KLIK: Kajian Ilmiah Informatika dan Komputer
ISSN 2723-3898 (Media Online)
Vol 4, No 5, April 2024, Hal 2637-2645
DOI 10.30865/klik.v4i5.1795
https://djournals.com/klik

Data Mining Classification Untuk Prediksi Jumlah Mahasiswa Aktif dan Cuti Angkatan 2020 Menggunakan Metode K-Nearest Neighbor

Sekar Ayu Fitria, Harma Oktafia Lingga Wijaya*, Davit Irawan

Fakultas Ilmu Teknik, Program Studi Sistem Informasi, Universitas Bina Insan, Lubuk Lingau, Indonesia
Email: [1]sekarayufitria1@gmail.com, [2,*]harmaoktafialingga@gmail.com, [3]davit_irawan@univbinainsan.ac.id
Email Penulis Korespondensi: harmaoktafialingga@gmail.com

Abstrak−Mahasiswa aktif adalah mahasiswa yang mengikuti perkuliahan, sedangkan mahasiswa cuti adalah mahasiswa yang tidak mengikuti perkuliahan. Agar tetap berstatus aktif, mahasiswa harus melakukan registrasi ulang pada awal semester. Pada penelitian kali ini peneliti menggunakan data mahasiswa tahun angkatan 2020 yang bersumber dari Universitas Bina Insan Kota Lubuklinggau. Metode yang digunakan adalah data mining. Data mining merupakan istilah yang digunakan untuk menggambarkan penemuan pengetahuan dalam basis data. Tujuan penelitian ini untuk mnegetahui berapa banyak jumlah mahasiswa yang aktif dan cuti pada angkatan 2022. Data akan dianalisis menggunakan bahasa pemrograman python dengan algoritma K-Nearest Neighbor. Dalam klasifikasi mahasiswa aktif dan cuti didapatkan nilai akurasi sebesar 89%.

Kata Kunci: Aktif dan Cuti; Data Mining; Classification; K-Nearest Neighbor

Abstract−Active students are students who attend lectures, while students on leave are students who do not attend lectures. To remain active, students must re-register at the start of the semester. In this study, researchers used student data from the class of 2020 sourced from Bina Insan University, Lubuklinggau City. The method used is data mining. Data mining is a term used to describe the discovery of knowledge in databases. Data will be analyzed using the Python programming language with the K-Nearest Neighbor algorithm. In the classification of active and leave students, an accuracy value of 89% was obtained.

Keywords: Aktif and Leave; Data Mining; Classification; K-Nearest Neighbor

1. PENDAHULUAN

Teknologi Informasi adalah suatu teknologi yang digunakan untuk mengolah data, termasuk memproses, mendapatkan, menyusun, menyimpan, memanipulasi data dalam berbagai cara untuk menghasilkan informasi yang berkualitas, yaitu informasi yang relevan, akurat dan tepat waktu, yang digunakan untuk keperluan pribadi, bisnis, dan pemerintahan dan merupakan informasi yang strategis untuk pengambilan keputusan. Seiring dengan semakin pesatnya perkembangan teknologi informasi tersebut, maka kebutuhan teknologi komputerisasi diharapkan juga semakin baik, guna menghasilkan suatu informasi[1]. Teknologi ini menggunakan seperangkat komputer untuk mengolah data, jaringan untuk menghubungkan satu komputer dengan komputer lainnya bila diperlukan, dan teknologi telekomunikasi yang digunakan agar data dapat disebarluaskan dan diakses secara global. Teknik menjadikan pembelajaran lebih praktis dan *flexibel*. Universitas Bina Insan Lubuklinggau merupakan salah satu perguruan tinggi yang ada di kota Lubuklinggau yang terletak di Jl. Jenderal Moh Suharto Km 13, Desa Lubuk Kupang, Kecamatan Lubuklinggau Selatan I, Kota Lubuklinggau[2]. Universitas Bina Insan didirikan pada tahun 1996 ketika Yayasan Dwi Tunggal Palembang pemilik STIE dan STMIK Musirawas menggabungkan dua perguruan tinggi menjadi satu universitas. Saat ini terdapat beberapa program studi yang telah mempunyai akreditasi baik dan sangat baik, dibuktikan dengan SK BAN-PT NO.7016/SK/BAN-PT/Akred/S/XI/2020. Saat ini Universitas Bina Insan menawarkan program studi Sistem Informasi, Informatika, Rekayasa Perangkat Lunak, Hukum, Manajemen, Akuntansi, Perikanan, Agroteknologi, dan program studi Pascasarjana Magister Manajemen. Salah satu permasalahan Universitas Bina Insan adalah masih adanya mahasiswa yang cuti dengan alasan yang tidak jelas dalam masa kuliah. Mahasiswa aktif adalah mahasiswa yang terdaftar atau telah melakukan registrasi dan pengisisan KRS (Kartu Rencana Studi) setiap semester dan mengikuti perkuliahan, sedangkan mahasiswa tidak aktif atau biasa disebut dengan mahasiswa cuti adalah mahasiswa yang tidak melakukan registrasi dan pengisian KRS tepat waktu serta tidak mengikuti perkuliahan[3].

