

**IMPLEMENTASI K-NEAREST NEIGHBOR DALAM PREDIKSI MAHASISWA  
BERHENTI KULIAH**



**DISUSUN OLEH :**

- **Fenny Jong** (535210001)
- **Kelvin Ferdinand** (535210027)
- **Jefri Jaya** (535210048)

**UNIVERSITAS TARUMANAGARA  
JAKARTA  
TAHUN 2023**

## **ABSTRAK**

Mahasiswa merupakan seorang pelajar yang duduk dan terdaftar pada salah satu perguruan tinggi baik negeri maupun swasta, menjadi seorang mahasiswa merupakan impian banyak pelajar diseluruh dunia dan menjadi mahasiswa merupakan gerbang awal untuk menentukan seseorang akan menggeluti dunia keilmuan dalam bidang apa, baik ilmu komputer, kedokteran, dunia pendidikan dan lain-lainnya. Namun sangat banyak penyebab mahasiswa memutuskan untuk berhenti mengikuti perkuliahan secara tiba-tiba yang diakibatkan oleh beberapa faktor, baik faktor external maupun faktor internal. Hal tersebut menyebabkan kerugian tersendiri yang akan dihadapi oleh kampus salah satunya berkurangnya kuantitas data mahasiswa dan mengakibatkan penumpukan data, diperlukan prediksi mahasiswa yang berpotensi berhenti kuliah secara sepihak dengan melihat beberapa kriteria dan menggali informasi terhadap data mahasiswa yang berpotensi untuk berhenti kuliah dengan menerapkan algoritma K-NN. Pada penelitian ini algoritma K-NN merekam data lama dan melihat kemiripan terhadap data baru dalam upaya pengenalan pola mahasiswa berhenti kuliah, hasil yang didapatkan dari data kuliah baru menunjukkan kemiripan data dengan data lama mahasiswa yang berhenti kuliah dengan kemiripan nilai terdekat dari kasus lainnya sehingga hasil yang didapatkan keputusan mahasiswa data baru tersebut memutuskan kemungkinan berhenti kuliah.

# **BAB I**

## **PENDAHULUAN**

### **1.1 Latar Belakang**

Pendidikan merupakan suatu proses yang membantu manusia untuk memperoleh pengetahuan, pengalaman, didikan etika, kehidupan dan masa depan, pendidikan merupakan jenjang dan tahapan yang banyak diterima oleh masyarakat luar dan pendidikan menjadi identitas diri manusia itu sendiri, dengan adanya pendidikan yang sudah diakui oleh banyak pihak maka pendidikan pada akhirnya menjadi patokan terhadap pengetahuan dan jabatan seseorang dikalangan masyarakat luas sehingga banyak masyarakat yang berlomba-lomba ingin memperoleh pendidikan sampai pendidikan tingkat tinggi. Banyak masyarakat yang mempersiapkan keuangan demi mendapatkan anaknya atau diri sendiri untuk memperoleh pendidikan tinggi, banyak rintangan ketika seseorang menduduki pendidikan khususnya pendidikan tinggi membuat peserta didik banyak yang disebut dengan mahasiswa memilih untuk berhenti melanjutkan pendidikannya.

Kasus mahasiswa berhenti kuliah merupakan kasus yang tidak asing lagi dan sering terjadi dalam perkuliahan. Berhentinya mahasiswa juga disebabkan banyak faktor, baik faktor akademik, non akademik ataupun dari personalnya yang menyebabkan serta mempengaruhi mahasiswa tersebut untuk berhenti kuliah, jenis berhenti kuliah ada beberapa tahapan seperti. Berhentinya mahasiswa dari perkuliahan tidak hanya merugikan diri mahasiswa itu secara personal saja, tetapi berdampak juga kepada perguruan tinggi seperti bertambahnya pengurangan mahasiswa, tidak diketahuinya mahasiswa yang memilih berhenti kuliah sehingga terjadinya penumpukan data hal tersebut juga akan mempersulit dosen dalam mengetahui mahasiswa kelas mana yang harus dibuang.

Perlu melakukan sebuah prediksi tentang faktor-faktor yang mempengaruhi berhentinya mahasiswa dari perkuliahan. Agar dapat mengurangi permasalahan tersebut dan juga dapat melakukan tindak lanjut terhadap mahasiswa yang memilih berhenti memutuskan perkuliahan salah satunya dengan menggunakan teknik prediksi dalam Algoritma K-NN, dimana Algoritma KNN merupakan sebuah proses untuk menemukan pengetahuan yang menarik dari data berhentinya mahasiswa.

