

# Malik, Hasan - cek.01

*by* Luqman Abdul Rahman Malik, Firman Noor Hasan

---

**Submission date:** 21-Jun-2023 08:05AM (UTC+0700)

**Submission ID:** 2119990975

**File name:** Malik,\_Hasan\_-\_cek.01.docx (2.06M)

**Word count:** 4192

**Character count:** 27162

## FAKTOR-FAKTOR YANG MEMPENGARUHI MINAT CALON MAHASISWA BARU MENDAFTAR PADA FTII UHAMKA MENGGUNAKAN ALGORITMA K-NEAREST NEIGHBOR (K-NN)

### Factors That Influence The Interest of Prospective New Students To Apply To FTII Uhamka Using K-NN Algorithm

Luqman Abdur Rahman Malik<sup>1</sup>, Firman Noor Hasan<sup>2</sup>

<sup>1,2</sup>Teknik Informatika, Universitas Muhammadiyah Prof. Dr. Hamka, Jakarta, Indonesia  
Correspondence email: [firman.noorhasan@uhamka.ac.id](mailto:firman.noorhasan@uhamka.ac.id)

#### ABSTRACT

In accepting new students at Prof. University. Dr. Hamka, many prospective students or parents of students are looking for registration information, this is a great opportunity for Uhamka to gain the sympathy of prospective students to register at Uhamka, especially the Faculty of Industrial and Informatics Technology. The problem in this study is that there is no data processing related to the factors that influence the interest of prospective new students to choose the Faculty of Industrial and Informatics Technology (FTII) Uhamka. The purpose of this study was to determine the factors that influence the interest of prospective new students in choosing majors at the Faculty of Industrial and Information Technology (FTII) Uhamka. The attributes used in this study were 10 attributes, namely full name, major, tuition fee, FTII location with domicile, presence of friends/family, accreditation, facilities, PMB services, PMB information, and information of interest. The method that researchers use in this study is the K-Nearest Neighbor Algorithm (K-NN). From the results of testing the researchers used the K-5 fold technique and the confusion matrix obtained an average accuracy of 72.5%, which means it is good.

**Keywords:** Student, Uhamka, FTII, K-Nearest Neighbor (K-NN).

#### ABSTRAK

Dalam penerimaan mahasiswa baru di Universitas Prof. Dr. Hamka, banyak calon mahasiswa ataupun orang tua mahasiswa/i yang mencari informasi pendaftaran, hal ini merupakan peluang besar bagi Uhamka dalam mengambil simpati calon mahasiswa/i untuk mendaftar di Uhamka khususnya Fakultas Teknologi Industri dan Informatika. Permasalahan yang hadir pada penelitian ini adalah ketiadaan pengolahan data terkait faktor-faktor yang mempengaruhi minat calon mahasiswa baru memutuskan daftar di Fakultas Teknologi Industri dan Informatika (FTII) Uhamka. Sasaran penelitian ini adalah untuk menyelidiki faktor-faktor yang mempengaruhi minat calon mahasiswa baru memilih jurusan pada Fakultas Teknologi Industri dan Informatika (FTII) Uhamka. Atribut yang diaplikasikan dalam penelitian ini terhitung 10 atribut berupa nama lengkap, jurusan, biaya kuliah, lokasi FTII dengan domisili, terdapat teman/ keluarga, akreditasi, fasilitas, pelayanan pmb, informasi pmb, dan keterangan minat. Metode yang peneliti terapkan adalah Algoritma K-Nearest Neighbor (K-NN). Pengujian yang menggunakan teknik K-5 fold dan confusion berdampak pada hasil rata-rata akurasi sebesar 72,5% yang artinya sudah baik.

**Kata Kunci:** Mahasiswa, Uhamka, FTII, K-Nearest Neighbor (K-NN).

#### PENDAHULUAN

Salah satu indikasi berkembang atau tidaknya suatu lembaga pendidikan perguruan tinggi adalah total perolehan pendaftaran mahasiswa baru dari tahun ke tahun. Terlebih lagi pada perguruan tinggi terutama swasta, perolehan mahasiswa setiap tahunnya menjadi

kunci utama dalam pertumbuhan dari perguruan tinggi tersebut (Nas, 2021). Dalam hal ini, mahasiswa menjadi komponen krusial bagi perguruan tinggi swasta untuk mendapatkan sumber pendapatan dalam menjalankan aktivitas pendidikan di perguruan tinggi. Semakin melonjaknya jumlah mahasiswa, maka akan semakin meningkatnya sumber pendapatan bagi perguruan tinggi, sehingga perguruan tinggi semakin lancar kegiatannya tanpa keterbatasan dana (Noviyanto & Fauzi, 2022).



Akan tetapi, kompetisi dalam penerimaan mahasiswa baru di perguruan tinggi negeri maupun swasta semakin meningkat setiap tahunnya sehingga setiap institusi pendidikan dituntut untuk selalu profesional (Tensao et al., 2022). Beberapa lembaga pendidikan mengeluarkan biaya besar untuk promosi dengan menggunakan berbagai program, seperti penawaran biaya perkuliahan yang terjangkau, program sertifikasi internasional, kemudahan dalam memperoleh pekerjaan setelah lulus, dan sebagainya (Wulandari & Iqbal, 2021). Hal ini senada dengan hasil dari wawancara pada pihak sekretariat Fakultas Teknologi Industri dan Informatika Uhamka (selanjutnya disebut FTII Uhamka) yang mana telah melakukan beberapa strategi promosi serupa.

Dalam Penerimaan Mahasiswa Baru (PMB), banyak calon mahasiswa ataupun orang tua mahasiswa yang mencari informasi pendaftaran di perguruan tinggi yang hendak dituju (Widyasari & Fitriani, 2021). Penyampaian informasi yang tepat dan sesuai juga sangat mempengaruhi paradigma calon mahasiswa baru untuk melakukan registrasi di perguruan tinggi yang dipilih (Situmorang et al., 2022).

