

## Analisis Klasifikasi Konsentrasi Mahasiswa Menggunakan Algoritma K-Nearest Neighbor

Almi Yulistia Alwanda<sup>1\*</sup>, Ema Utami<sup>2</sup>, Ainul Yaqin<sup>3</sup>

<sup>1,2,3</sup>Program Studi Magister Teknik Informatika, Universitas Amikom Yogyakarta

\* [almi.yulistia@students.amikom.ac.id](mailto:almi.yulistia@students.amikom.ac.id)

Universitas Hamzanwadi di Nusa Tenggara Barat memiliki Fakultas Teknik yang merupakan salah satu pilar penting dalam menyediakan pendidikan tinggi berkualitas. Fakultas Teknik memiliki empat program studi yang sangat diminati oleh mahasiswa, yaitu Teknik Informatika, Sistem Informasi, Teknik Komputer, dan Teknik Lingkungan. Pertumbuhan jumlah mahasiswa baru di fakultas ini mencerminkan keberhasilannya, terutama karena prospek kerja yang semakin besar dalam bidang-bidang tersebut. Namun, mahasiswa sering menghadapi tantangan kompleks dalam memilih konsentrasi studi yang sesuai dengan minat dan potensi mereka. Dalam respons terhadap kebutuhan ini, Fakultas Teknik Universitas Hamzanwadi menyediakan beragam opsi konsentrasi studi atau peminatan seperti Data Science, RPL dan Multimedia. Dalam rangka membantu mahasiswa membuat keputusan yang lebih baik tentang konsentrasi studi, penelitian ini menggunakan algoritma K-Nearest Neighbor (KNN) dalam analisis klasifikasi konsentrasi mahasiswa. Metode penelitian ini bersifat eksperimental dan menggunakan teknik pengumpulan data seperti observasi, wawancara, dan kuisioner. Data yang digunakan ada tujuh atribut seperti NIM, Jenis Kelamin, IPK, Nilai Mata Kuliah Data Science, RPL dan Multimedia yang diproses menggunakan algoritma KNN melalui Google Colab. Hasil penelitian menunjukkan bahwa dengan  $k=2$  dan 8-fold cross-validation, akurasi yang diperoleh adalah 67%..

**Kata kunci :** Klasifikasi, Konsentrasi, K-Nearest Neighbor

### Abstract

*Hamzanwadi University located in West Nusa Tenggara boasts a Faculty of Engineering, which plays a pivotal role in delivering top-notch higher education. This faculty offers four highly coveted programs—Informatics Engineering, Information Systems, Computer Engineering, and Environmental Engineering—that attract a significant number of students. The increase in enrollment in these programs underscores the faculty's success, particularly owing to the promising job opportunities in these fields. Nevertheless, students often encounter intricate challenges when selecting their area of specialization that resonates with their interests and capabilities. In addressing this concern, the Faculty of Engineering at Hamzanwadi University provides diverse concentration options like Data Science, RPL, and Multimedia. To aid students in making informed decisions regarding their study concentration, this study employs the K-Nearest Neighbor (KNN) algorithm to analyze the classification of student concentrations. This research adopts an experimental approach and utilizes data collection methods such as observation, interviews, and surveys. The dataset comprises seven attributes including NIM, Gender, GPA, Data Science Course Grade, RPL Course Grade, and Multimedia Course Grade, processed using the KNN algorithm through Google Colab. The research outcomes reveal that with  $k=2$  and 8-fold cross-validation, the achieved accuracy stands at 67%.*

**Keywords :** Classification, Concentration, K-Nearest Neighbor

### 1. Pendahuluan

Universitas Hamzanwadi, yang berlokasi di Nusa Tenggara Barat, telah membuktikan dirinya sebagai salah satu institusi pendidikan tinggi terkemuka di wilayah tersebut. Dengan

enam fakultas yang mencakup berbagai bidang studi, Fakultas Teknik menjadi salah satu tiang penting dalam menyelenggarakan pendidikan tinggi yang bermutu. Fakultas Teknik memiliki peran yang sangat penting dengan

menyelenggarakan empat program studi, yakni Teknik Informatika, Sistem Informasi, Teknik Komputer, dan Teknik Lingkungan.