Registrasi akademik adalah kegiatan mendaftarkan diri menjadi mahasiswa peserta kegiatan akademik tiap semester berupa pengisian Kartu Rencana Studi. Beberapa faktor penyebab status yang tidak jelas tersebut diantaranya adalah faktor ekonomi, kemampuan akademik, dan lain-lain. Manajemen perguruan tinggi perlu mengidentifikasi serta melakukan tindakan terhadap mahasiswa yang mempunyai status "tidak diharapkan" untuk mengetahui faktor munculnya masalah tersebut perlu dilakukan evaluasi saat pertengahan masa studi mahasiswa guna mencegah sedini mungkin munculnya mahasiswa yang diindikasi terdapat status cuti untuk mengurangi dampak yang ditimbulkan. Saat ini, untuk evaluasi biasanya dilakukan dengan melakukan rapat evaluasi mengenai aktivitas kuliah mahasiswa melibatkan unsur pimpinan yaitu Rektor, Wakil Rektor I, Wakil Rektor II, Dekan, Kaprodi, kemudian Kabag Akademik untuk melihat ada berapa jumlah mahasiswa yang cuti, yang nanti kedepannya akan dilakukan penelusuran untuk meriset alasan mereka cuti itu apa, apakah bisa aktif lagi, lalu dipandu mulai aktif kuliah lagi, sehingga mengurangi jumlah mahasiswa cuti dari semester sebelumnya[4].

Pada penelitian terdahulu yang dilakukan oleh [5] Mei Lestari, tentang Hasil penelitian diperoleh nilai akurasi sebesar 70% serta nilai AUC sebesar 0.875 yang masuk kedalam klasifikasi baik. Pada Penelitian selanjutnya yang dilakukan oleh [6] Anna Baita dkk, Hasil penelitian menunjukkan bahwa algoritma SVM memiliki performa yang lebih baik jika dibandingkan dengan algoritma KNN. Akurasi algoritma SVM sebesar 70%, sedangkan akurasi algoritma KNN

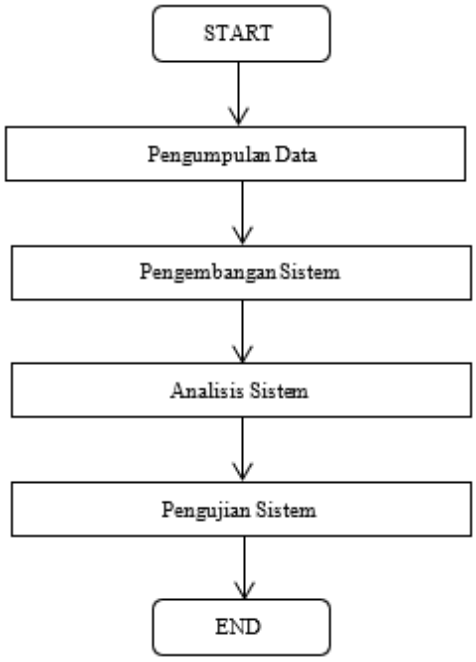
Copyright © 2024, Sekar Ayu Fitria. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

sebesar 56%. Pada Penelitian selanjutnya yang dilakukan oleh [7] Ari Putra Wibowo dkk, tentang hasil akurasi sebesar 70,46%, untuk algoritma Naïve Bayes dan akurasi sebesar 73,33% untuk algoritma K-Nearest Neighbor. Pada Penelitian selanjutnya yang dilakukan oleh [8] Rahmat Robi dkk, Klasifikasi citra kayu jati yang berasal dari Sulawesi dengan Metode K- Nearest Neighbor sebesar 76%, dan Metode Naive Bayes paling baik dengan persentase tingkat akurasi sebesar 82,7%. Hasil penelitian ini yaitu analisis menggunakan metode *K-Nearest Neighbor* (KNN) secara sistematis yang digunakan untuk melakukan klasifikasi untuk prediksi jumlah mahasiswa aktif dan cuti. Untuk itu penelitian ini dilakukan untuk menggali informasi tentang mahasiswa yang aktif dan cuti dengan penerapan data mining dengan metode *K-Nearest Neighbor*. Data yang diteliti berasal dari Universitas Bina Insan (UNIVBI) Kota Lubuklinggau. Hasil penelitian memungkinkan untuk mendapatkan aturan klasifikasi untuk data siswa yang aktif dan cuti dengan tingkat akurasi yang sangat baik. Oleh karena itu, penelitian ini dapat digunakan untuk mencegah mahasiswa putus kuliah secara sepihak. Tujuan penelitian ini untuk mengetahui jumlah mahasiswa yang masuk dalam kategori cuti dan aktif pada angkatan 2022.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Tahapan penelitian merupakan tahapan apa saja yang akan di lakukan pada penelitian[9]. Ada 4 tahap yang akan dilakukan dalam penelitian ini. Dapat dilihat pada gambar 1 Flowchart Tahapan Penelitian.