Penelitian terdahulu memperlihatkan algoritma KNN sangat membantu dalam banyak permasalahan seperti memperlihatkan klasifikasi tingkat stres seseorang melalui tweet yang ada pada akun twitter, pada tahapan ini data dilakukan seleksi fitur terlebih dahulu terhadap data status

yang ada pada twitter dan dilakukan klasifikasi terhadap tingkat stress pemilik akun dengan metode KNN, terlihat metode KNN mampu menyelesaikan permasalahan klasifikasi dengan baik.

Penelitian terdahulu memperlihatkan algoritma KNN dalam menentukan pemberian pinjaman kredit pensiunan yang memberikan hasil terbaik dari setiap keputusan dan mampu mengklasifikasikan data berdasarkan kriteria yang sudah ditentukan. Penelitian lainnya metode ini juga digunakan dalam memprediksi cuaca di kota Palembang dimana jejak data rekaman cuaca yang ada di kota Palembang dilakukan klasifikasi menggunakan algoritma KNN untuk mempermudah memprediksi kejadian terhadap cuaca dimasa yang akan mendatang.

Pada penelitian ini metode K-NN digunakan sebagai tahapan klasifikasi dalam menentukan mahasiswa yang akan berhenti kuliah secara sepihak untuk membantu perguruan tinggi dalam menghindari permasalahan yang telah disebutkan sebelumnya, menggunakan metode KNN dikarenakan banyak penelitian yang lainnya menyatakan tingkat akurasi dan pengenalan pola yang baik terhadap kasus yang lama dalam menyelesaikan permasalahan baru.

## **BAB II**

### **METODE PENELITIAN**

#### **2.1 Pengumpulan dan Pengolahan Data**

Tahapan penelitian ini melakukan pengumpulan dan data dengan cara sebagai berikut ini :

1. Observasi

Observasi merupakan tahapan penelitian dalam pengumpulan data secara langsung, mengamati langsung untuk mendapatkan hasil yang terpercaya.

2. Dokumentasi

Hal ini perlu dilakukan untuk mendapatkan data yang direkam di dalam sebuah data file, record dan lain-lain, fungsi dari data ini merupakan bukti dari adanya hal yang menjadi landasan dari penelitian ini.

3. Studi Pustaka

Studi pustaka untuk melengkapi data teoritis dan data yang memperlihatkan penelitian terdahulu sebagai arahan kemana penelitian ini akan dilakukan.

4. Analisa

Analisa pada penelitian ini melihat kepada permasalahan yang menjadi topik utama untuk diketahui tahapan apa yang selanjutnya akan dilakukan, sebagai bentuk penyelesaian permasalahan dengan memanfaatkan ilmu komputer dan memilih tahapan dan metode yang akan dilakukan pada proses penyelesaian permasalahan.

5. Tahapan pengujian

Tahapan pengujian digunakan metodologi kuantitatif dimana metode ini menampilkan hasil matematik. Hasil yang didapatkan merupakan hasil dari penggunaan metode K-Nearest Neighbor, hasil sebagai keputusan pada penelitian ini.

6. Kesimpulan

Kesimpulan dilakukan dengan memberitahu hasil dan poin-poin penting dalam penelitian secara terperinci, tepat, singkat dan padat.

#### **2.2 K-Nearest Neighbor (KNN)**

Adapun tahapan dalam penyelesaian permasalahan menggunakan metode K-nearest Neighbor, sebagai berikut ini:

1. Menentukan parameter K, parameter ini nantinya akan menjadi bahan utama dalam batasan yang akan dilakukan pencarian dengan menggunakan metode ini, agar penelitian dan penelitian ini memiliki pondasi awal dari gambaran hal yang ingin ditemukan.
2. Menghitung kuadrat jarak euclidean masing-masing obyek terhadap data sampel yang diberikan
3. Mengurutkan objek-objek tersebut ke dalam kelompok yang mempunyai jarak euclid terkecil
4. Mengumpulkan kategori Y
5. Dengan menggunakan kategori mayoritas, maka dapat diprediksikan nilai query instance yang telah dihitung

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

Keterangan ;,

- $d(x, y)$  adalah jarak antara data  $x$  ke data  $y$
- $x_i$  adalah data testing ke- $i$
- $y_i$  adalah data training ke- $i$

### BAB III

#### HASIL DAN PEMBAHASAN

##### 3.1 Studi Kasus dan Penyelesaian

Pertama: tentukan terlebih dahulu jumlah jarak yang akan menjadi parameter, pada penelitian ini menjadi 3 (Tiga) Jumlah tetangga terdekat ( $K=3$ ).