Keberhasilan Fakultas Teknik, terutama dalam program-program studi tersebut, tercermin dari pertumbuhan tahunan jumlah mahasiswa baru. Prospek pekerjaan yang semakin luas di bidang-bidang ini menjadi alasan utama di balik peningkatan yang signifikan ini. Dengan demikian, minat yang semakin meningkat dari mahasiswa untuk bergabung dengan program studi di Fakultas Teknik menimbulkan kebutuhan akan penelitian yang dapat membantu mahasiswa dalam membuat keputusan yang tepat mengenai konsentrasi studi atau peminatan yang paling sesuai dengan minat dan potensi mereka.

Namun, proses pengambilan keputusan mengenai konsentrasi studi seringkali merupakan tantangan yang kompleks bagi mahasiswa. Faktor-faktor seperti minat pribadi, potensi akademis, dan tuntutan industri menjadi pertimbangan yang perlu dipertimbangkan. Sebagai respons terhadap kebutuhan ini, institusi pendidikan tinggi, khususnya Fakultas Teknik Universitas Hamzanwadi menyediakan beragam opsi konsentrasi studi atau peminatan yang dapat dipilih oleh mahasiswa.

Perkembangan teknologi data mining saat ini merupakan sebuah alat yang sangat efisien dalam ranah teknologi informasi, terutama di

dalam konteks bisnis yang penuh persaingan, terutama ketika masyarakat memasuki era Big Data. Dalam beberapa tahun belakangan, penerapan data mining dalam sektor pendidikan telah menjadi semakin umum di perguruan tinggi, di mana teknologi ini digunakan untuk melakukan prediksi terkait mahasiswa, termasuk evaluasi kinerja mereka dalam proses pembelajaran, analisis kelulusan tepat waktu, mengidentifikasi potensi mahasiswa yang berisiko drop out, mengelola penerimaan mahasiswa baru, menentukan minat mahasiswa, dan berbagai aspek lainnya.

Klasifikasi konsentrasi mahasiswa bukanlah tugas yang sederhana, mengingat keragaman preferensi dan faktor-faktor yang mempengaruhi keputusan mahasiswa.. Penelitian terdahulu telah mencoba berbagai metode untuk mengatasi tantangan ini, namun demikian, ada potensi untuk meningkatkan keakuratan dan efisiensi analisis. Dalam konteks ini, penelitian sebelumnya telah mencoba memanfaatkan algoritma machine learning, yaitu Support Vector Machine (SVM), untuk memprediksi konsentrasi studi mahasiswa. Meskipun demikian, hasil penelitian sebelumnya menunjukkan tingkat akurasi yang belum memuaskan, sekitar 66%, terutama saat menggunakan metode validasi silang (cross-validation) dengan 5 fold [1].

Ketidaktimalan ini menggugah kebutuhan akan penyelidikan lebih lanjut untuk meningkatkan performa model klasifikasi. Penelitian ini bertujuan untuk menggali potensii algoritma KNN dalam analisis klasifikasi konsentrasi mahasiswa. Dengan mengintegrasikan data mahasiswa dan karakteristik konsentrasi program studi, kita dapat memperoleh wawasan mendalam tentang pola preferensi mahasiswa dan faktor-faktor yang memengaruhi pilihan mereka.

Algoritma KNN merupakan salah satu pendekatan yang menjanjikan dalam analisis klasifikasi karena sifatnya yang sederhana dan intuitif. KNN mengklasifikasikan data berdasarkan kemiripan dengan tetangga terdekat, yang membuatnya cocok untuk analisis konsentrasi mahasiswa di mana pola dan tren mungkin terjadi dalam konteks lokal

## 2. Tinjauan Pustaka

### 2.1. Penelitian Terkait

Dalam menjalankan penelitian ini, penulis merujuk pada penelitian terdahulu yang berkaitan dengan topik penelitian ini antar lain :

- Yogie Indra dan rekan-rekannya melakukan studi yang berjudul "Klasifikasi Penentuan Pengajuan Kartu Kredit Menggunakan Metode K-Nearest Neighbor". Tujuan dari penelitian ini adalah untuk mempermudah pihak bank atau analis

dalam menentukan kategori kartu kredit yang sesuai untuk nasabah bank dengan lebih akurat. Mereka menggunakan metode K-Nearest Neighbor untuk mengidentifikasi pola-pola dalam data nasabah, dengan variabel pendukung seperti jenis kelamin, status rumah, status, jumlah tanggungan (anak), profesi, dan penghasilan per tahun. Hasil penelitian menunjukkan bahwa nilai presisi rata-rata adalah 92%, nilai recall adalah 83%, dan nilai akurasi adalah 93%.[2].

