pada *Classification Report, seperti Recall, Precision, Accuracy dan F1-Score* [7].

Recall: tingkat probabilitas dengan hasil yang positif dan ditentukan dengan benar. Rumus dari *recall* adalah

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

Precision: tingkat probabilitas dengan hasil yang positif dan ditentukan dengan benar. Rumus dari *precision* adalah:

$$Precision = \frac{\text{TP}}{\text{TP+FP}} \times 100\% \quad (3)$$

Accuracy: nilai hasil untuk menentukan keakuratan model. Rumus dari *accuracy* adalah:

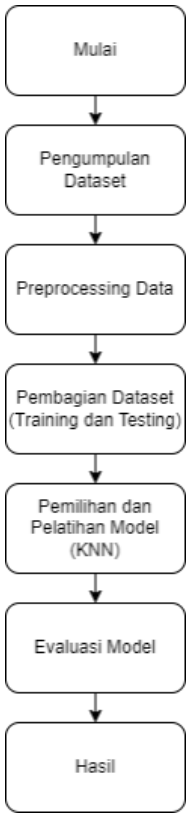
$$accuracy = \frac{\text{TP+TN}}{\text{TP+FN+FP+FN}} \times 100\% \quad (4)$$

F1 Score: adalah nilai hasil untuk mengetahui nilai rata-rata dari perbandingan keluaran *precision* maupun keluaran *Recall*. Rumus dari *F1 Score* adalah:

$$F1\ Score = 2 \times \frac{Recall \times Precision}{Recall + Precision} \times 100\% \quad (5)$$

3. METODE PENELITIAN

Pada penelitian ini dilakukan beberapa tahapan seperti yang terlihat pada gambar 1 yaitu identifikasi masalah, pengumpulan serta preprocessing dataset, persiapan data awal yang meliputi training dan testing, pengujian dengan algoritma K-Nearest Neighbor, evaluasi, serta hasil.



Gambar 1. Diagram Alir Metode Penelitian

3.1. Pengumpulan Dataset

Data yang digunakan pada penelitian ini adalah hasil survei dari beberapa mahasiswa Universitas Halu Oleo (UHO) yang ingin mengajukan keringanan UKT periode tahun dan data rujukan disetujui atau tidak

disetujui pengajuan keringanan UKT di ambil dari beberapa jurnal yang melakukan penelitian yang sama. Beberapa syarat yang harus dipenuhi untuk mendapatkan keringanan UKT bagi mahasiswa yaitu sebagai berikut.

a. Tempat Tinggal: Aspek tempat tinggal yang dimaksud adalah tempat tinggal mahasiswa saat menempuh bangku kuliah. Apakah iatinggal bersama orang tua atau menyewa kosan dapat memberikan gambaran tentang dukungan finansial yang mereka terima dan kemandirian mereka dalam memenuhi kebutuhan hidup.
b. Pekerjaan Orang Tua: Aspek pekerjaan orang tua beban mencerminkan kesejahteraan keuangan dari mahasiswa tersebut. Hal ini berpengaruh terhadap kelayakan pemberian keringanan UKT bagi Mahasiswa tersebut.
c. Pendapatan Orang Tua: Pada aspek pendapatan, mahasiswa wajib meberitahukan besaran pendapatan yang didapat per bulannya. Hal ini berpengaruh terhadap kelayakan pemberian keringanan UKT bagi Mahasiswa tersebut.
d. Jumlah Tanggungan: Aspek tanggungan mencerminkan beban biaya yang ditanggung pihak orang tua mahasiswa dalam kesehariannya. Jika jumlah tanggungan semakin besar maka memiliki kemungkinan disetujui untuk menerima keringanan UKT.
e. Kendaraan: Aspek kepemilikan kendaraan dapat memberikan gambaran tentang kondisi keuangan keluarga mahasiswa dan dapat dipertimbangkan sebagai salah satu faktor yang mempengaruhi kelayakan pemberian keringanan UKT.

3.2. *Preprocessing* Data

Preprocessing data adalah tahap yang dilakukan sebelum memulai proses pembelajaran pada data [8]. Pada tahap ini, dilakukan pemilihan variabel, dan juga penanganan *missing value*. Tidak semua variabel pada data dibutuhkan dalam mencapai tujuan penelitian, sehingga perlu dipilih beberapa variabel yang akan digunakan sebagai variabel *input* dan variabel target. Selain itu, dilakukan data *cleaning*, yang mencakup penghapusan data duplikat dan penanganan *missing value* karena data yang hilang dapat mempengaruhi hasil analisis. *Missing value* adalah kondisi di mana terdapat nilai yang tidak lengkap atau kosong pada satu atau lebih kriteria. *Missing value* merupakan masalah yang tidak diinginkan dalam *machine learning* dan data *mining*, karena dapat menyebabkan berbagai kendala dalam analisis dan pengolahan data [9].