Gambar 1. Flowchart Tahapan Penelitian

Adapun tahapan penelitian tersebut sebagai berikut :

a. Pengumpulan Data

Pengumpulan data dengan cara melakukan pengamatan dan pencatatan langsung pada tempat penelitian atau pihak-pihak yang terkait dalam penelitian, Penulis melakukan observasi pengamatan langsung pada Toko Ratna Siahan, Serta pengamatan permasalahan yang berhubungan dengan pelayanan dan pengolahan data.

b. Pengembangan Sistem

Pengembangan Sistem adalah proses membuat system yang akan dibuat lalu dikembangkan agar dapat digunakan secara efektif dan berguna.

c. Analisis Sistem

Analisis Sistem Merupakan Proses untuk mengidentifikasi system yang telah dibuat apakah sudah sesuai dengan yang diharapkan.

d. Pengujian Sistem

Pada tahapan ini merupakan pengujian system yang akan diuji coba apakah sudah bisa digunakan dan dipakai oleh user atau pengguna.

2.2 Metode Analisa

Metode analisis yang digunakan penulis dalam penelitian ini adalah Metode Pengembangan Data Mining yang digunakan untuk menganalisis data dalam penerapannya menggunakan fase proses *Knowledge Discovery in Database* (KDD) dalam *database* meliputi data, transformasi data (*data transformation*), pemilihan data (*data selection*), penambangan data (*data mining*), pemodelan evolusi (*pattern evolution*) dan presentasi pengetahuan (*knowledge presentation*)[10].

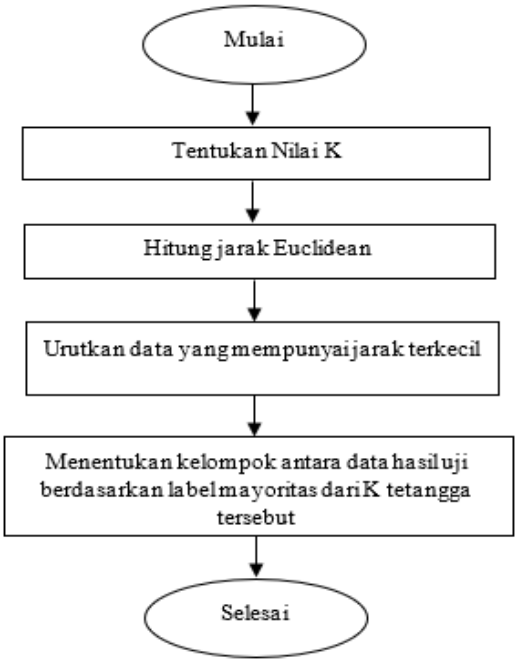
2.3 Metode *K-Nearest Neighbor (*KNN)

Dalam penelitian ini juga menggunakan metode Metode K-Nearest Neighbors (KNN) yaitu sebuah algoritma pembelajaran yang berdasarkan konsep bahwa data dengan atribut yang serupa cenderung berkumpul dalam ruang fitur yang sama. Untuk melakukan klasifikasi, KNN mencari k titik data terdekat dalam ruang fitur untuk data baru dan menentukan kategori mayoritas dari tetangga-tetangga tersebut sebagai prediksi. Proses ini bergantung pada perhitungan jarak antara data baru dan titik-titik data yang ada, yang umumnya menggunakan metrik seperti Euclidean distance atau Manhattan distance. Setelah tetangga terdekat teridentifikasi, data baru diklasifikasikan sesuai dengan mayoritas kategori dari tetangga-tetangga tersebut [11][12][13].

Keunggulan KNN termasuk kemudahan implementasi dan interpretasi serta kemampuannya untuk menangani dataset dengan kompleksitas yang tinggi dan tidak terstruktur. Namun, KNN juga memiliki beberapa kelemahan, seperti sensitivitas terhadap data pencilan dan kinerja yang kurang optimal pada dataset dengan dimensi tinggi karena kebutuhan akan perhitungan jarak yang kompleks. Oleh karena itu, pemilihan nilai k yang tepat dan pemrosesan data yang cermat diperlukan untuk memperoleh hasil yang optimal dengan metode KNN. Algoritma *K-Nearest Neighbor* memiliki kelebihan salah satunya ketangguhan terhadap data *training* yang memiliki banyak *noise* dan data dalam jumlah yang besar. Berikut ini alur penerapan metode *K-Nearest Neighbor* dalam klasifikasi untuk prediksi jumlah mahasiswa yang aktif dan cuti tahun angkatan 2020 Universitas Bina Insan.

a. Tetapkan parameter K
b. Hitung jarak antara data yang dievaluasi dengan seluruh pelatihan
c. Urutkan jarak yang terbentuk (urutan menaik)
d. Tentukan jarak terpendek ke barisan K
e. Kecocokan asosiasi kelas
f. Temukan jumlah kelas tetangga terdekat dan tetapkan lapisan tersebut sebagai lapisan data yang akan dievaluasi.