Sedangkan, dari penyebaran informasi tentang Uhamka didapat dari teman yang sedang kuliah atau alumni Uhamka. Tim promosi juga menyosialisasikan ke sekolah-sekolah yang ada di daerah Jakarta dan sekitarnya, diikuti dengan memasang iklan di sosial media. Selain informasi dari tim promosi dan sosial media, calon mahasiswa juga dapat mencari informasi melalui website Uhamka ataupun pada website FTII Uhamka.

Meskipun telah menyebarkan informasi dan beberapa program promosi, berdasarkan pengumpulan data melalui wawancara terhadap salah satu staff sekretariat FTII Uhamka, terjadi penurunan jumlah pendaftaran calon mahasiswa baru. Pada ajaran tahun 2020/2021, sebanyak 344 mahasiswa baru terdaftar di FTII Uhamka. Namun, pada tahun ajaran 2021/2022 tercatat sejumlah 328 mahasiswa baru yang terdaftar. Disini terjadi penurunan sebanyak 16 calon mahasiswa baru. Lalu pada tahun ajaran 2022/2023, mahasiswa baru yang terdaftar hanya 260. Ini berarti terjadi penurunan sebanyak 68 calon mahasiswa baru dari tahun ajaran kemarin. Dengan begitu, tercatat bahwa total jumlah penurunan calon mahasiswa baru dari tahun ajaran 2020/2021 hingga tahun ajaran 2022/2023 sebanyak 84 calon mahasiswa baru. Tentunya penurunan jumlah pendaftaran menjadi masalah yang serius yang harus dicari penyebab penurunan minat daftar calon mahasiswa baru (Noviyanto & Fauzi, 2022).

Permasalahan lain pada penelitian ini adalah belum adanya pengolahan data terkait faktor-faktor yang berdampak terhadap minat calon mahasiswa baru

mendaftar ke FTII Uhamka. Terlebih, mayoritas mahasiswa baru akan menelusuri informasi mengenai tempat kuliah yang diincarnya, dalam hal ini berupa lokasi perguruan tinggi, kegiatan mahasiswa, biaya kuliah, sarana dan prasarana, akreditasi (brand) serta program studi yang tersedia pada perguruan tinggi tersebut (Nasution, 2020). Diluar itu, pengaruh orang terdekat seperti orang tua, teman, saudara, beasiswa, peluang kerja dan lain-lain juga menyumbang dampak pada penyusunan keputusan dalam perekrutan calon mahasiswa semaksimal mungkin (Nas, 2021).

Dengan terkumpulnya informasi dan faktor tersebut, maka berpotensi untuk diolah agar mengetahui seberapa besar minat calon mahasiswa memilih FTII Uhamka. Kecenderungan dan antusiasme yang tinggi pada sesuatu dapat dimaknakan sebagai minat. Minat dapat timbul karena daya tarik dari luar dan juga datang dari hati sanubari (Prasetyo et al., 2019). Oleh karena itu, diperlukan tindakan berupa analisis dan pengolahan data menggunakan *data mining* untuk menelusuri faktor pendukung dari ketertarikan calon mahasiswa untuk registrasi. Penelitian ini memiliki target untuk menyelidiki faktor-faktor yang mempengaruhi minat calon mahasiswa baru mendaftar di FTII Uhamka menggunakan algoritma *K-Nearest Neighbor* (K-NN) (Fahlevi, 2019).

Selanjutnya, terdapat algoritma yang memiliki sebuah metode dengan mengklasifikasikan objek yang dipengaruhi data pembelajaran berjarak terdekat terhadap objek tersebut sehingga dikenal sebagai algoritma *K-Nearest Neighbor* (selanjutnya disebut K-NN). Selain itu, K-NN memiliki kesederhanaan, berjalan dengan berorientasi pada jarak terpendek dari sampel uji (*testing sample*) ke sampel latihan (*training sample*) untuk menentukan nilai daripada *K-Nearest Neighbor* nya. Apabila telah terkumpul nilai *K-Nearest Neighbor*, maka akan dilanjutkan dengan dipilih mayoritas dari *K-Nearest Neighbor* (K-NN) untuk ditetapkan sebagai prediksi dari sampel uji. Ditambah lagi, K-NN berkemampuan kuat terkait *training data* yang *noise* dan praktis terhadap data latihan berjumlah besar (Mubassiran, 2020). Dalam siklus *training*, K-NN berfungsi dalam penyimpanan vektor-vektor fitur dan penggolongan data *training sample*. Lalu, ditahapan klasifikasi dilakukan *testing data* berdasarkan perhitungan terhadap kesamaan fitur-fitur atau yang klasifikasinya belum diketahui (Hasanal et al., 2022). Tindakan selanjutnya yaitu menghitung jarak dari vektor baru yang ini kepada kesemua vektor *training sample* dilanjut dengan pengambilan terdekat dari sejumlah k buah. Titik baru diprediksi akan tergolong dalam klasifikasi yang paling dominan di antara titik-titik tersebut. Ditambah lagi, keberadaan dari fitur-fitur yang tidak terkait atau bobot

fitur berbeda dengan relevansinya akan mempengaruhi keakurasian (Bode, 2017).

Penelitian sebelumnya yang memprediksi program studi bagi calon mahasiswa baru menggunakan algoritma K-NN menghasilkan akurasi tinggi sebesar 74% dari total 379 data (Rahmawati Rahayu & Lidiawati, 2021).

Penggunaan algoritma K-NN juga menyimpulkan bahwa dapat memprediksi pemilihan jurusan untuk calon siswa baru. Hasil akurasi yang didapat sangat tinggi, yaitu 92,86% (Saepudin et al., 2019).

Algoritma K-NN juga dipakai dalam penelitian terkait prediksi waktu kelulusan siswa. Kesimpulan penelitian ini, K-NN menghasilkan akurasi sebesar 98,5% dari total 60 data (Muliono et al., 2020).