- Riska Sagita bersama timnya melakukan penelitian menggunakan KNN untuk mendapatkan hasil berita clickbait atau bukan clickbait yang dimana waktu proses pada KNN lebih cepat daripada waktu proses yang lainnya. Dari hasil penelitian yang dilakukan untuk pengklasifikasian berita clickbait, didapatkan hasil terbaik pada jumlah  $k=11$  dengan menggunakan scenario 1 pada pembagian data dengan jumlah data sebanyak 800 data latih dan 200 data uji yang menghasilkan akurasi sebesar 71%, Precision 72% dan Recall 71%. Hal ini menunjukkan bahwa klasifikasi berita clickbait dapat di klasifikasikan menggunakan K-Nearest Neighbor[3].
- Aria Pratama dkk melakukan penelitian dengan judul "Klasifikasi Penerima Beasiswa Dengan Menggunakan Algoritma

K-Nearest Neighbor". Penelitian ini dilakukan untuk membantu mengambil keputusan dalam menentukan siapa yang layak mendapatkan beasiswa menggunakan K-Nearest Neighbor. Data diolah sebanyak 383 peserta didik dengan parameter Nama Siswa, alamat, alat transportasi, nama orang tua, pekerjaan, penghasilan, nama ibu, pekerjaan ibu, penghasilan ibu. Dari hasil pengujian kemudian dihasilkan akurasi sebesar 78.45%[4].

- Luqman Abdur Rahman dkk melakukan penelitian dengan judul "Faktor-Faktor Yang Mempengaruhi Minat Calon Mahasiswa Baru Mendaftar Pada FTII UHAMKA Menggunakan Algoritma K-Nearest Neighbor". Penelitian ini dilakukan akibat ketiadaan pengolahan data terkait factor-faktor yang mempengaruhi minat calon mahasiswa baru memutuskan daftar di Fakultas Teknologi Industri dan Informatika. Parameter yang digunakan seperti nama, jurusan, biaya kuliah, lokasi FTII dengan domisili, terdapat teman/keluarga, akreditasi, fasilitas, pelayanan pmb, informasi pmb dan keterangan minat. Dari hasil pengujian dengan menggunakan metode KNN dihasilkan akurasi sebesar 72.5% dengan teknik K-5 Fold[5].

- Nora Lizarti dan rekan-rekannya melakukan penelitian berikutnya yang bertujuan untuk menentukan konsentrasi studi di STMIK Amik Riau memanfaatkan algoritma KNN Pengujian dilakukan menggunakan perangkat RapidMiner untuk mengukur kinerja algoritma. Hasil pengujian terhadap 183 data latih dan 100 data uji menunjukkan bahwa algoritma KNN memiliki kinerja yang sangat baik, dengan akurasi sebesar 98%, recall 100%, presisi 100%, F-measure 91.67%, dan tingkat kesalahan klasifikasi yaitu 2%[6].

## 2.2. Landasan Teori

### 1. Data Mining

Data mining merujuk pada bidang ilmu yang bertujuan untuk menganalisis dan mengolah data dalam jumlah besar guna menghasilkan informasi yang berarti. Pendekatan data mining memanfaatkan teknik statistik, matematika, dan kecerdasan buatan untuk mengekstraksi serta mengidentifikasi informasi penting dan pengetahuan yang tersembunyi dari berbagai sumber data besar. Bidang ini terus berkembang pesat, terutama seiring dengan meningkatnya kebutuhan akan nilai tambah dari basis data yang semakin besar sejalan dengan kemajuan teknologi informasi. Dalam praktiknya, data mining melibatkan berbagai teknik, metode, dan algoritma yang beragam yang dapat

dikombinasikan secara tepat untuk menghasilkan hasil analisis yang akurat.

Dalam proses klasifikasi, terdapat variabel target yang memiliki sifat kategoris dan dibagi ke dalam berbagai kelas yang sudah ditentukan Seperti kategori nasabah yang mengalami masalah atau tidak, atau hewan yang termasuk dalam kategori reptil, amfibi, mamalia, burung, atau ikan dalam klasifikasi. Klasifikasi proses penentuan model pola dengan menggambarkan perbedaan kelas data sehingga bisa digunakan untuk mencapai tujuan prediksi kelas yang belum diketahui (Annur Haditsah, 2018) [7]. Setiap algoritma bisa menghasilkan klasifikasi yang berbeda. Kualitas algoritma dapat dinilai dari seberapa baik model mampu mengklasifikasikan data sesuai dengan kebenaran yang sebenarnya, atau seberapa tepat model dapat memprediksi kelas klasifikasi. Teknik pada klasifikasi yang sering digunakan adalah seperti Naïve Bayes, K-Nearest Neighbor (KNN) dan Support Vector Machine (SVM).