Dibawah ini merupakan alur dari metode *K-Nearest Neighbor* (KKN).



Gambar 2. Alur KNN

2.4 Pengujian Data

Penulis menggunakan metode pengujian *Library Confusion Matrix* yang termasuk dalam metode *K-Nearest Neighbor*[14]. Secara umum terdapat istilah-istilah dalam proses klasifikasi, khususnya sebagai berikut:

a. Akurasi merupakan ukuran yang menentukan tingkat kemiripan antara hasil pengukuran dengan nilai pengukuran sebenarnya. Berikut merupakan rumus akurasi:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

Keterangan:

TN = *True Negative*
TP = *True Positive*
FN = *False Negative*
FN = *False Negative*

b. Presisi merupakan sebuat tingkat kesamaan dari data yang diambil dengan informasi yang ada. Berikut merupakan rumus presisi:

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

c. *Recall* ialah tingkat keberhasilan sistem dalam mengambil informasi. Berikut merupakan rumus *recall*:

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

2.5 Pengolahan Data

Dalam memprediksi jumlah mahasiswa yang aktif dan cuti dengan menerapkan klasifikasi data mining menggunakan metode *K-Nearest Neighbor* dengan ini diperlukan data status mahasiswa yang aktif dan cuti tahun angkatan 2020[15]. Data yang digunakan yaitu sebanyak 624 data mahasiswa. Data tersebut meliputi nama mahasiswa, nomor induk mahasiswa (NIM), Indeks Prestasi Kumulatif (IPK), dan status mahasiswa yang aktif dan cuti sebagai labelnya[16]. Tabel 1 menunjukan beberapa data status mahasiswa yang dijadikan sampel penelitian.

Tabel 1. Data Status Mahasiswa

No	Nama	Nim	Program Studi	IPK	Label
1	Lokahita	2003010001	Agroteknologi	3.00	Aktif
2	Vito Janu Anggara	2001020021	Akuntansi	3.20	Aktif
3	Suci Nabilah Barokah	2004010022	Hukum	3.22	Cuti
4	Ari Hutama Batubara	2003020001	Ilmu Perikanan	3.32	Aktif
5	Reni Rahayu Lestari	2002020040	Informatika	3.10	Aktif
6	Muhammad Agung	2001010015	Manajemen	2.84	Cuti
…..	…..	….	….	…	…
…....	…..	….	…	…	…
621	…	…	…	…	…
622	Wendi Aprizal	2002010058	Rekayasa Sistem Komputer	1.80	Cuti
623	Agung Sanjaya	2002030015	Sistem Informasi	3.33	Aktif
624	Anjani	2002030030	Sistem Informasi	3.20	Aktif

Tabel 2 dibawah ini menunjukan konversi data Program Studi kedalm bentuk angka

Tabel 2. Konversi Data Program Studi

Program Studi	Konversi Angka
S1 Agroteknologi	1
S1 Akuntansi	2
S1 Hukum	3
S1 Ilmu Perikanan	4
S1 Informatika	5
S1 Manajemen	6
S1 Rekayasa Sistem Komputer	7
S1 Sistem Informasi	8
S2 Manajemen	9

Tabel 3 menunjukan data konversi IPK kedalam angka

Tabel 3. Konversi Data IPK

IPK	Konversi Angka
1.00 - 1.99	1
2.00 - 2.99	2
3.00 - 4.00	3

3. HASIL DAN PEMBAHASAN

Data yang digunakan yaitu sebanyak 624 data mahasiswa. Data tersebut meliputi nama mahasiswa, nomor induk mahasiswa (NIM), Indeks Prestasi Kumulatif (IPK), dan status mahasiswa yang aktif dan cuti sebagai labelnya. Selanjutnay data tersebut di konversikan berdasarkan tabel 1, 2 dan 3 yang telah di jelaskan sebelumnya.

3.1 Data Set

Tabel 4 menunjukan data set Status Mahasiswa yang sudah di konversi kedalam angka.