3.3. Pembagian Data

Tahap ini *dataset* yang telah melewati *preprocessing* data dibagi menjadi dua bagian, yaitu sebagian dialokasikan pada data *training* dan sebagian pada data *testing*. Data *training* digunakan untuk membentuk model K-Nearest Neighbor, sedangkan data *testing* akan digunakan untuk klasifikasi berdasarkan model yang telah terbentuk. Pembagiannya dilakukan secara acak dengan teknik *random*. Pada sistem pengambilan keputusan proporsi yang dipakai 80% data *training* dan 20% data *testing*.

3.4. Pelatihan Model

Pada tahap ini, model K-Nearest Neighbor (KNN) digunakan untuk membangun sistem klasifikasi kelayakan keringanan UKT. KNN dipilih karena kesederhanaannya dan kemampuannya dalam menangani berbagai jenis data, baik numerik maupun kategori. Algoritma KNN bekerja dengan cara mencari k tetangga terdekat dari sebuah data baru berdasarkan jarak, dan menentukan kelas berdasarkan mayoritas tetangga tersebut.

Langkah-langkah dalam pelatihan model adalah sebagai berikut:

a. Penentuan Parameter K: Parameter K menunjukkan jumlah tetangga terdekat yang akan dipertimbangkan dalam pengambilan keputusan klasifikasi. Dalam penelitian ini, dilakukan eksperimen untuk menemukan nilai K yang optimal dengan menggunakan *cross-validation*.
b. Pemilihan Metode Pengukuran Jarak: Jarak antar data dihitung menggunakan *Euclidean Distance*, yang merupakan metode umum dalam KNN. *Euclidean Distance* mengukur jarak geometris antara dua titik dalam ruang multidimensi, yang dalam hal ini adalah variabel-variabel seperti tempat tinggal, pekerjaan orang tua, pendapatan orang tua, jumlah tanggungan, dan kepemilikan kendaraan.
c. Pelatihan Model: Dataset *training* yang telah dibagi sebelumnya digunakan untuk melatih model KNN. Model akan mempelajari hubungan antara fitur-fitur input dan kelas target (disetujui atau tidak disetujui keringanan UKT) berdasarkan data mahasiswa.
d. Normalisasi Data: Sebelum diterapkan ke model, dilakukan normalisasi data untuk memastikan bahwa variabel yang memiliki skala yang berbeda tidak mendominasi perhitungan jarak. Normalisasi dilakukan agar semua variabel berada dalam rentang yang sama.

3.5. Evaluasi Model

Setelah model KNN dilatih, langkah selanjutnya adalah melakukan evaluasi untuk mengetahui seberapa baik model tersebut dalam melakukan prediksi. Tahapan evaluasi dilakukan dengan menggunakan data testing yang telah disisihkan sebelumnya. Evaluasi dilakukan melalui beberapa metrik berikut:

a. Akurasi: Akurasi mengukur persentase prediksi yang benar dari total prediksi yang dilakukan oleh model. Akurasi merupakan salah satu metrik dasar yang menunjukkan seberapa efektif model dalam melakukan klasifikasi keringanan UKT.
b. *Precision* dan *Recall*: *Precision* mengukur seberapa tepat model dalam memprediksi

mahasiswa yang layak mendapatkan keringanan UKT dari seluruh prediksi yang dikategorikan layak. Sedangkan *recall* mengukur seberapa baik model dalam menangkap seluruh mahasiswa yang benar-benar layak mendapatkan keringanan UKT dari data aktual.

c. F1-*Score*: F1-*Score* adalah metrik yang menggabungkan precision dan recall, dan memberikan penilaian lebih seimbang ketika ada ketidakseimbangan dalam jumlah data di setiap kelas (layak dan tidak layak).
d. *Confusion Matrix*: *Confusion matrix* digunakan untuk memvisualisasikan jumlah prediksi yang benar dan salah dari setiap kelas (layak dan tidak layak). Matriks ini menunjukkan jumlah *True Positives* (TP), *True Negatives* (TN), *False Positives* (FP), dan *False Negative* (FN) yang membantu memahami jenis kesalahan prediksi yang dilakukan oleh model.