## 2. K-Nearest Neighbor

K-NN, singkatan dari K-Nearest Neighbors, merupakan sebuah teknik pengklasifikasian data di mana data tersebut telah sebelumnya diklasifikasikan. Tingkat akurasi algoritma K-NN sangat dipengaruhi oleh berbagai karakteristik, terutama ketika nilai fitur tidak sesuai dengan perkiraan yang ada. Beberapa penelitian yang

menggunakan algoritma K-NN secara khusus fokus pada pemilihan fitur dan pembobotan untuk meningkatkan efisiensi algoritma dalam proses klasifikasi.

Algoritma K-NN termasuk dalam kategori instance-based learning, di mana data training disimpan untuk memungkinkan pencarian klasifikasi untuk rekaman baru yang belum terklasifikasi. Hal ini dilakukan dengan membandingkan sebagian besar persamaan atau kesamaan dengan data training yang ada. Peran utama K-NN adalah menemukan jarak terdekat antar data, diikuti dengan mengevaluasi tetangga terdekat sebanyak K pada data latih. Data training direpresentasikan dalam ruang multidimensi, masing-masing mencerminkan karakteristik data [8].

Berikut adalah langkah-langkah perhitungan dalam algoritma K-NN:

- Menentukan Nilai K yang berupa bilangan bulat positif
- Menghitung jarak antar data baru dan dengan data training

$$D = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2}$$

Diketahui:

- $x$  : data sampel
- $y$  : data pengujian jangkauan
- Menentukan urutan jarak (jangkauan) paling dekat dengan jarak min pada K
- Label dari tetangga terdekat dimanfaatkan untuk melakukan prediksi terhadap data

baru. Sedangkan persamaan *K-NN* dalam menghitung nilai prediksi.

Algoritma *K-NN* memiliki beberapa kelebihan, di antaranya keefektifan dalam menghadapi data berjumlah banyak dan kemampuan untuk menghasilkan prediksi dengan tingkat akurasi yang tinggi. Meskipun demikian, terdapat juga beberapa kekurangan yang perlu diperhatikan. Salah satu kekurangan utama dari algoritma *K-NN* adalah kebutuhan untuk menentukan nilai *k* yang optimal. Pemilihan nilai *k* yang tidak tepat dapat mempengaruhi performa algoritma dan menghasilkan prediksi yang kurang akurat. Oleh karena itu, proses pemilihan nilai *k* menjadi kritis dalam implementasi *K-NN*.

Selain itu, algoritma *K-NN* juga memerlukan biaya komputasi yang tinggi, terutama dalam menghitung jarak antara instance pada query dengan data latih. Proses ini dapat menjadi memakan waktu, terutama jika datasetnya sangat besar atau memiliki dimensi yang tinggi. Oleh karena itu, perlu mempertimbangkan secara hati-hati trade-off antara akurasi dan biaya komputasi saat menggunakan algoritma *K-NN*.

### 3. Klasifikasi

Klasifikasi adalah sebuah teknik dalam data mining yang sering digunakan dalam bidang pembelajaran di mana proses analisis data dilakukan untuk Membuat model-model yang

menggambarkan kategori-kategori dari data yang sedang dievaluasi [9].

#### 4. K-Fold Cross Validation

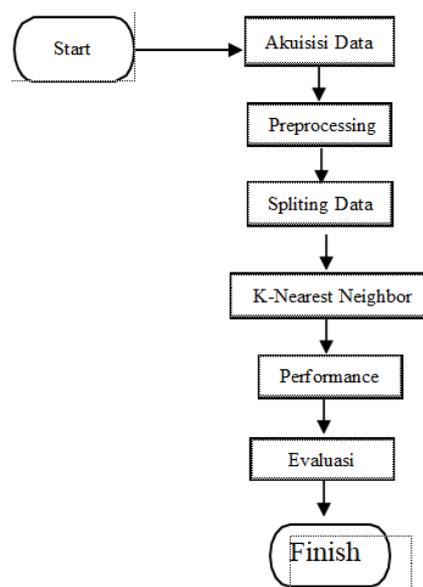
*K-Fold Cross Validation* merupakan suatu teknik yang digunakan untuk melatih dan menguji dataset dengan pembagian yang efisien menggunakan ukuran *K* yang sama untuk mengukur akurasi. Dalam proses ini, nilai *K* ditetapkan sebagai sepuluh partisi dan diulang sepuluh kali untuk mencari akurasi dan kinerja model secara rata-rata dari sepuluh kali pengujian [10].