Tabel 4. Data Set Status Mahasiswa Yang di Konversi

	Nama	Nim	Program Studi	IPK	Label
1	Lokahita	2003010001	1	3	Aktif
2	Vito Janu Anggara	2001020021	2	3	Aktif
3	Suci Nabilah Barokah	2004010022	3	3	Cuti
4	Ari Hutama Batubara	2003020001	4	3	Aktif
5	Reni Rahayu Lestari	2002020040	5	3	Aktif
6	Muhammad Agung	2001010015	6	2	Cuti
7	…..	….	….	…	…
8	…..	….	…	…	…
621	…	…	…	…	…
622	Wendi Aprizal	2002010058	7	1	Cuti
623	Agung Sanjaya	2002030015	8	3	Aktif
624	Anjani	2002030030	8	3	Aktif

Setelah data di konversi maka di terpaakn metode KKN dengan langkah-langkah berikut ini. Langkah Penyelesaian:

a. Mentukan Parameter K sebagai banyaknya jumlah
b. Tetangga terdekat dengan objek baru.
   K=3
c. Kita hitung jarak antara data baru dengan semua data training. Kita menggunakan *Euclidean Distance*.
   Hitung seperti berikut:

$$K\ 1 = \sqrt{(1-9)^2 + (3-3)^2} = \sqrt{(-8)^2 + (0)^2} = \sqrt{64+0} = \sqrt{64} = 8$$

$$K\ 2 = \sqrt{(2-9)^2 + (3-3)^2} = \sqrt{(-7)^2 + (0)^2} = \sqrt{49+0} = \sqrt{49} = 7$$

$$K\ 3 = \sqrt{(3-9)^2 + (3-3)^2} = \sqrt{(-6)^2 + (0)^2} = \sqrt{36+0} = \sqrt{36} = 6$$

$$K\ 4 = \sqrt{(4-9)^2 + (3-3)^2} = \sqrt{(-5)^2 + (0)^2} = \sqrt{25+0} = \sqrt{25} = 5$$

$$K\ 5 = \sqrt{(5-9)^2 + (3-3)^2} = \sqrt{(-4)^2 + (0)^2} = \sqrt{16+0} = \sqrt{16} = 4$$

$$K\ 6 = \sqrt{(6-9)^2 + (2-3)^2} = \sqrt{(-3)^2 + (-1)^2} = \sqrt{9+1} = 10 = 3.162$$

$$K\ 7 = \sqrt{(7-9)^2 + (1-3)^2} = \sqrt{(-2)^2 + (-2)^2} = \sqrt{4+4} = \sqrt{8} = 2.828$$

$$K\ 8 = \sqrt{(8-9)^2 + (3-3)^2} = \sqrt{(-1)^2 + (0)^2} = \sqrt{1+0} = \sqrt{1} = 1$$

$$K\ 9 = \sqrt{(8-9)^2 + (3-3)^2} = \sqrt{(-1)^2 + (0)^2} = \sqrt{1+0} = \sqrt{1} = 1$$

d. Langkah selajutnya adalah mengurutkan jarak dari data baru dengan data training dan menentukan tetangga terdekat berdasarkan jarak minimum K. Tabel 5 menunjukan nilai urtu dari mahasiswa berdasarkan jarak.

Tabel 5. Mengurutkan Nilai

No	Nama	Nim	Program Studi	IPK	Encludien Distance	Urutan Jarak	Label
1	Lokahita	2003010001	1	3	8	8	Aktif
2	Vito Janu Anggara	2001020021	2	3	7	7	Aktif
3	Suci Nabilah Barokah	2004010022	3	3	6	6	Cuti
4	Ari Hutama Batubara	2003020001	4	3	5	5	Aktif
5	Reni Rahayu Lestari	2002020040	5	3	4	4	Aktif
6	Muhammad Agung	2001010015	6	2	3.162	3	Cuti
7	…..	….	….	…	…	…..	….
8	…..	….	…	…	…	…..	….
621	…	…	…	…	…	…	…
622	Wendi Aprizal	2002010058	7	1	2.828	2	Cuti
623	Agung Sanjaya	2002030015	8	3	1	1	Aktif
624	Anjani	2002030030	8	3	1	1	Aktif

e. Selanjutnya menentuka kategori dari tetangga terdekat. Kategori **Ya** diambil jika nilai **k<=3**. Sehingga baris 6, 7, 8, 9 termasuk dalam kategori **Ya**. Tabel 6 menunjukan tabel kategori.

Tabel 6. Menentukan Kategori

Encludien Distance	Urutan Jarak	Apakah Termasuk 3-NN	Kategori Ya Untuk 3-NN
8	8	Tidak (k > 3)	-
7	7	Tidak (k > 3)	-
6	6	Tidak (k > 3)	-
5	5	Tidak (k > 3)	-
4	4	Tidak (k > 3)	-
3.162	3	Ya (k = 3)	Cuti
2.828	2	Ya (k < 3)	Cuti
1	1	Ya (k < 3)	Aktif
1	1	Ya (k < 3)	Aktif

f. Dalam menentukan nilai prediksi, digunakan kategori mayoritas yang sederhana dari tetangga terdekat tersebut. Tabel 7 menunjkan nilai kategori tetanga terdekat