4. HASIL DAN PEMBAHASAN

4.1. Pengumpulan Dataset

Dataset yang digunakan dalam penilitian berasal dari daata yang dikumpulkan melauli survei *online* (*Google Form*) kepada para mahasiswa. *Dataset* terdiri dari 100 data UKT Mahasiswa dan data memiliki 6 atribut sebagaimana dideskripsikan pada tabel berikut.

Tabel 2. Dataset Mahasiswa

No	Atribut	Keterangan	Contoh Data
1	Tempat Tinggal	Tempat tinggal mahasiswa ketika menempuk pendidikan	0 = Bersama Orang Tua 1 = Kos/Kontrakan
2	Pekerjaan Orang Tua	Jenis pekerjaan yang dimiliki orang tua mahasiswa	PNS
3	Penghasilan Orang Tua	Besar penghasilan orang tua mahasiswa perbulan (Rupiah)	2.000.000
4	Jumlah Tanggungan Orang Tua	Jumlah anak/keluarga yang ditanggung oleh orang tua (orang)	4
5	Kendaraan	Jumlah kendaraan dari keluarga mahasiswa tersebut	2
6	Kelayakan Keringanan UKT	Hasil pengajuan kelayakan keringan UKT	0 = Layak 1 = Tidak Layak

Pada tabel diatas terdapat 6 atribut yakni 5 atribut yang digunakan sebagai atribut utama dalam menentukan klasifikasi kelayakan pemberian keringanan UKT dan 1 atribut hasil yang menjadi hasil penentuan layak atau tidaknya mahasiswa tersebut dalam mendapatkan keringanan UKT.

4.2. *Preprocessing* Data

Sebelum memulai pembentukan model, dilakukan beberapa tahapan *preprocessing* untuk memastikan kualitas data dan performa model yang optimal.

a. Pemilihan variabel yaitu dengan memiilih 5 atribut utama sebagai fitur *input* dan 1 atribut sebagai target (kelayakan keringanan UKT).
b. Penanganan *missing value* yaitu data yang hilang dapat diatasi dengan menghapus entri yang tidak lengkap.
c. Data *cleaning* yaitu melakukan penghapusan data duplikat dan normalisasi variabel-variabel numerik untuk memastikan skala yang konsisten.
d. *Feature scaling* merupakan proses normalisasi atau standarisasi nilai-nilai fitur pada dataset. Hal ini penting dilakukan agar setiap fitur memiliki skala yang serupa, sehingga tidak ada fitur yang mendominasi dalam proses pembelajaran model. Proses *scaling* dilakukan dengan *metode min-max scaling* yang mengubah nilai-nilai fitur ke dalam rentang [0,1].

4.3. Pembagian Data

Salah satu metode pembagian data adalah *splitting* data. *Spliting* data adalah proses membagi dataset menjadi subset data pelatihan (*training* set) dan subset data pengujian (test set). Splitting data dilakukan dengan tujuan untuk membagi dataset sehingga dapat digunakan untuk melatih model dan menguji kinerjanya [10]. Pada sistem pengambilan keputusan proporsi yang dipakai 80% data *training* dan 20% data *testing*.

4.4. Hasil Pelatihan Model menggunakan KNN

Model K-Nearest Neighbor (KNN) dilatih menggunakan metode supervised learning di mana 80% data digunakan sebagai training set dan 20% sebagai testing set. Tahapan pelatihan model KNN meliputi beberapa langkah berikut:

a. Pemilihan Parameter K: Nilai k yang digunakan adalah 2, yang dipilih berdasarkan hasil eksperimen sebelumnya. Nilai k ini menentukan jumlah tetangga terdekat yang dipertimbangkan oleh model dalam melakukan klasifikasi.
b. Perhitungan Jarak Euclidean: Pada tahap pelatihan, model KNN mengukur kedekatan setiap titik data baru terhadap titik-titik dalam data training. Jarak antara dua titik data dihitung menggunakan rumus jarak Euclidean, yang merupakan akar kuadrat dari jumlah kuadrat perbedaan antara koordinat titik-titik tersebut
c. Penentuan Kelas Berdasarkan Tetangga Terdekat: mengidentifikasi 2 tetangga terdekat (n*earest neighbors*) dari titik data baru. Kelas

dari titik data baru ditentukan berdasarkan mayoritas kelas dari tetangga-tetangga terdekat ini. Jika mayoritas tetangga berasal dari kelas "Layak", maka titik data tersebut akan diklasifikasikan sebagai "Layal", dan sebaliknya.