#### 5. Confusion Matrix

*Confusion Matrix* merupakan suatu teknik yang digunakan untuk mengevaluasi hasil kinerja dari metode klasifikasi [11].

### 2.3. Tahapan Penelitian

Adapun alur proses pada penelitian ini dapat terlihat pada gambar 1 dibawah ini :



Gambar 1 Alur proses penelitian

Alur proses penelitian diatas terdiri dari beberapa tahapan yaitu:

1. Akuisisi data merupakan tahap mengumpulkan data yang akan diolahData bisa didapatkan melalui beragam cara, seperti interaksi langsung melalui wawancara atau pengisian kuisisioner., observasi atau pengambilan data dari sumber yang ada. Metode yang digunakan untuk memperoleh data harus konsisten dengan tujuan penelitian dan asumsi yang telah ditentukan sebelumnya. Dataset yang dipakai dalam penelitian ini diperoleh langsung dari bagian Staff akademik Fakultas Teknik Universitas Hamzanwadi.
2. Preprocessing merupakan adalah tahap yang digunakan untuk mempersiapkan data sebelum analisis dilakukan. Terdapat serangkaian tahapan yang perlu diselesaikan. Ini mencakup langkah-langkah seperti membersihkan data, mengubah format data, dan mengintegrasikan data. Tahap ini merupakan tahap awal sebelum pengujian model algoritma yang digunakan. Dataset yang tersedia akan dimodifikasi menjadi data yang siap diuji, tujuannya adalah untuk menguji karakteristik data terhadap model algoritma klasifikasi. Proses pembersihan data melibatkan penghapusan entri data yang kosong atau tidak relevan, sementara transformasi data melibatkan

perubahan struktur data untuk memenuhi kebutuhan penelitian.

3. Pemisahan data adalah proses membagi dataset yang digunakan dalam penelitian menjadi dua atau lebih subset., biasanya digunakan untuk menguji model atau algoritma. *Dataset* umumnya dibagi menjadi data (*training*) dan data uji (*testing*). Data latih digunakan melatih algoritma, sementara data uji untuk evaluasi kinerja pada algoritma.
4. Kemudian data yang sudah siap dapat diproses menggunakan Google Colab dengan algoritma K-Nearest Neighbor, disini dilakukan proses pemilihan atribut dan target untuk output yang diinginkan
5. Performance merupakan adalah Metrik atau ukuran yang diterapkan untuk menilai kualitas atau ketepatan model atau algoritma yang diterapkan dalam penelitian. [12]. Kinerja dapat dinilai melalui matriks kebingungan dengan menggunakan berbagai metrik seperti akurasi, presisi, recall, dan lainnya sesuai dengan tujuan dan konsep penelitian yang telah ditetapkan. Evaluasi merupakan tahapan yang digunakan untuk menilai kualitas atau keakuratan model atau algoritma dalam penelitian, yang bisa dilakukan melalui metode uji coba dengan menggunakan data uji. Metode yang digunakan harus sesuai

dengan hipotesis, tujuan, dan kerangka penelitian yang telah ditetapkan sebelumnya.

### **3. Metode Penelitian**

Metode penelitian ini adalah bersifat eksperimental, karena penelitian dilakukan dengan eksperimen agar mendapatkan hasil akurasi yang terbaik agar dapat digunakan sebagai nilai untuk memprediksi. Adapun beberapa teknik dalam melakukan penelitian.

#### **3.1. Teknik Pengumpulan Data**

Pada penelitian ini data yang digunakan diperoleh dari Fakultas Teknik Universitas Hamzanwadi yang termasuk juga ke dalam data private yang belum digunakan dalam penelitian-penelitian sebelumnya. Data diperoleh dengan cara observasi dan wawancara serta penyebaran kuisioner kepada mahasiswa.

##### **1. Teknik Observasi**

Salah satu teknik pengumpulan data yang dilakukan dengan pengamatan langsung ke lokasi penelitian yaitu ke Fakultas Teknik Universitas Hamzanwadi.

##### **2. Teknik Wawancara**

Wawancara dilakukan di lokasi penelitian dengan staff bagian akademik di Fakultas Teknik Universitas Hamzanwadi.