Tabel berikut merupakan data mahasiswa yang berhenti kuliah

**Tabel 1.** Data Mahasiswa Yg Berhenti Kuliah

No	Nama	Semester	Penghasilan/Jam	IPK	Keterangan
1	Lilis	2	Rp.2500	3.55	Lanjut
2	Yeni	4	Rp.2600	2.92	Berhenti
3	Putri	4	Rp.2700	3.40	Berhenti
4	Putra	2	Rp.2800	2.95	Berhenti
5	Roni	4	Rp.2900	3.20	Lanjut

Pertanyaan seorang calon mahasiswa baru dengan data sebagai berikut:

**Tabel 2.** calon mahasiswa baru

No	Nama	Jurusan	Penghasilan	IPK	Keterangan
1	Santi	T.Informatika	Rp.2400	3.4	??

Berdasarkan data di atas, apakah mahasiswa tersebut lanjut atau berhenti kuliah.

Algoritma K-NN ini tidak dapat memproses nilai atribut kriteria dalam bentuk huruf oleh karena itu, nilai atribut jurusan diubah ke bentuk variabel angka seperti berikut:

**Tabel 3.** Transformasi data

No	Nama	Semester	Penghasilan	IPK	Keterangan
1	Santi	2	Rp.2400	3.4	??

Kedua : Proses perhitungan jarak menggunakan persamaan euclidean Distan

$$\begin{aligned} d(x3, C) &= \sqrt{(a - ca)^2 + (b - cb)^2 + (c1 - c2)^2 + (d2 - cd)^2} \\ &= \sqrt{(2500 - 2400)^2 + (1 - 1)^2 + (3 - 3)^2 + (2 - 4)^2} = 100,02 \end{aligned}$$

$$\begin{aligned} d(x2, C) &= \sqrt{(a - ca)^2 + (b - cb)^2 + (c1 - c2)^2 + (d2 - cd)^2} \\ &= \sqrt{(2600 - 2400)^2 + (2 - 1)^2 + (2 - 3)^2 + (2 - 4)^2} = 200,015 \end{aligned}$$

$$\begin{aligned} d(x4, C) &= \sqrt{(a - ca)^2 + (b - cb)^2 + (c1 - c2)^2 + (d2 - cd)^2} \\ &= \sqrt{(2700 - 2400)^2 + (1 - 1)^2 + (3 - 3)^2 + (2 - 4)^2} = 300,007 \end{aligned}$$

$$\begin{aligned} d(x5, C) &= \sqrt{(a - ca)^2 + (b - cb)^2 + (c1 - c2)^2 + (d2 - cd)^2} \\ &= \sqrt{(2800 - 2400)^2 + (1 - 1)^2 + (3 - 3)^2 + (2 - 4)^2} = 400,005 \end{aligned}$$

$$\begin{aligned} d(x6, C) &= \sqrt{(a - ca)^2 + (b - cb)^2 + (c1 - c2)^2 + (d2 - cd)^2} \\ &= \sqrt{(2900 - 2400)^2 + (1 - 1)^2 + (3 - 3)^2 + (2 - 4)^2} = 500,004 \end{aligned}$$

Setelah dihitung selanjutnya urutkan nilai jaraknya dari terkecil ke nilai yang terbesar seperti berikut:

**Tabel 4.** Mengurutkan Data Berdasarkan Jarak Terkecil

No	Nama	Semester	Penghasilan	IPK	Keterangan	Jarak Data Training ke Data Testing
1	Lilis	2	Rp.2500	3.55	Lanjut	100,02
2	Yeni	4	Rp.2600	2.92	Berhenti	200,015
3	Putri	4	Rp.2700	3.40	Berhenti	300,007
4	Putra	2	Rp.2800	2.95	Berhenti	400,005
5	Roni	4	Rp.2900	3.20	Lanjut	500,004



**Tabel 5. Hasil**

No	Nama	Semester	Penghasilan	Ipk	Keterangan
1	Santi	2	Rp.2400	3.4	Berhenti

Pada tabel hasil diatas menunjukan kecenderungan dari nilai terdekat denga nilai lainnya berakhir kepada hasil mahasiswa yang bernama santi tidak melanjutkan perkuliahnya.