Penelitian terdahulu yang menggunakan K-NN untuk mengklasifikasi seleksi penerimaan beasiswa menunjukkan hasil akurasi tinggi sebesar 90,5% dari total 89 data (Cholil et al., 2021).

Penelitian terkait prediksi yang menggunakan K-NN juga menyimpulkan bahwa K-NN dapat memprediksi jumlah mahasiswa yang lulus untuk jurusan teknik informatika sebanyak 148 mahasiswa, dan jumlah yang tidak lulus sebanyak 67 mahasiswa (Manullang & Sianturi, 2021).

## METODE PENELITIAN

Peneliti menerapkan beberapa tindakan atau metode yang diperlukan, yaitu yang pertama pengumpulan data atau informasi (Hasan & Sudaryana, 2022). Langkah kedua berupa pengolahan data menggunakan algoritma K-NN dilanjut dengan menerapkan *confusion matrix* sebagai evaluasi dan validasi.

### Metode Pengumpulan Data

Merujuk pada (Sulaiman et al., 2022) dan (Muhammad Sopandi, Agus Efendi, 2017), dilakukan akuisisi data melalui:

1. Observasi  
Peneliti melakukan observasi secara langsung pada Fakultas Teknologi Industri dan Informatika (FTII) Uhamka.
2. Kuesioner  
Peneliti melakukan penyebaran kuesioner secara *online* kepada mahasiswa Angkatan 2022 dan calon mahasiswa yang mau mendaftar pada Fakultas Teknologi Industri dan Informatika (FTII) Uhamka.
3. Wawancara  
Peneliti melakukan tanya jawab secara tatap muka terhadap salah satu staff sekretariat FTII Uhamka dan staff promosi.
4. Studi Pustaka

Peneliti melakukan pengkajian terhadap literatur atau studi pustaka yang bernada sama dengan pokok bahasan beserta teknik penelitian yang digunakan.

### Metode Penelitian yang Dipakai

Peneliti menggunakan K-NN untuk memahami seberapa besar akurasi dari faktor-faktor yang berimbang pada minat calon mahasiswa baru melakukan pendaftaran di FTII Uhamka (Saputra & Hasan, 2023).

K-NN memiliki prinsip kerja dengan mencari jarak terkecil antara data yang sedang dievaluasi dengan ketangganya di dalam data pelatihan (Cholil et al., 2021). *Euclidean distance* dan *manhattan distance* (yang paling lazim dipakai *euclidean distance*) merupakan sekian dari banyaknya alternatif dalam mencari jarak kedekatan antara data baru dengan data lama. *Euclidean distance* memiliki rumus sebagai berikut (Kurniawan & Saputra, 2019):

$$X = \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2 + \dots + (a_n - b_n)^2} \dots (1)$$





**HASIL DAN PEMBAHASAN**

**Pengolahan Data**

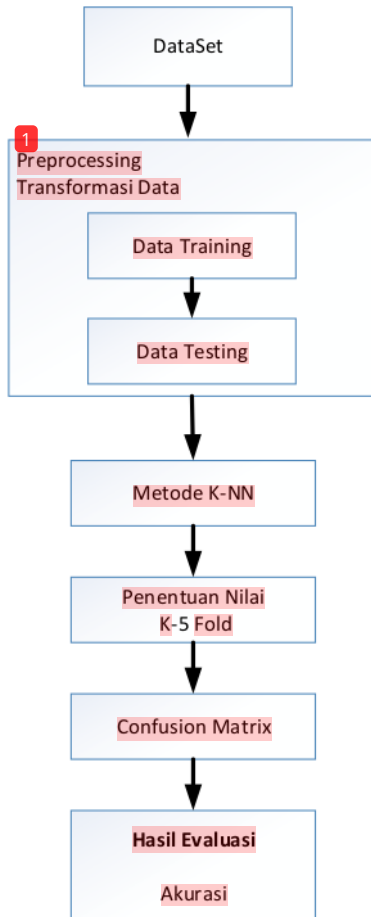
Untuk bagian ini, diawali dengan persiapan data awal, pemilihan atribut yang akan dikaji, dilanjutkan dengan validasi variabel yang tersedia dan persiapan transformasi data. Tahap persiapan data ini menghasilkan beberapa hal sebagai berikut:

Dataset yang peneliti gunakan sebanyak 119 data. Kemudian dataset ditransformasikan menggunakan *split* data dengan perbandingan 0.8 (80%) data *training* sebanyak 95 data dan 0.2 (80%) data *testing* sebanyak 24 data (Ode Nurhayah Kadir et al., 2019).

Kemudian, sejumlah sebanyak 95 data diikuti oleh data *testing* sejumlah 24 data, dilakukan data *training*. Atribut yang digunakan terdiri dari 10 atribut yaitu dapat dilihat pada tabel 1 sebagai berikut:

Tabel 1. Nilai Dataset Diikuti Atributnya

No	Atribut	Nilai
1	Nama Lengkap	
2	Jurusan	- Teknik Instrumentasi - Logam - Multimedia - IPA - IPS - Teknik Instalasi tenaga listrik (TITL) - Broadcasting dan Perfilman - Akuntansi - TKJ - Teknik Mesin - Teknik Kimia - Kimia Industri - Teknik Kendaraan Ringan - Teknik Komputer
3	Biaya Kuliah	- Terjangkau - Mahal
4	Lokasi Ftit Dengan Domisili	- Murah - Dekat - Terjangkau - Jauh
5	Terdapat Teman/ Keluarga	- Banyak - Cukup - Tidak Ada
6	Akreditasi	- Sangat Baik - Baik - Cukup
7	Fasilitas	- Sangat Baik - Memadai - Cukup
8	Pelayanan Pmb	- Sangat Baik - Baik - Cukup
9	Informasi Pmb	- Sosial Media (Facebook, Instagram, Youtube, dsb) - Teman, Kerabat, Keluarga



Sumber: (Gumanti et al., 2022)