##### **3. Kuisioner**

Teknik angket ini dilakukan untuk mempermudah saat melakukan wawancara dan juga sebagai alat bukti untuk penginputan data.

Penjelasan terinci mengenai metode penelitian diperlukan. Bagian ini minimal harus menjelaskan desain penelitian yang dipilih, alasan di balik pemilihan desain tersebut, prosedur penelitian yang dilaksanakan, populasi dan sampel penelitian atau partisipan penelitian, instrumen penelitian yang digunakan, teknik pengumpulan data, teknik analisis data, dan tempat dilaksanakannya penelitian.

##### **3.2. Teknik Analisa Data**

Metode analisis data yang digunakan dalam penelitian ini adalah pendekatan kuantitatif, yang dipilih karena sesuai dengan lingkup penelitian yang dilakukan. Data yang diperoleh berasal dari bagian akademik Fakultas Teknik Universitas Hamzanwadi, dan instrumen penelitian terdiri dari data serta wawancara yang dilakukan di lokasi penelitian.

##### **3.3. Teknik Pengolahan Data**

Dalam pengelolaan data, perangkat lunak yang dimanfaatkan adalah Google Colab dengan penerapan metode klasifikasi serta pemanfaatan algoritma K-Nearest Neighbor.

##### **3.4. Model Yang diusulkan**

Studi ini merekomendasikan penggunaan metode klasifikasi menggunakan algoritma K-Nearest Neighbor. K-Nearest Neighbors (KNN)

menjadi pilihan yang relevan karena pendekatan ini berfokus pada kesamaan antar data. Dalam konteks konsentrasi mahasiswa, KNN dapat mengidentifikasi pola dan hubungan antara mahasiswa yang memiliki konsentrasi serupa berdasarkan atribut tertentu seperti minat atau nilai akademis. KNN memberikan fleksibilitas dan dapat memberikan wawasan tentang bagaimana preferensi konsentrasi mahasiswa berkumpul dalam ruang atribut

#### 4. Hasil dan Pembahasan

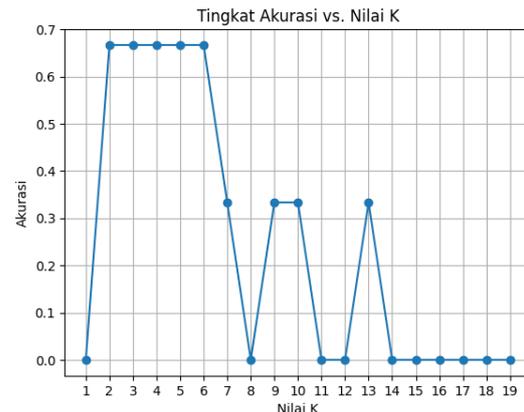
Dalam penelitian ini, algoritma K-Nearest Neighbor diterapkan untuk melakukan klasifikasi terhadap konsentrasi mahasiswa dalam program studi Sistem Informasi di Fakultas Teknik Universitas Hamzanwadi. Pengolahan data dilakukan menggunakan Google Colab dengan metode klasifikasi dan penerapan algoritma K-Nearest Neighbor. Dari data yang baru terkumpul atribut yang digunakan sebanyak 6 (enam) atribut, antara lain: NIM, Jenis Kelamin, IPK, Nilai Mata Kuliah Data Science, Nilai Mata Kuliah RPL dan Nilai Mata Kuliah Multimedia. Dari tujuh atribut tersebut dapat berfungsi sebagai faktor penentu dalam pembuatan keputusan akhir dari konsentrasi mahasiswa.

Tabel 1 Atribut yang digunakan

No	Atribut
1	NIM

No	Atribut
2	Jenis Kelamin
3	IPK
4	Nilai Mata Kuliah Data Science
5	Nilai Mata Kuliah RPL
6	Nilai Mata Kuliah Multimedia

Optimisasi nilai k dilakukan melalui metode k-fold cross-validation. Hasil dari 8-fold cross-validation menunjukkan bahwa nilai akurasi yang diperoleh adalah 67%, dengan nilai k sebesar 2 ( $k=2$ ). Nilai ini akan digunakan sebagai parameter k untuk proses KNN. Berikut adalah hasil lengkap dari 8-fold cross-validation.