d. Pelatihan Model: Model KNN mempelajari pola-pola dalam data training, yaitu bagaimana data terdistribusi di ruang fitur dan bagaimana setiap titik data diklasifikasikan berdasarkan jarak ke tetangga terdekatnya. Tidak ada fase pembobotan seperti pada algoritma lain karena KNN langsung menggunakan data training pada saat proses prediksi.
e. Pengujian Model: Setelah tahap pelatihan selesai, model diuji menggunakan data testing yang belum pernah dilihat model sebelumnya. Proses ini bertujuan untuk mengevaluasi sejauh mana model dapat memprediksi label secara akurat pada data baru.

4.5. Evaluasi Model

Evaluasi model *K-Nearest Neighbor* (KNN) dilakukan untuk menilai performa model dalam mengklasifikasikan data.

4.6. Classification Report

Classification Report adalah pengklasifikasian dengan memberikan informasi yang lebih terperinci tentang kinerja model dalam memprediksi setiap kelas secara individu. Laporan klasifikasi mencakup beberapa metrik evaluasi, termasuk *Precision*, *Recall*, dan f1-score, untuk setiap kelas dan rata-rata tertimbang (*weighted average*).

Adapun hasil laporan klasifikasi untuk setiap kelas adalah sebagai berikut.

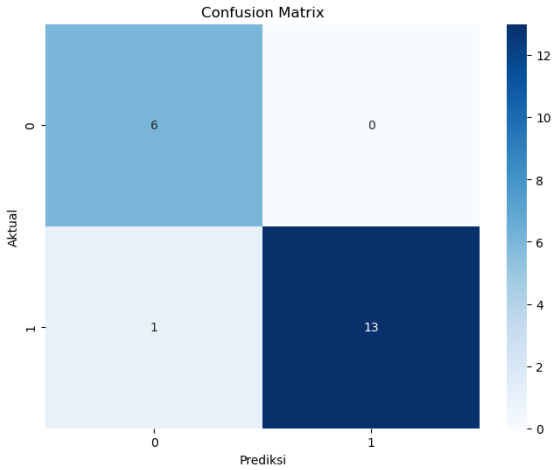
	precision	recall	f1-score	support
0	1.00	0.67	0.80	6
1	0.88	1.00	0.93	14
accuracy			0.90	20
macro avg	0.94	0.83	0.87	20
weighted avg	0.91	0.90	0.89	20

Gambar 2. *Clasification Report*

a. *Precision* untuk kelas 0 adalah 1.00 yang berarti 100% dari prediksi kelas 0 yang dilakukan oleh model adalah benar. *Precision* untuk kelas 1 adalah 0.88, yang berarti 88% dari prediksi kelas 1 adalah benar.
b. *Recall* untuk kelas 0 adalah 0.67, yang berarti 67% dari total sampel kelas 0 telah diidentifikasi dengan benar oleh model. *Recall* untuk kelas 1 adalah 1.00, yang berarti 100% dari total sampel kelas 1 telah diidentifikasi dengan benar oleh model.
c. *F1-score* untuk kelas 0 adalah 0.80, sedangkan untuk kelas 1 adalah 0.93.
d. *Support* menunjukkan jumlah sampel dalam set pengujian yang termasuk dalam setiap kelas. Dalam kasus ini, terdapat 6 sampel kelas 0 dan 14 sampel kelas 1.
e. *Accuracy* (akurasi) model adalah 0.90, yang berarti sebanyak 94% dari total sampel dalam set pengujian diprediksi dengan benar oleh model.
f. *Macro average* adalah rata-rata dari metrik evaluasi (*Precision*, *Recall*, *F1-score*) dihitung untuk setiap kelas secara terpisah, kemudian diambil rata-ratanya. *Macro average Precision*, *Recall*, dan f1-score dalam kasus ini adalah 0.88.
g. *Weighted average* adalah rata-rata dari metrik evaluasi (*Precision*, *Recall*, *f1-score*) dihitung untuk setiap kelas dengan memperhitungkan bobotnya berdasarkan jumlah sampel dalam setiap kelas. *Weighted average Precision*, *Recall*, dan f1-score dalam kasus ini adalah 0.90.