Gambar 1. Tahapan Penelitian Menggunakan K-Nearest Neighbors

Merujuk pada gambar 1, peneliti menggunakan dataset yang didapatkan dari hasil kuesioner sebanyak 119 data (Wibisono et al., 2022). Langkah selanjutnya, yaitu diperoleh *data training* dan *data testing* dari proses transformasi dataset, lalu merencanakan proses dataset dengan algoritma K-NN, dan tidak lupa menentukan nilai *k-fold*, sebagaimana penelitian ini menerapkan 5 kali proses pengujian. Terakhir, ditutup dengan *confusion matrix* yang berfungsi sebagai evaluasi dan validasi algoritma sehingga akan didapatkan nilai akurasi dari hasil penelitian. Rumus *confusion matrix*, yaitu (Gumanti et al., 2022):

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \dots \dots (2)$$

- Kunjungan Langsung ke Sekolah
- Guru(wali kelas, guru bk, dll)
- Brosur/Poster
- Ya
- Tidak

10 Keterangan Minat

**Pengelompokan Data**

Peneliti melakukan klasifikasi data untuk data *testing* sebanyak 24. Pengelompokan data ini berdasarkan 10 atribut yang telah dipilih sesuai dengan table 1.

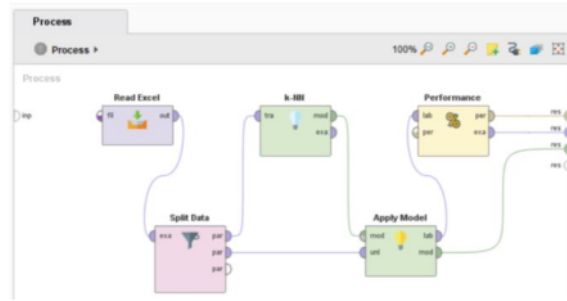
Berikut tabel 2 yang merupakan contoh sampel 4 data *testing* dari 24 data *testing*:

Tabel 2. Pengelompokan Data Testing

Nama Lengkap	Jurusan	Biaya Kuliahan	Lokasi FTTH Dengan Domisili	Terdaftar / Keluar	Akreditasi	Fasilitas	Penyamanan	Informasi
Anggia Rinjani	Teknik Instrumentasi Logam	Terjangkau	Jauh (>10 Km)	Cukup (1 - 3 Orang)	Baik	Mendatangi	Sangat Baik	Teman, Kerabat, Keluar
Abdillah Ismail Adha	RPL	Terjangkau	Jauh (>10 Km)	Tidak Ada	Baik	Mendatangi	Baik	Kunjungan Langsung ke Sekolah
Adi Tri Siswanto	TITL	Mahal	Jauh (>10 Km)	Tidak Ada	Baik	Mendatangi	Baik	Teman, Kerabat, Keluar
Adrian Maulana Saputra	IPA	Mahal	Jauh (>10 Km)	Tidak Ada	Baik	Cukup	Cukup	Teman, Kerabat, Keluar

**Evaluasi**

Berikutnya, Peneliti menerapkan *tools RapidMiner* sebagai evaluasi model. Desain model mengikuti gambar 2 sebagai berikut ini :



Sumber: (Saepudin et al., 2019)  
Gambar 2. Tahapan Penelitian Menggunakan K-Nearest Neighbors

Berdasarkan gambar 2, evaluasi model yang dilakukan, yaitu (Saepudin et al., 2019):

1. *Read Excel* : dataset *spreadsheet* dari *Microsoft Excel* akan terbaca dengan aksi ini.
2. *Split Data* : data latih sebanyak 0.8 dibarengin dengan data uji sebesar 0.2 akan dipisah secara acak dari kumpulan dataset.
3. *K-Nearest Neighbors* : merupakan teknik *data mining* yang diterapkan dalam penelitian.
4. *Apply Model* : berfungsi sebagai penerapan model yang telah dilatih sebelumnya dengan memakai data *training* pada *unlabeled data*(data *testing*).
5. *Performance* : aksi yang bertugas mengukur akurasi dari model atau metode K-NN.

**Ke-5 Fold dan Confusion Matrix Ditetapkan Sebagai Validasi Model**

Pengujian K-1  
Peneliti melakukan proses validasi model k-1 *fold* dengan menggunakan data *testing* sebanyak 24 data. Proses tersebut membuahkan hasil berupa gambar 3 berikut ini :

accuracy: 66.67%			
	true YA	true TIDAK	class precision
pred. YA	14	4	77.78%
pred. TIDAK	4	2	33.33%
class recall	77.78%	33.33%	

Gambar 3. Pengujian K-1 Fold

Matrix *Accuracy* menorehkan akurasi sebesar 66.67% yang didapat dari siklus evaluasi model dan proses validasi K-NN.

Langkah berikutnya yaitu menerapkan *confusion matrix* dengan rumus (2) yang digunakan sebagai validasi perhitungan akurasi dari gambar 3:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} = \frac{14+2}{14+2+4+4} = \frac{16}{24} = 0,6666 = 66,67\%$$

Dari hasil prediksi akurasi diatas dan gambar 3 menggunakan rapidminer dan *confusion matrix* yaitu :

- True Positive*(TP) yaitu kasus dimana faktor penyebab yang mendorong minat calon mahasiswa baru mendaftar pada FTII Uhamka adalah yang diprediksi (Positif) Ya minat, memang benar (True) Ya minat sebesar 14 orang.
- True Negatif*(TN) yaitu kasus dimana faktor penyebab yang mendorong minat calon mahasiswa baru mendaftar pada FTII Uhamka adalah yang diprediksi (Negatif) Tidak minat, sebenarnya calon mahasiswa tersebut memang (True) Tidak Minat sebesar 2 orang.
- False Positive*(FP) yaitu kasus dimana faktor penyebab yang mendorong minat calon mahasiswa baru mendaftar pada FTII Uhamka adalah yang diprediksi (Positif) Ya minat, Ternyata Tidak Minat. Prediksinya salah (False) sebesar 4 orang.
- False Negatif*(FN) yaitu kasus dimana faktor penyebab yang mendorong minat calon mahasiswa baru mendaftar pada FTII Uhamka adalah yang diprediksi (Negatif) Tidak Minat, tetapi ternyata sebenarnya (True) Ya Minat Sebesar 4 orang.