Tabel 7. Kategori Mayoritas Tetangga Terdekat

No	Nama	Nim	Program Studi	IPK	Label
1	Lokahita	2003010001	1	3	Aktif
2	Vito Janu Anggara	2001020021	2	3	Aktif
3	Suci Nabilah Barokah	2004010022	3	3	Cuti
4	Ari Hutama Batubara	2003020001	4	3	Aktif
5	Reni Rahayu Lestari	2002020040	5	3	Aktif
6	Muhammad Agung	2001010015	6	2	Cuti
7	…..	….	….	…	…
8	…..	….	…	…	…
621	…	…	…	…	…
622	Wendi Aprizal	2002010058	7	1	Cuti
623	Agung Sanjaya	2002030015	8	3	Aktif
624	Anjani	2002030030	8	3	Aktif

Data yang dimiliki pada baris 1 sampai 10 kategori yaitu Aktif dan Cuti. Dari jumlah mayoritas (Aktif > Cuti) tersebut maka didapatkan hasil nilai prediksi berupa kategori pada data baru adalah termasuk kategori **Aktif**.

3.2 Pemisahan atribut dan label

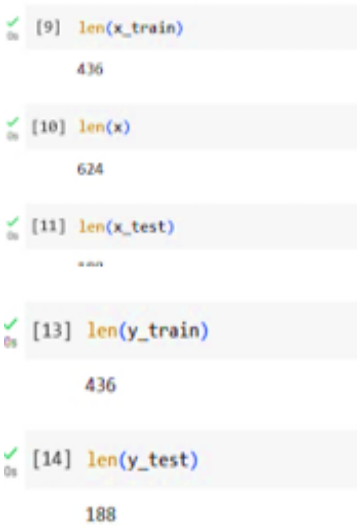
Dari data set yang ada maka dilakukan pemisan atribut sel. Pada langkah ini, akan dilakukan pemisahan atribut dan label yang akan digunakan[17]. Pada gambar 3 dibawah ini menunjukan proses pemisahan atribut dna label:

```
[1]  import numpy as np
     import matplotlib.pyplot as plt
     import pandas as pd

[4]  x = dataset.iloc[:, [2, 3]].values
     y = dataset.iloc[:, -1].values
```

Gambar 3. Pemisahan atribut dan label

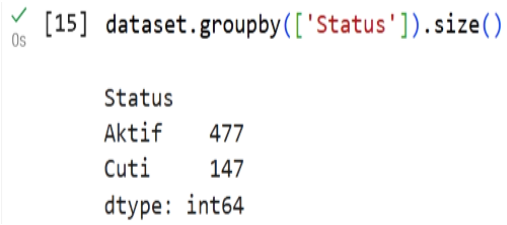
Setelah melakukan seleksi data, maka data siap dibagi menjadi data training dan data testing[18]. Pengujian ini dilakukan menggunakan split data dengan data training sebesar 70% dan data testing 30%. Dari hasil split data didapatlah data training sebanyak 436 data dan data testing sebanyak 188[19]. Pada gambar 4 dibawah ini :



Gambar 4. Data training dan data testing

3.3 Jumlah Data Per Kelas

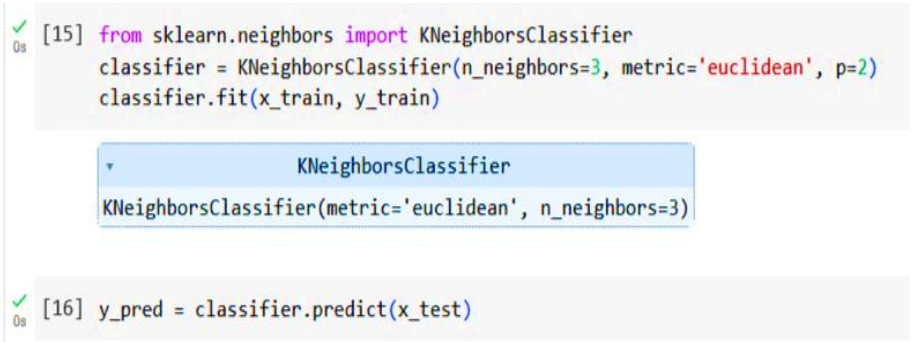
Selanjutnya, menemukan jumlah data untuk setiap kelas. Didapatlah data kelas Aktif sebanyak 477 mahasiswa dan kelas Cuti sebanyak 147 mahasiswa[20]. Pada gambar 5 dibawah ini:



Gambar 5. Jumlah data per kelas

3.4 Klasifikasi *K-Nearest Neighbor*

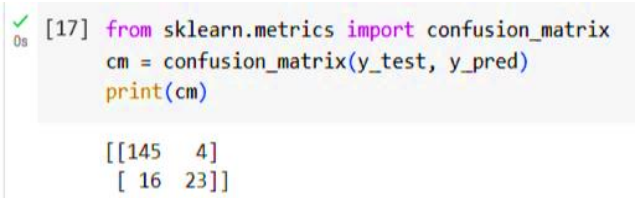
Penelitian ini menggunakan metode klasifikasi *K-Nearest Neighbor* dengan nilai *neighbor*=3 menggunakan *metric Euclidean* dan parameter 2. Pada gambar 6 dibawah ini :