4.7. Hasil *Confusion Matrix*

Confusion matrix adalah representasi tabel yang menunjukkan jumlah prediksi yang benar dan salah yang dibuat oleh model. Matriks konfusi memiliki empat elemen utama: *True negative* (TN), *False positive* (FP), *False negative* (FN), dan *True positive* (TP). Adapun hasil *confusion matrix* untuk data uji yang telah dilakukan adalah sebagai berikut.

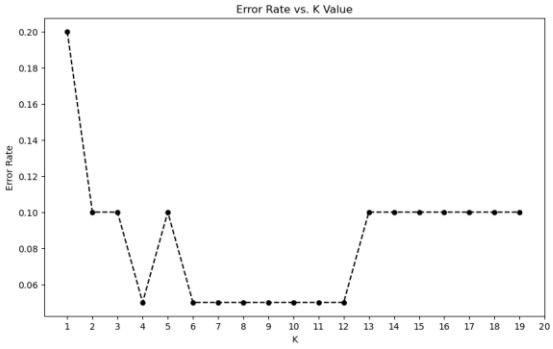


Gambar 3. *Confusion Matrix*

Berdasarkan gambar diatas dapat dilihat bahwa model memprediksi 6 sampel dengan benar sebagai kelas 0 (*true negative*), 0 sampel salah diprediksi sebagai kelas 1 (*false positive*), memprediksi 1 sampel dengan salah sebagai kelas 0 (*false negative*), 13 sampel benar diprediksi sebagai kelas 1 (*true positive*).

4.8. *Error Rate*

Error rate adalah ukuran yang digunakan untuk mengukur sejauh mana prediksi model yang dibuat berbeda dari nilai yang sebenarnya dalam dataset. *Error rate* sering digunakan dalam konteks pemodelan statistik, pembelajaran mesin, dan pengujian hipotesis. Adapun grafik yang menampilkan hubungan antara nilai K (jumlah tetangga) dengan tingkat kesalahan (*error rate*) yang dihasilkan adalah sebagai berikut.



Gambar 4. Grafik Nilai *Error Rate*

Grafik tersebut menggambarkan hubungan antara tingkat kesalahan (error rate) dan nilai K pada model K-Nearest Neighbors (KNN). Pada sumbu X, ditampilkan berbagai nilai K yang digunakan dalam model, mulai dari 1 hingga 20, sementara sumbu Y menunjukkan tingkat kesalahan yang terkait dengan setiap nilai K. Berdasarkan grafik, terlihat bahwa pada K=1, tingkat kesalahan sangat tinggi, sekitar 0,2. Namun, tingkat kesalahan ini menurun drastis setelah K=2, dengan penurunan signifikan pada nilai K=3 dan K=5 di mana tingkat kesalahan hampir mencapai nol, menunjukkan performa terbaik model. Setelah nilai K=10, tingkat kesalahan stabil di sekitar 0,1 dan tidak mengalami perubahan signifikan meskipun nilai K meningkat. Dengan demikian, dapat disimpulkan bahwa model bekerja paling baik pada K=3 atau K=5, dan peningkatan nilai K setelah itu tidak memberikan perbaikan yang berarti terhadap kinerja model.

5. KESIMPULAN DAN SARAN

Berdasarkan tahapan-tahapan yang telah dilakukan dalam penerapan algoritma K-Nearest Neighbor (KNN) untuk memprediksi kelayakan pemberian Uang Kuliah Tunggal (UKT), dapat disimpulkan bahwa penggunaan tetangga terdekat dengan nilai k = 2 menghasilkan akurasi sebesar 90% dan error rate terendah yaitu 0.1, sementara penggunaan k = 1 menunjukkan akurasi terendah sebesar 80% dan error rate tertinggi 0.2. Kesimpulan ini diperoleh dari analisis confusion matrix yang menunjukkan nilai True Negative (TN) = 6, False Positive (FP) = 0, False Negative (FN) = 1, dan True Positive (TP) = 13. Oleh karena itu, disarankan untuk mempertimbangkan penggunaan nilai k = 2 dalam penerapan model KNN ini untuk meningkatkan keakuratan prediksi dalam penentuan kelayakan UKT, serta melanjutkan penelitian dengan mempertimbangkan faktor-faktor lain yang mungkin memengaruhi hasil klasifikasi.