2. Pengujian K-2

Peneliti melakukan proses validasi model k-2 *fold* dengan menggunakan data testing sebanyak 24 data. Proses tersebut membuahkan hasil berupa gambar 4 berikut ini :

accuracy: 70.83%			
	true YA	true TIDAK	class precision
pred YA	15	4	78.95%
pred TIDAK	3	2	40.00%
class recall	83.33%	33.33%	

Gambar 4. Pengujian K-2 *Fold*

Matrix *Accuracy* menorehkan akurasi sebesar 70.83% yang didapat dari siklus evaluasi model dan proses validasi K-NN.

Langkah berikutnya yaitu menerapkan *confusion matrix* dengan rumus (2) yang digunakan sebagai validasi perhitungan akurasi dari gambar 4:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} = \frac{15+2}{15+2+3+4} = \frac{17}{24} = 0,7083 = 70,83\%$$

Dari hasil prediksi akurasi diatas dan gambar 4 menggunakan rapidminer dan *confusion matrix* yaitu :

- True Positive*(TP) yaitu kasus dimana faktor penyebab yang mendorong minat calon mahasiswa baru mendaftar pada FTII Uhamka adalah yang diprediksi (Positif) Ya minat, memang benar (True) Ya minat sebesar 15 orang.
  - True Negatif*(TN) yaitu kasus dimana faktor penyebab yang mendorong minat calon mahasiswa baru mendaftar pada FTII Uhamka adalah yang diprediksi (Negatif) Tidak minat, sebenarnya calon mahasiswa tersebut memang (True) Tidak Minat sebesar 2 orang.
  - False Positive*(FP) yaitu kasus dimana faktor penyebab yang mendorong minat calon mahasiswa baru mendaftar pada FTII Uhamka adalah yang diprediksi (Positif) Ya minat, Ternyata Tidak Minat. Prediksinya salah (False) sebesar 3 orang.
  - False Negatif*(FN) yaitu kasus dimana faktor penyebab yang mendorong minat calon mahasiswa baru mendaftar pada FTII Uhamka adalah yang diprediksi (Negatif) Tidak Minat, Tetapi ternyata sebenarnya (True) Ya Minat Sebesar 4 orang.
3. Pengujian K-3

Peneliti melakukan proses validasi model k-3 *fold* dengan menggunakan data testing sebanyak 24 data. Hasilnya ditunjukkan pada gambar 5 berikut ini:

accuracy: 75.00%			
	true YA	true TIDAK	class precision
pred YA	16	4	80.00%
pred TIDAK	2	2	50.00%
class recall	88.89%	33.33%	

Gambar 5. Pengujian K-3 *Fold*

Matrix *Accuracy* menorehkan akurasi sebesar 75.00% yang didapat dari siklus evaluasi model dan proses validasi K-NN.

Langkah berikutnya yaitu menerapkan *confusion matrix* dengan rumus (2) yang digunakan sebagai validasi perhitungan akurasi dari gambar 5:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} = \frac{16+1}{16+2+2+4} = \frac{18}{24} = 0,75 = 75\%$$

Dari hasil prediksi akurasi diatas dan gambar 5 menggunakan rapidminer dan *confusion matrix* yaitu :

- True Positive*(TP) yaitu kasus dimana faktor penyebab yang mendorong minat calon mahasiswa baru mendaftar pada FTII Uhamka adalah yang diprediksi (Positif) Ya minat, memang benar (True) Ya minat sebesar 16 orang.
- True Negatif*(TN) yaitu kasus dimana faktor penyebab yang mendorong minat calon mahasiswa baru mendaftar pada FTII Uhamka adalah yang

diprediksi (Negatif) Tidak minat, sebenarnya calon mahasiswa tersebut memang (True) Tidak Minat sebesar 2 orang.

- c. *False Positive*(FP) yaitu kasus dimana faktor penyebab yang mendorong minat calon mahasiswa baru mendaftar pada FTII Uhamka adalah yang diprediksi (Positif) Ya minat, Ternyata Tidak Minat. Prediksinya salah (False) sebesar 2 orang.
- d. *False Negatif*(FN) yaitu kasus dimana faktor penyebab yang mendorong minat calon mahasiswa baru mendaftar pada FTII Uhamka adalah yang diprediksi (Negatif) Tidak Minat, tetapi ternyata sebenarnya (True) Ya Minat Sebesar 4 orang.

4. Pengujian K-4

Peneliti melakukan proses validasi model k-4 *fold* dengan menggunakan data testing sebanyak 24 data. Hasilnya ditunjukkan pada gambar 5 berikut ini:

accuracy: 75.00%			
	true YA	true TIDAK	class precision
pred YA	16	4	80.00%
pred TIDAK	2	2	50.00%
class recall	88.89%	33.33%	

Gambar 6. Pengujian K-4 Fold

Matrix *Accuracy* menorehkan akurasi sebesar 75.00% yang didapat dari siklus evaluasi model dan proses validasi K-NN.

Langkah berikutnya yaitu menerapkan *confusion matrix* dengan rumus (2) yang digunakan sebagai validasi perhitungan akurasi dari gambar 6:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} = \frac{16+1}{16+2+2+4} = \frac{18}{24} = 0,75 = 75\%$$

Dari hasil prediksi akurasi diatas dan gambar 6 menggunakan rapidminer dan *confusion matrix* yaitu :

- a. *True Positive*(TP) yaitu kasus dimana faktor penyebab yang mendorong minat calon mahasiswa baru mendaftar pada FTII Uhamka adalah yang diprediksi (Positif) Ya minat, memang benar (True) Ya minat sebesar 16 orang.
- b. *True Negatif*(TN) yaitu kasus dimana faktor penyebab yang mendorong minat calon mahasiswa baru mendaftar pada FTII Uhamka adalah yang diprediksi (Negatif) Tidak minat, sebenarnya calon mahasiswa tersebut memang (True) Tidak Minat sebesar 2 orang.
- c. *False Positive*(FP) yaitu kasus dimana faktor penyebab yang mendorong minat calon mahasiswa baru mendaftar pada FTII Uhamka adalah yang diprediksi (Positif) Ya minat, Ternyata Tidak Minat. Prediksinya salah (False) sebesar 2 orang.