Gambar 7. Klasifikasi *K-Nearest Neighbor*

3.5 Pengujian hasil Analisa

Untuk mendapatkan akurasi, presisi, dan *recall* dilakukanlah pengujian hasil pada penelitian ini menggunakan *Confusion Matrix*. Pada gambar 8 dibawah ini:



Gambar 8. *Confusion Matrix*

Pada gambar diatas dapat dijelaskan bahwa ada 145 data yang diprediksi benar dan kemudian hasilnya benar. Sedangkan, ada 23 data yang diprediksi menghasilkan salah dan ternyata adalah salah. Kemudian 16 dan 4 adalah prediksi yang kurang tepat.

```
[18] from sklearn.metrics import classification_report
     akurasi = classification_report(y_test,y_pred)
     print(akurasi)
```

	precision	recall	f1-score	support
Aktif	0.90	0.97	0.94	149
Cuti	0.85	0.59	0.70	39
accuracy			0.89	188
macro avg	0.88	0.78	0.82	188
weighted avg	0.89	0.89	0.89	188

```
[19] from sklearn.metrics import accuracy_score
     akurasi = accuracy_score(y_test,y_pred)
     print("Tingkat Akurasi :%d persen"%(akurasi*100))

     Tingkat Akurasi :89 persen
```

Gambar 9. Tingkat akurasi

Pada gambar 9 diatas, tingkat akurasi setelah dilakukan pengujian menggunakan *confusion matrix* adalah dengan nilai presisi Aktif sebesar 90% dan Cuti sebesar 85%, nilai *recall* Aktif sebesar 97% dan Cuti sebesar 59%, dan nilai akurasi sebesar 89%.

4. KESIMPULAN

Berdasarkan dari hasil penelitian dan pembahasan yang telah dilakukan, hasil klasifikasi status mahasiswa di Universitas Bina Insan pada Angkatan 2020 berdasarkan data sebanyak 624 data dengan menggunakan metode *K-Nearest Neighbor* terdapat 477 mahasiswa yang aktif dan 147 mahasiswa yang cuti dengan nilai presisi Aktif sebesar 90% dan Cuti sebesar 85%, nilai *recall* Aktif sebesar 97% dan Cuti sebesar 59%, dan nilai akurasi sebesar 89% maka dari itu metode *K-Nearest Neighbor* efektif digunakan untuk klasifikasi dengan jumlah data yang besar. Untuk mendapatkan aturan klasifikasi untuk data siswa yang aktif dan cuti dengan tingkat akurasi yang sangat baik. Oleh karena itu, penelitian ini dapat digunakan untuk mencegah mahasiswa putus kuliah secara sepihak.

REFERENCES

[1] S. Diansyah, "Klasifikasi Tingkat Kepuasan Pengguna dengan Menggunakan Metode K-Nearest Neighbour (KNN)," *J. Sistim Inf. dan Teknol.*, vol. 4, pp. 7–12, 2022, doi: 10.37034/jsisfotek.v4i1.114.

[2] H. Harianto and D. Rosiyadi, "Komparasi Algortima C4.5, Naïve Bayes dan k-Nearest Neighbor Sebagai Sistem Pendukung Keputusan Menaikkan Jumlah Peserta Didik," *J. Inform.*, vol. 7, no. 1, pp. 55–61, 2020, doi: 10.31311/ji.v7i1.7250.

[3] A. Salam, F. B. Nugroho, and J. Zeniarja, "Implementasi Algoritma K-Nearest Neighbor Berbasis Forward Selection Untuk Prediksi Mahasiswa Non Aktif Universitas Dian Nuswantoro Semarang," *JOINS (Journal Inf. Syst.*, vol. 5, no. 1, pp. 69–76, 2020, doi: 10.33633/joins.v5i1.3351.

[4] A. Damuri, U. Riyanto, H. Rusdianto, and M. Aminudin, "Implementasi Data Mining dengan Algoritma Naïve Bayes Untuk Klasifikasi Kelayakan Penerima Bantuan Sembako," *JURIKOM (Jurnal Ris. Komputer)*, vol. 8, no. 6, p. 219, 2021, doi: 10.30865/jurikom.v8i6.3655.

[5] Heliyanti Susana, "Penerapan Model Klasifikasi Metode Naive Bayes Terhadap Penggunaan Akses Internet," *J. Ris. Sist. Inf. dan Teknol. Inf.*, vol. 4, no. 1, pp. 1–8, 2022, doi: 10.52005/jursistekni.v4i1.96.