Gambar 2. Tingkat Akurasi Vs. Nilai K

Proses KNN dilakukan menggunakan Google Colab. Tahap awal melibatkan evaluasi tingkat akurasi melalui validasi 8-fold dengan  $k=2$ . Rincian langkah-langkah validasi silang menggunakan Google Colab dapat dilihat pada Gambar 3 berikut.

```

[8] from sklearn.model_selection import KFold
     from sklearn.neighbors import KNeighborsClassifier
     from sklearn.metrics import accuracy_score

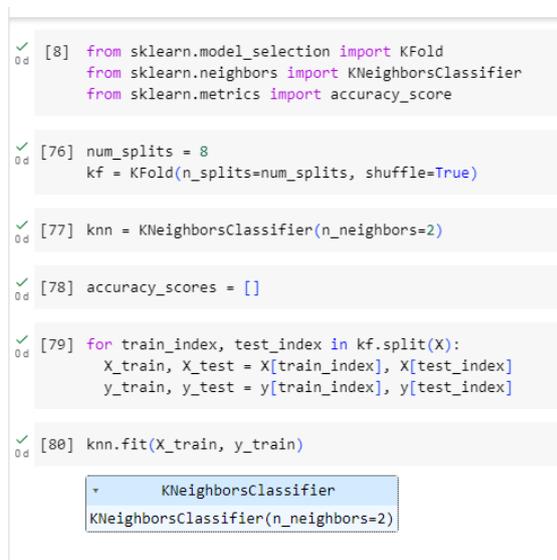
[76] num_splits = 8
     kf = KFold(n_splits=num_splits, shuffle=True)

[77] knn = KNeighborsClassifier(n_neighbors=2)

[78] accuracy_scores = []

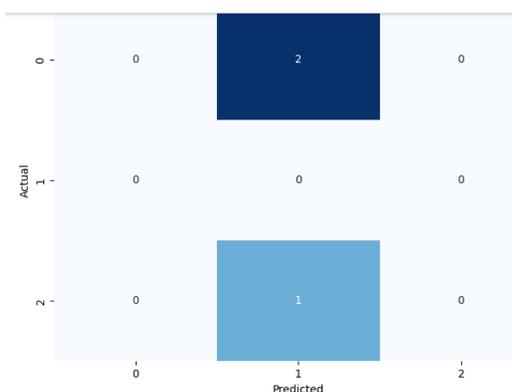
[79] for train_index, test_index in kf.split(X):
     X_train, X_test = X[train_index], X[test_index]
     y_train, y_test = y[train_index], y[test_index]

[80] knn.fit(X_train, y_train)
  
```



Gambar 3. Tahapan di Google Colab

Dapat dilihat pada gambar 2 proses atau tahapan menentukan cross validation. Dimulai dari import library yang dibutuhkan kemudian menentukan jumlah cross validation yang dimana pada penelitian ini menggunakan 8-fold cross validation, selanjutnya menentukan k yang akan digunakan dan di penelitian ini menggunakan k=2. Dari hasil 8-fold validation dan k=2 didapatkan akurasi sebesar 67% dengan confusion matrix yang dapat dilihat pada gambar 4.



Gambar 4. Confusion Matrix

## 5. Kesimpulan

Dari hasil penelitian yang telah dilakukan, Hasil penelitian menunjukkan bahwa dengan k=2 dan 8-fold cross-validation, akurasi yang diperoleh adalah 67%. Berdasarkan hasil penelitian yang telah dilakukan, kesimpulannya adalah bahwa penggunaan metode k-Nearest Neighbors (k=2) dengan menerapkan 8-fold cross-validation, akurasi model klasifikasi untuk memprediksi konsentrasi studi mahasiswa mencapai 67%. Ini menunjukkan bahwa model tersebut memberikan tingkat keakuratan yang moderat dalam memprediksi preferensi konsentrasi studi mahasiswa

## 6. Daftar Pustaka

- [1] Wattimury, F.K., & Seniwati, E. (2019). Penentuan Peminatan Mahasiswa Prodi Informatika di Universitas AMIKOM Yogyakarta Menggunakan SVM. *INTECHNO Journal*, 1(4), 15. e-ISSN: 2655-1438. p-ISSN: 2655-1632.
- [2] Kurniawan, Y. I., & Barokah, T. I. (2020). Klasifikasi Penentuan Pengajuan Kartu Kredit Menggunakan K-Nearest Neighbor. *Jurnal Ilmiah MATRIK*, 22(1), 73. ISSN: 1411-1624. e-ISSN: 2621-8089.
- [3] Sagita, R., Enri, U., & Primajaya, A. (2020). Klasifikasi Berita Clickbait Menggunakan K-Nearest Neighbor (KNN). *Journal of Information System*, 5(2), 230-239. DOI: 10.33633/joins.v5i2.3705.
- [4] Pratama, A., Ma'ruf, F. A., Rinaldi, I. A. R., & Faturrhohman. (2021). Klasifikasi Penerima Beasiswa Dengan Menggunakan Algoritma K Nearest