- d. *False Negatif*(FN) yaitu kasus dimana faktor penyebab yang mendorong minat calon mahasiswa baru mendaftar pada FTII Uhamka adalah yang diprediksi (Negatif) Tidak Minat, tetapi ternyata sebenarnya (True) Ya Minat Sebesar 4 orang.

5. Pengujian K-5

Peneliti melakukan proses validasi model k-5 *fold* 17 gan menggunakan data testing sebanyak 24 data. Dapat kita lihat pada gambar 7 berikut ini :

accuracy: 75.00%			
	true YA	true TIDAK	class precision
pred YA	17	5	77.27%
pred TIDAK	1	1	50.00%
class recall	94.44%	16.67%	

Gambar 7. Pengujian K-5 Fold

Matrix *Accuracy* menorehkan akurasi sebesar 75.00% yang didapat dari siklus evaluasi model dan proses validasi K-NN.

Langkah berikutnya yaitu menerapkan *confusion matrix* dengan rumus (2) yang digunakan sebagai validasi perhitungan akurasi dari gambar 7:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} = \frac{17+1}{17+1+1+5} = \frac{18}{24} = 0,75 = 75\%$$

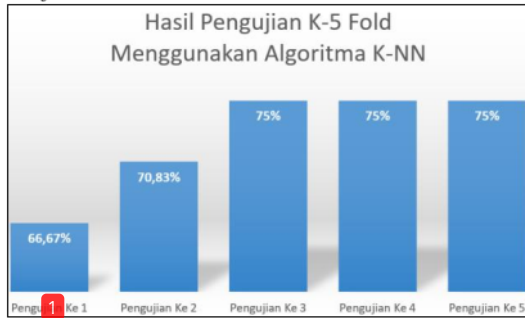
Dari hasil prediksi akurasi diatas dan gambar 7 menggunakan rapidminer dan *confusion matrix* yaitu :

- a. *True Positive*(TP) yaitu kasus dimana faktor penyebab yang mendorong minat calon mahasiswa baru mendaftar pada FTII Uhamka adalah yang diprediksi (Positif) Ya minat, memang benar (True) Ya minat sebesar 17 orang.
- b. *True Negatif*(TN) yaitu kasus dimana faktor penyebab yang mendorong minat calon mahasiswa baru mendaftar pada FTII Uhamka adalah yang diprediksi (Negatif) Tidak minat, sebenarnya calon mahasiswa tersebut memang (True) Tidak Minat sebesar 1 orang.
- c. *False Positive*(FP) yaitu kasus dimana faktor penyebab yang mendorong minat calon mahasiswa baru mendaftar pada FTII Uhamka adalah yang diprediksi (Positif) Ya minat, Ternyata Tidak Minat. Prediksinya salah (False) sebesar 1 orang.
- d. *False Negatif*(FN) yaitu kasus dimana faktor penyebab yang mendorong minat calon mahasiswa baru mendaftar pada FTII Uhamka adalah yang diprediksi (Negatif) Tidak Minat, tetapi ternyata sebenarnya (True) Ya Minat Sebesar 5 orang.





### Hasil Pengujian Menggunakan K-5 Fold dan Confusion Matrix



Gambar 8. Diagram Chart Pengujian Algoritma K-NN

Sesuai dengan rujukan dari gambar 8, dapat dipahami bahwa *diagram chart* menunjukkan setiap pengujian algoritma K-NN yang dimulai dari pengujian ke 1 dengan perolehan akurasi sebesar pengujian sebesar 66,67%, pengujian ke 2 membuahkan keakuratan sebesar 70,83%, pengujian ke 3 membuahkan keakuratan sebesar 75%, pengujian ke 4 membuahkan keakuratan sebesar 75%, dan pengujian terakhir dengan akurasi tinggi sebesar 75%. Dari hasil pengujian menggunakan teknik K-5 *fold* tersebut diperoleh hasil rata-rata akurasinya sebesar 72,5%.

### KESIMPULAN

Atribut yang digunakan dalam penelitian ini sebanyak 10 atribut yaitu nama lengkap, jurusan, biaya kuliah, lokasi fti dengan domisili, terdapat teman/keluarga, akreditasi, fasilitas, pelayanan pmb, informasi pmb, dan keterangan minat. Dataset yang peneliti gunakan sebanyak 119 data. Kemudian dataset di transformasikan menggunakan *split* data dengan perbandingan 0.8 (80%) data *training* sebanyak 95 data dan 0.2 (80%) data *testing* sebanyak 24 data. Dari hasil pengujian peneliti menggunakan teknik K-5 *fold* dan *confusion matrix* diperoleh hasil rata-rata akurasinya sebesar 72,5% yang artinya sudah baik.

### DAFTAR PUSTAKA

Bode, A. (2017). K-Nearest Neighbor Dengan Feature Selection Menggunakan Backward Elimination Untuk Prediksi Harga Komoditi Kopi Arabika. *ILKOM Jurnal Ilmiah*, 9(2), 188–195. <https://doi.org/10.33096/ilkom.v9i2.139.188-195>

Cholil, S. R., Handayani, T., Prathivi, R., & Ardianita, T. (2021). Implementasi Algoritma Klasifikasi K-

Nearest Neighbor (KNN) Untuk Klasifikasi Seleksi Penerima Beasiswa. *IJCIT: Indonesian Journal on Computer and Information Technology*, 6(2), 118–127. <https://doi.org/10.31294/ijcit.v6i2.10438>

Fahlevi, J. H. J. dan S. A. (2019). *Klasifikasi Calon Mahasiswa Bidikmisi dengan Algoritma K-Nearest Neighbor*. 5(1), 1–5.