[6] D. Irawan and Z. Novianto, "Perancangan E-Learning Pada Sman 1 Kota Lubuklinggau Menggunakan Framework Codeigniter (Ci)," *J. Digit. Teknol. Inf.*, vol. 3, no. 2, p. 53, 2020, doi: 10.32502/digital.v3i2.2690.

[7] R. Shinta, P. Fairul, and G. Saputri, "Penerapan Data Mining Untuk Prediksi Penjualan Cat Menggunakan Metode Naïve (Studi Kasus : Mitra 10 Gading Serpong)," *Log. J. Ilmu Komput. dan …*, vol. 1, no. 3, pp. 571–578, 2023.

[8] M. Mega and J. Jasmir, "Prediksi Masa Studi Mahasiswa Unama Jambi Menggunakan Metode Algoritma C4.5," *J. Manaj. Sist. Inf.*, vol. 8, no. 1, pp. 140–151, 2023, doi: 10.33998/jurnalmsi.2023.8.1.770.

[9] A. Baita and N. Cahyono, "Analisis Sentimen Mengenai Vaksin Sinovac Menggunakan Algoritma Support Vector Machine (Svm) Dan K-Nearest Neighbor (Knn)," *Infos*, vol. 4, no. 2, pp. 42–42, 2021.

[10] J. Astri, J. Karman, and N. K. Daulay, "Prediksi Kelulusan Mahasiswa Menggunakan Metode K-Nearest Neigbor (KNN) pada

Fakultas Ilmu Teknik, Univeritas Bina Insan," *J. Ris. Sist. Inf. Dan Tek. Inform.*, vol. 8, pp. 169–173, 2022.
[11] A. P. Wibowo, W. Darmawan, and N. Amalia, "Komparasi Metode Naïve Bayes Dan K-Nearest Neighbor Terhadap Analisis Sentimen Pengguna Aplikasi Pedulilindungi," *IC-Tech*, vol. 17, no. 1, pp. 18–23, 2022, doi: 10.47775/ictech.v17i1.234.
[12] E. Irawan and I. Gunawan, "Penerapan C4.5 pada Keaktifan Mahasiswa dalam Pengumpulan Berkas di Biro Akademik," *REMIK (Riset dan E-Jurnal Manaj. Inform. Komputer)*, vol. 3, no. 2, p. 87, 2019, doi: 10.33395/remik.v3i2.10132.
[13] D. Cahyanti, A. Rahmayani, and S. A. Husniar, "Analisis performa metode Knn pada Dataset pasien pengidap Kanker Payudara," *Indones. J. Data Sci.*, vol. 1, no. 2, pp. 39–43, 2020, doi: 10.33096/ijodas.v1i2.13.
[14] R. A. Soleha, "Klasifikasi Kemampuan Ekonomi Calon Siswa Baru dengan Metode K-Nearest Neighbor pada SMA Negeri 1 Musi Rawas," *Jurasik (Jurnal Ris. Sist. Inf. dan Tek. Inform.*, vol. 8, no. 1, pp. 62–69, 2023.
[15] Q. A. A'yuniyah and M. Reza, "Penerapan Algoritma K-Nearest Neighbor Untuk Klasifikasi Jurusan Siswa Di Sma Negeri 15 Pekanbaru," *Indones. J. Inform. Res. Softw. Eng.*, vol. 3, no. 1, pp. 39–45, 2023, doi: 10.57152/ijirse.v3i1.484.
[16] M. M. Baharuddin, H. Azis, and T. Hasanuddin, "Analisis Performa Metode K-Nearest Neighbor Untuk Identifikasi Jenis Kaca," *Ilk. J. Ilm.*, vol. 11, no. 3, pp. 269–274, 2019, doi: 10.33096/ilkom.v11i3.489.269-274.
[17] S. K. P. Loka and A. Marsal, "Perbandingan Algoritma K-Nearest Neighbor dan Naïve Bayes Classifier untuk Klasifikasi Status Gizi Pada Balita," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 3, no. 1, pp. 8–14, 2023, doi: 10.57152/malcom.v3i1.474.
[18] A. A. D. Halim and S. Anraeni, "Analisis Klasifikasi Dataset Citra Penyakit Pneumonia menggunakan Metode K-Nearest Neighbor (KNN)," *Indones. J. Data Sci.*, vol. 2, no. 1, pp. 01–12, 2021, doi: 10.33096/ijodas.v2i1.23.
[19] A. M. Argina, "Penerapan Metode Klasifikasi K-Nearest Neigbor pada Dataset Penderita Penyakit Diabetes," *Indones. J. Data Sci.*, vol. 1, no. 2, pp. 29–33, 2020, doi: 10.33096/ijodas.v1i2.11.
[20] Z. Zulfikar, E. S. Podungge, M. I. Saleh, and ..., "Penerapan Data Mining Untuk Memprediksi Tingkat Kelulusan Siswa Menggunakan Algoritma Neural Network," *J. Elektron. Sist. ...*, vol. 5, no. 1, pp. 7–13, 2022.