## Lampiran hasil code KNN

```
import pandas as pd
from sklearn.neighbors import KNeighborsClassifier
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score

dataset = pd.read_csv('mahasiswa.csv', delimiter=',')
dataset
```

	No	Nama	Semester	Penghasilan	IPK	keterangan
0	1	Lilis	2	2500	3.55	1
1	2	Yeni	4	2600	2.92	0
2	3	Putri	4	2700	3.40	0
3	4	Putra	2	2800	2.95	0
4	5	Roni	4	2900	3.20	1

```
from sklearn.model_selection import train_test_split
X = dataset.drop(columns = "keterangan")
y = dataset.keterangan

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.2, random_state = 0)
X_train.shape, X_test.shape, y_train.shape, y_test.shape
```

```
((4, 5), (1, 5), (4,), (1,))
```

```
from sklearn.compose import ColumnTransformer
from jcopml.pipeline import num_pipe, cat_pipe
preprocessor = ColumnTransformer([
    ('numeric', num_pipe(), ['No', 'Semester', 'Penghasilan', 'IPK']),
    ('categorical', cat_pipe(encoder = 'onehot'), ['Nama']),
])
```

```
from sklearn.pipeline import Pipeline
pipeline = Pipeline([
    ('prep', preprocessor),
    ('algo', KNeighborsClassifier(n_neighbors=3, p=2, metric = 'euclidean')),
])
```

```
pipeline.fit(X_train, y_train)
```

```
Pipeline(steps=[('prep',
                  ColumnTransformer(transformers=[('numeric',
                                                  Pipeline(steps=[('imputer',
                                                                    SimpleImputer(strategy='median'))]),
                                                  ['No', 'Semester',
                                                   'Penghasilan', 'IPK']),
                                                  ('categorical',
                                                  Pipeline(steps=[('imputer',
                                                                    SimpleImputer(strategy='most_frequent')),
                                                                    ('onehot',
                                                                    OneHotEncoder(handle_unknown='ignore'))]),
                                                  ['Nama'])])),
                  ('algo',
                  KNeighborsClassifier(metric='euclidean', n_neighbors=3))])
```

```
y_pred = pipeline.predict(X_test)
```

```
y_pred
```

```
array([0], dtype=int64)
```

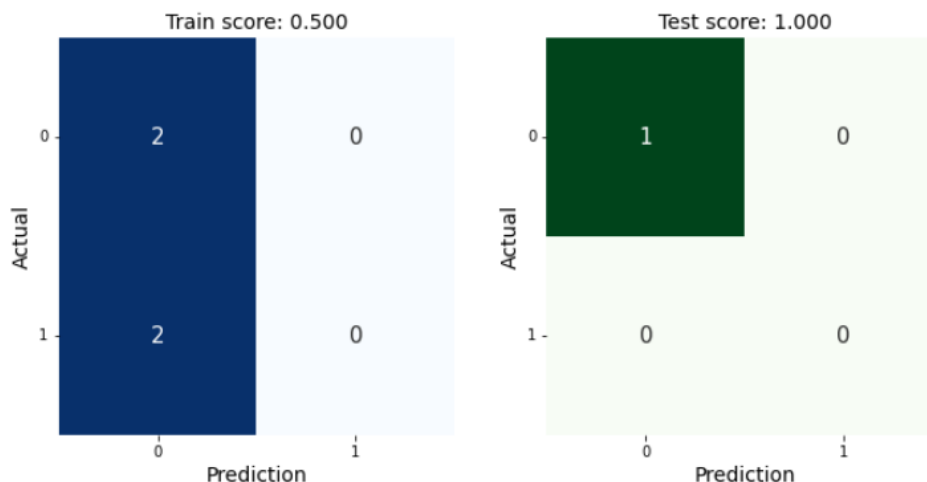
```
accuracy = accuracy_score(y_test, y_pred) * 100
```

```
print("Akurasi: {:.2f}%".format(accuracy))
```

```
Akurasi: 100.00%
```

```
from jcopml.plot import plot_confusion_matrix
```

```
plot_confusion_matrix(X_train, y_train, X_test, y_test, pipeline)
```



## **BAB IV**

### **KESIMPULAN**

Dengan dibuatnya penelitian ini, kita dapat mengetahui tingkat prediksi mahasiswa berhenti kuliah, yang cukup sering terjadi pada lingkungan masyarakat, baik itu dari segi pendapatan dan pembiayaan hidup serta kebutuhan lain yang meningkatkan faktor mahasiswa memutuskan untuk berhenti kuliah. Dengan digunakannya metode ini, prediksi mahasiswa berhenti kuliah yang tentunya dapat mempermudah banyak kalangan untuk lebih mengetahui penyebab mahasiswa yang berhenti kuliah secara sepihak dan sering kali membuat pihak kampus mengalami kerugian bahkan keterlambatan dalam pelaporan data mahasiswa. Pada hasil data di atas memperlihatkan perkiraan mahasiswa akan berhenti kuliah, ternyata lebih cenderung kepada tetap lanjut perkuliahan.