Gumanti, A. E., Handayani, S., & Toresa, D. (2022). Penerapan Algoritma K-Nearest Neighbor Untuk Klasifikasi Topik Skripsi Mahasiswa di Fakultas Ilmu Komputer. *JITaCS: Journal of Information Technology and Computer Science*, 1(2), 44–50. <https://journal.unilak.ac.id/index.php/JITaCS/article/view/10037>

Hasan, F. N., & Sudaryana, I. K. (2022). Penerapan Business Intelligence & Online Analytical Processing untuk Data-Data Penelitian dan Luarannya pada Perguruan Tinggi Menggunakan Pentaho. *Infotech: Journal of Technology Information*, 8(2), 85–92. <https://doi.org/10.37365/jti.v8i2.143>

Hasanah, F., Suprpti, T., Rahaningsih, N., & Ali, I. (2022). Implementasi Algoritma K-Nearest Neighbor dalam Menentukan Buku Berdasarkan Peminatan. *Jurnal Accounting Information System (AIMS)*, 5(1), 102–111. <https://doi.org/10.32627/aims.v5i1.467>

Kurniawan, D., & Saputra, A. (2019). Penerapan K-Nearest Neighbour dalam Penerimaan Peserta Didik dengan Sistem Zonasi. *Jurnal Sistem Informasi Bisnis*, 9(2), 212. <https://doi.org/10.21456/vol9iss2pp212-219>

Manullang, R. A., & Sianturi, F. A. (2021). Penerapan Algoritma K-Nearest Neighbour Untuk Memprediksi Kelulusan Mahasiswa. *Jurnal Ilmu Komputer Dan Sistem Informasi (JIKOMSI)*, 4(2), 42–50. <https://doi.org/10.9767/jikoms.v4i2.155>

Mubassiran, I. C. (2020). PENERAPAN ALGORITMA K-NEAREST NEIGHBORS UNTUK MENENTUKAN POLA PENERIMAAN MAHASISWA BARU (Studi Kasus: Politeknik Pos Indonesia). *Ilmiah Manajemen Informatika*, 12(2), 1–70.

Muhammad Sopandi, Agus Efendi, B. (2017). Jurnal Ilmiah Pendidikan Teknik Kejuruan ( JIPTEK ). *Jurnal Ilmiah Pendidikan Teknik Kejuruan*, X(2), <https://jurnal.uns.ac.id/jptk>.

Muliono, R., Lubis, J. H., & Khairina, N. (2020). Analysis K-Nearest Neighbor Algorithm for Improving Prediction Student Graduation Time. *Sinkron: Jurnal Dan Penelitian Teknik Informatika*, 4(2), 42–46. <https://doi.org/10.33395/sinkron.v4i2.10480>

- Nas, C. (2021). Data Mining Prediksi Minat Calon Mahasiswa Memilih Perguruan Tinggi Menggunakan Algoritma C4.5. *Jurnal Manajemen Informatika (JAMIKA)*, 11(2), 131–145. <https://doi.org/10.34010/jamika.v11i2.5506>
- Nasution, T. (2020). Implementasi Algoritma K-Nearest Neighbor Untuk Penentuan Kelulusan Mahasiswa Tepat Waktu. *Jurnal Perangkat Lunak*, 2(1), 1–14. <https://doi.org/10.32520/jupel.v2i1.944>
- Noviyanto, H., & Fauzi, A. (2022). Implementasi Data Mining Untuk Mengetahui Minat Siswa Terhadap Perguruan Tinggi. *Jurnal Teknik Informatika Dan Teknologi Informasi*, 2(1), 51–55. <https://doi.org/10.55606/jutiti.v2i1.450>
- Ode Nurhayah Kadir, W., Pramono, B., & Statiswaty. (2019). Penerapan Data Mining Dengan Metode K-Nearest Neighbor(KNN) Untuk Mengelompokan Minat Konsumen Asuransi (Pt. Jasarharja Putera). *SemanTIK*, 5(1), 97–104. <https://doi.org/10.5281/zenodo.3116132>
- Prasetyo, A., Kursini, & Rudyanto Arief, M. (2019). Penerapan Algoritma K Nearest Neighbor Untuk Rekomendasi Minat Konsentrasi Di Program Studi Teknik Informatika Universitas Pgrri Yogyakarta. *Jurnal Informasi Interaktif*, 4(1), 1–6.
- Rahmawati Rahayu, R., & Lidiawati. (2021). Implementasi Algoritma K-Nearest Neighbor Untuk Memprediksi Program Studi Bagi Calon Mahasiswa Baru. *Jurnal Informatika Dan Teknologi*, 4(2), 131–141. <https://doi.org/10.29408/jit.v4i2.3546>
- Saepudin, S., Muslih, M., & Sihabudin. (2019). Pemilihan Jurusan dengan Metode K-Nearest Neighbor Untuk Calon Siswa Baru. *Jurnal Rekayasa Teknologi Nusa Putra*, 5(2), 15–19. <https://doi.org/10.52005/rekayasa.v5i2.89>
- Saputra, A., & Hasan, F. N. (2023). Analisis Sentimen Terhadap Aplikasi Coffee Meets Bagel Dengan Algoritma Naïve Bayes Classifier. *Sibatik Journal / Volume*, 2(2), 465–474.
- Situmorang, Z., Mandasari, S., Franciska, Y., Andriyani, K., & Ramadhan, P. S. (2022). Algoritma C45 Dalam Memprediksi Minat Calon Mahasiswa. *Journal of Science and Social Research*, 5(1), 125–133. <https://doi.org/10.54314/jssr.v5i1.809>
- Sulaiman, R., Raya, A. M., & Cahya Saputra, K. E. (2022). ADVICE PLANNING SYSTEM PADA DINAS PEKERJAAN UMUM DAN PENATAAN RUANG TATA KOTA PANGKALPINANG BERBASIS ANDROID. *Infotech: Journal of Technology Information*, 8(2), 69–76. <https://doi.org/10.37365/jti.v8i2.133>
- Tensao, O. O., Wijaya, N. Y. A., & Fredlina, K. Q. (2022). Analisa Data Mining dengan Algoritma K-Means Clustering Untuk Menentukan Strategi Promosi Mahasiswa Baru Pada STMIK Primakara. *Jurnal Informatika Dan Sistem Informasi*, 14(1). <https://doi.org/10.37424/informasi.v14i1.135>
- Wibowo, A., Noor Hasan, F., Nurhayati, R., & Wibowo, A. (2022). Analisis Sentimen Opini Masyarakat Terhadap Keefektifan Pembelajaran Daring Selama Pandemi COVID-19 Menggunakan Naïve Bayes Classifier. *Jurnal Asimetri: Jurnal Ilmiah Rekayasa Dan Inovasi*, 4(2), 239–248. <https://doi.org/10.35814/asimetri.v4i1.3577>
- Widyasari, V., & Fitriani, A. S. (2021). Penerapan Data Mining Dengan Metode Klasifikasi Pada Promosi Penerimaan Mahasiswa Baru Universitas Muhammadiyah Sidoarjo Menggunakan Algoritma Naïve Bayes Berbasis Web. *Procedia of Engineering and Life Science*, 1(2). <https://doi.org/10.21070/pels.v1i2.1062>
- Wulandari, S., & Iqbal, M. (2021). Analisis Kelayakan Lokasi Promosi Dalam Penerimaan Mahasiswa Baru (PMB) Dengan Algoritma Naïve Bayes & Decision Tree C4.5. *Kilat*, 10(1). <https://doi.org/10.33322/kilat.v10i1.1169>