DAFTAR PUSTAKA

[1] F. Claudya Lahinta, S. N. Rumokoy, S. Pendukung Keputusan Penentuan Penerimaan Pengurangan UKT, S. P. Junaedy, and A. S. Luntungan, “Sistem Pendukung Keputusan Penentuan Penerimaan Pengurangan UKT Mahasiswa Politeknik Negeri Manado Di Era Pandemi Covid-19 Menggunakan Metode SAW,” *TEKNO Jurnal Teknologi Elektro dan Kejuruan*, vol. 32, no. 2, pp. 334–345, 2022, doi: 10.17977/um034v32i2p334-345.

[2] R. Susetyoko, W. Yuwono, E. Purwantini, and N. Ramadijanti, “Perbandingan Metode Random Forest, Regresi Logistik, Naïve Bayes, dan Multilayer Perceptron Pada Klasifikasi Uang Kuliah Tunggal (UKT),” *Jurnal Infomedia: Teknik Informatika, Multimedia & Jaringan*, vol. 7, no. 1, pp. 8–16, 2022, doi: 10.30811/jim.v7i1.2916.

[3] S. R. Cholil, T. Handayani, R. Prathivi, and T. Ardianita, “Implementasi Algoritma Klasifikasi K-Nearest Neighbor (KNN) Untuk Klasifikasi Seleksi Penerima Beasiswa,” *IJCIT (Indonesian Journal on Computer and Information Technology)*, vol. 6, no. 2, pp. 118–127, 2021, doi: 10.31294/ijcit.v6i2.10438.

[4] A. Syarifah, A. A. Riadi, and A. Susanto, “Klasifikasi Tingkat Kematangan Jambu Bol Berbasis Pengolahan Citra Digital Menggunakan Metode K-Nearest Neighbor,” *JIMP : Jurnal Informatika Merdeka Pasuruan*, vol. 7, no. 1, pp. 27–34, 2022, doi: 10.37438/jimp.v7i1.417.

[5] F. Marisa *et al.*, “Pengukuran Tingkat Kematangan Kopi Arabika Menggunakan Algoritma K-Nearest Neighbour,” *JIMP : Jurnal Informatika Merdeka Pasuruan*, vol. 6, no. 3, p. 1, 2021, doi: 10.51213/jimp.v6i3.280.

[6] R. N. Amalda, N. Millah, and I. Fitria, “IMPLEMENTASI ALGORITMA C5.0 DALAM MENGANALISA KELAYAKAN PENERIMA KERINGANAN UKT MAHASISWA ITK,” *Teorema: Teori dan Riset Matematika*, vol. 7, no. 1, pp. 101–116, Mar. 2022, doi: 10.25157/teorema.v7i1.6692.

[7] A. Razaki, Y. Herry Chrisnanto, and Melina, “PENANGANAN OUTLIER PADA METODE ALGORITMA K-NEAREST NEIGHBORS (KNN) DENGAN METODE KERNEL DENSITY ESTIMATION PADA KASUS PENYAKIT DIABETES HANDLING OUTLIERS IN THE K-NEAREST NEIGHBORS (KNN) ALGORITHM USING KERNEL DENSITY ESTIMATION IN DIABETES CASES,” *Journal of Information Technology and Computer Science (INTECOMS)*, vol. 7, no. 4, p. 1177, 2024, doi: 10.31539/intecoms.v7i4.10866.

[8] F. Rozi, F. Sukmana, and M. N. Adani, “Pengelompokkan Judul Buku dengan Menggunakan Algoritma K-Nearest Neighbor (K-NN) dan Term Frequency – Inverse Document Frequency (TF-IDF),” *JIMP : Jurnal Informatika Merdeka Pasuruan*, vol. 6, no. 3, pp. 1–5, 2021, doi: 10.51213/jimp.v6i3.346.

[9] W. Sudrajat and I. Cholid, “K-NEAREST NEIGHBOR (K-NN) UNTUK PENANGANAN MISSING VALUE PADA DATA UMKM,”

Jurnal Rekayasa Sistem Informasi dan Teknologi, vol. 1, no. 2, pp. 54–63, 2023, doi: 10.59407/jrsit.v1i2.77.

[10] J. M. A. S. Dachi and P. Sitompul, "Analisis Perbandingan Algoritma XGBoost dan Algoritma Random Forest Ensemble Learning pada Klasifikasi Keputusan Kredit," *JURNAL RISET RUMPUN MATEMATIKA DAN ILMU PENGETAHUAN ALAM (JURRIMIPA)*, vol. 2, no. 2, pp. 87–1003, Oct. 2023, doi: 10.55606/jurrimipa.v2i2.1336.