- Neighbor. *Jurnal Data Science & Informatika*, 1(1), 11-15.
- [5] Malik, L. A. R., Kamayani, M., & Hasan, F. N. (2023). Faktor-Faktor yang Mempengaruhi Minat Calon Mahasiswa Baru Mendaftar pada FTII UHAMKA Menggunakan Algoritma K-Nearest Neighbor (K-NN). *Infotech: Journal of Technology Information*, 9(1), 85.
- [6] Lizarti, N., & Ulfah, A.N. (2019). Penerapan Algoritma K-Nearest Neighbor Untuk Penentuan Peminatan Studi STMIK Amik Riau. *Fountain of Informatics Journal*, 4(1), 1. ISSN: 2541-4313 (Print) / 2548-5113 (Online).
- [7] Putri, A., Hardiana, C. S., Novfuja, E., Siregar, F. T. P., Rahmadden, Y., Fatma, Y., & Wahyuni, R. (2023). Comparison of K-NN, Naive Bayes and SVM Algorithms for Final-Year Student Graduation Prediction. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, Vol. 3, Iss. 1, April 2023, pp: 20-26. ISSN(P): 2797-2313, ISSN(E): 2775-8575.
- [8] Wibowo, A.M., Kasih, P., & Farida, I.N. (2024). Sistem Bantu Penentuan Konsentrasi Mahasiswa Menggunakan Metode K-Nearest Neighbor Classification. Dalam *Prosiding Seminar Nasional Teknologi dan Sains Tahun 2024*, Vol. 1. Program Studi Teknik Informatika, Universitas Nusantara PGRI Kediri. e-ISSN: 2828-299X.
- [9] A. Y. Perdana, R. Latuconsina, A. Dinimaharawati, and U. Telkom, "Prediksi Stunting Pada Balita Dengan Algoritma Random," vol. 8, no. 5, pp. 6650-6656, 2021.
- [10] M. Asrol, P. Papilo, and F. E. Gunawan, "ScienceDirect Procedia ScienceDirect Support Vector Machine with K-fold Validation to Improve the Support Vector Machine with K-fold Validation to Improve the Industry ' s Sustainability Performance Classification Industry ' s Sustainability," *Procedia Comput. Sci.*, vol. 179, no. 2020, pp. 854-862, 2021, doi: 10.1016/j.procs.2021.01.074.
- [11] A. Residencia, "Klasifikasi Kelayakan Peserta Pengajuan Kredit Rumah Dengan Algoritma Naïve Bayes Di Perumahan Azzura Residencia," vol. 9, pp. 43-48, 2019.
- [12] M. Azhari, "Perbandingan Akurasi, Recall, dan Presisi Klasifikasi pada Algoritma C4.5, Random Forest, SVM, dan Naive Bayes," vol. 5, no. April, pp. 640-651, 2021, doi: 10.30865/mib.v5i2.2937.
- [13] Y. Yahya and H. Bahtiar, "Pengaruh Pertumbuhan Ekonomi Terhadap Tingkat Kesejahteraan Masyarakat Kecamatan Selong Kabupaten Lombok Timur – Nusa Tenggara Barat Menggunakan Algoritma Naive Bayes," *Infotek J. Inform. dan Teknol.*, vol. 4, no. 1, pp. 20-28, 2021, doi: 10.29408/jit.v4i1.2981.
- [14] M. Saiful1\*, Syamsuddin2, and Moh. Farid Wajdi3, "Implementasi Algoritma Naive Bayes Untuk Memprediksi Predikat Ketuntasan Belajar Siswa Pasca Pandemi Covid 19," *Infotek J. Inform. dan Teknol.*, vol. 4, no. 1, pp. 96-104, 2021.
- [15] B. A. Candra Permana and I. K. Dewi Patwari, "Komparasi Metode Klasifikasi Data Mining Decision Tree dan Naive Bayes Untuk Prediksi Penyakit Diabetes," *Infotek J. Inform. dan Teknol.*, vol. 4, no. 1, pp. 63-69, 2021, doi: 10.29408/jit.v4i1.2994.