#### PENTING UNTUK DIBACA

- ☑ Naskah dikirimkan melalui **E-Journal STMIK Widuri**, dengan alamat website <http://jurnal.kampuswiduri.ac.id/index.php/infotech> Jika anda belum memiliki akun diharuskan untuk registrasi terlebih dulu.
- ☑ Isi tulisan bukan merupakan tanggung jawab redaksi. Redaksi berhak mengedit redaksional tanpa mengubah arti.
- ☑ **Redaksi berhak menolak naskah yang tidak memenuhi syarat dan akan dikembalikan.**
- ☑ Hal-hal yang belum jelas dapat menghubungi Tim Redaksi Jurnal Infotech STMIK Widuri dengan alamat: **Jl. Palmerah Barat No. 353 Jakarta 12210. Telp: (021) 5330977**



## ORIGINALITY REPORT

---

14%

SIMILARITY INDEX

14%

INTERNET SOURCES

3%

PUBLICATIONS

2%

STUDENT PAPERS

---

## PRIMARY SOURCES

---

1	<a href="http://jurnal.kampuswiduri.ac.id">jurnal.kampuswiduri.ac.id</a> Internet Source	4%
2	<a href="http://id.123dok.com">id.123dok.com</a> Internet Source	2%
3	<a href="http://www.researchgate.net">www.researchgate.net</a> Internet Source	1%
4	<a href="http://ojs.unikom.ac.id">ojs.unikom.ac.id</a> Internet Source	1%
5	<a href="http://pdfs.semanticscholar.org">pdfs.semanticscholar.org</a> Internet Source	1%
6	<a href="http://repository.iti.ac.id">repository.iti.ac.id</a> Internet Source	1%
7	<a href="http://ejournal.sisfokomtek.org">ejournal.sisfokomtek.org</a> Internet Source	1%
8	<a href="http://methonomi.net">methonomi.net</a> Internet Source	<1%
9	<a href="http://media.neliti.com">media.neliti.com</a> Internet Source	<1%

---

10	<a href="https://repository.uinjambi.ac.id">repository.uinjambi.ac.id</a> Internet Source	<1 %
11	Submitted to Universitas Dian Nuswantoro Student Paper	<1 %
12	S Ulfah, S A Awalludin, Wahidin. "Advection-diffusion model for the simulation of air pollution distribution from a point source emission", Journal of Physics: Conference Series, 2018 Publication	<1 %
13	<a href="https://skripsi.tunasbangsa.ac.id">skripsi.tunasbangsa.ac.id</a> Internet Source	<1 %
14	<a href="https://core.ac.uk">core.ac.uk</a> Internet Source	<1 %
15	<a href="https://smart.stmikplk.ac.id">smart.stmikplk.ac.id</a> Internet Source	<1 %
16	<a href="https://stamina.ppj.unp.ac.id">stamina.ppj.unp.ac.id</a> Internet Source	<1 %
17	<a href="https://thohamuhammad.blogspot.com">thohamuhammad.blogspot.com</a> Internet Source	<1 %
18	Jeremy Andre Septian, Tresna Maulana Fachrudin, Aryo Nugroho. "Analisis Sentimen Pengguna Twitter Terhadap Polemik Persepakbolaan Indonesia Menggunakan Pembobotan TF-IDF dan K-Nearest Neighbor",	<1 %



# Journal of Intelligent System and Computation, 2019

Publication

---

19	<a href="http://ejournal.bsi.ac.id">ejournal.bsi.ac.id</a> Internet Source	<1 %
20	<a href="http://ejurnal.politeknikpratama.ac.id">ejurnal.politeknikpratama.ac.id</a> Internet Source	<1 %
21	<a href="http://ejurnal.stmik-budidarma.ac.id">ejurnal.stmik-budidarma.ac.id</a> Internet Source	<1 %
22	<a href="http://zombiedoc.com">zombiedoc.com</a> Internet Source	<1 %

---

Exclude quotes      On

Exclude matches      Off

Exclude bibliography      On