## DAFTAR PUSTAKA

- [1] K. Arizona, Z. Abidin, and R. Rumansyah, "Pembelajaran Online Berbasis Proyek Salah Satu Solusi Kegiatan Belajar Mengajar Di Tengah Pandemi Covid-19," *J. Ilm. Profesi Pendidik.*, vol. 5, no. 1, pp. 64–70, 2020.
- [2] M. A. Firmansyah, "Analisis Hambatan Belajar Mahasiswa Pada Mata Kuliah Statistika," *J. Penelit. dan Pembelajaran Mat.*, vol. 10, no. 2, 2017.
- [3] M. Muhamad, A. P. Windarto, and S. Suhada, "Penerapan Algoritma C4.5 Pada Klasifikasi Potensi Siswa Drop Out," *KOMIK (Konferensi Nas. Teknol. Inf. dan Komputer)*, vol. 3, no. 1, pp. 1–8, 2019.
- [4] D. A. Anggoro and N. D. Kurnia, "Comparison of accuracy level of support vector machine (SVM) and K-nearest neighbors (KNN) algorithms in predicting heart disease," *Int. J. Emerg. Trends Eng. Res.*, vol. 8, no. 5, pp. 1689–1694, 2020.
- [5] E. P. W. Mandala, "Data Mining Algoritma Nearest Neighbor Untuk Memprediksi Tingkat Resiko Pinjaman Dana Di Bank Perkreditan Rakyat," *JIK J. Ilmu Komput.*, vol. 1, no. 2, pp. 91–103, 2016.
- [6] M. I. Maulana and A. A. Soebroto, "Klasifikasi Tingkat Stres Berdasarkan Tweet pada Akun Twitter menggunakan Metode Improved k-Nearest Neighbor dan Seleksi Fitur Chi-square," vol. 3, no. 7, pp. 6662–6669, 2019.
- [7] V. Alfani, "Data Mining Untuk Klasifikasi Pinjaman Kredit Pensiunan Menggunakan Algoritma K-Nearest Neighbor," *J. Pelita Inform.*, vol. 18, no. April, pp. 281–286, 2019.
- [8] J. T. Informatika and U. Sriwijaya, "Prediksi Cuaca di Kota Palembang Berbasis," pp. 9–18.
- [9] Albi Anggito and Johan Setiawan, *Metodologi Penelitian Kuantitatif*. Jawa Barat: CV Jejak, 2018.
- [10] G. Hendro, T. B. Adji, and N. A. Setiawan, "Penggunaan Metodologi Analisa Komponen Utama ( PCA ) untuk Mereduksi Faktor-Faktor yang Mempengaruhi Penyakit Jantung Koroner," *Semin. Nas. ScrETec*, pp. 1–5, 2012.
- [11] E. W. Winarni, *Teori dan Praktik Penelitian Kualitatif dan Kuantitatif PTK dan R&D*. Jakarta: Bumi Aksara, 2018.
- [12] E. Buulolo, *Data Mining Untuk Perguruan Tinggi*. Deepublish, 2020.
- [13] M. P. Simatupang and D. P. Utomo, "Analisa Testimonial Dengan Menggunakan Algoritma Text Mining Dan Term Frequency- Inverse Document Frequence (Tf-Idf) Pada Toko Allmeeart," *KOMIK (Konferensi Nas. Teknol. Inf. dan Komputer)*, vol. 3, no. 1, pp. 808–814, 2019.

- [14] Y. Zhang, G. Cao, B. Wang, and X. Li, "A novel ensemble method for k-nearest neighbor," *Pattern Recognit.*, vol. 85, pp. 13–25, 2019.
- [15] M. R. Islam, A. R. M. Kamal, N. Sultana, R. Islam, M. A. Moni, and A. Ulhaq, "Detecting Depression Using K-Nearest Neighbors (KNN) Classification Technique," *Int. Conf. Comput. Commun. Chem. Mater. Electron. Eng. IC4ME2 2018*, no. February, pp. 1–4, 2018.
- [16] M. T. Masud, M. A. Mamun, K. Thapa, D. H. Lee, M. D. Griffiths, and S. H. Yang, "Unobtrusive monitoring of behavior and movement patterns to detect clinical depression severity level via smartphone," *J. Biomed. Inform.*, vol. 103, 2020.
- [17] A. Adiwijaya, M. N. Aulia, M. S. Mubarak, U. N. W, and F. Nhita, "A Comparative Study of MFCC-KNN and LPC-KNN for Hijaiyyah Letters Pronunciation Classification System," in *International Conference on Information and Communication Technology (ICoICT)*, 2017.