



PREDIKSI KELULUSAN MAHASISWA TEKNIK INFORMATIKA UNIVERSITAS KRISTEN WIRA WACANA SUMBA MENGUNAKAN ALGORITMA K-NEAREST NEIGHBOR

Prediction of Graduation of Information Technology Students at Wira Wacana Christian University Using the K-Nearest Neighbor Algorithm

Angelita Saylistiawati Radjah¹, Yustina Rada² dan Tri Sari Dewi Novyanti Bertha Mira³

^{1,2,3}Program Studi Teknik Informatika, Fakultas Sains Dan Teknologi, Universitas Kristen Wira Wacana Sumba, Indonesia

E-mail : angelitasaylistiawataradjah@gmail.com yustinarada@unkriswina.ac.id tri@unkriswina.ac.id

Jl.R.Suprato No.35,Prailiu,Kec.Kota Waingapu,Kabupaten Sumba Timur,Nusa Tenggara Timur

Corresponding author: angelitasaylistiawataradjah@gmail.com

ABSTRACT

The phenomenon of students failing to graduate on time is a challenge faced by many universities in Indonesia, including the Informatics Engineering Study Program at Wira Wacana Christian University of Sumba. This study aims to develop a student graduation prediction model to identify early those at risk of not graduating on time, allowing for targeted academic interventions. The method used is the K-Nearest Neighbors (KNN) algorithm with a parameter of $K=5$, utilizing academic data of students from the 2016–2020 cohorts, including total cumulative credits (SKS), GPA, semester GPA, semester, and graduation period. The data was processed through pre-processing stages consisting of data cleaning, labeling, normalization, and splitting into training and testing datasets with an 80:20 ratio. Model evaluation was conducted using accuracy, precision, recall, and f1-score metrics. The implementation results show that the KNN model achieved an accuracy of 81.18%, with the best performance in the On-Time Graduation class (precision 0.81; recall 0.90; f1-score 0.85) and performance that still needs improvement in the Not-On-Time Graduation class (precision 0.81; recall 0.68; f1-score 0.74). These findings indicate that KNN can be used as a reliable graduation prediction tool, while also providing opportunities for universities to carry out early interventions in efforts to reduce the rate of delayed graduations.

Keywords: Graduation Prediction, K-Nearest Neighbor, Academic Data, Informatics Engineering, Google Colab

ABSTRAK

Fenomena ketidaklulusan mahasiswa tepat waktu menjadi tantangan yang dihadapi banyak perguruan tinggi di Indonesia, termasuk Program Studi Teknik Informatika Universitas Kristen Wira Wacana Sumba. Penelitian ini bertujuan untuk membangun model prediksi kelulusan mahasiswa guna mengidentifikasi secara dini mahasiswa yang berisiko tidak lulus tepat waktu, sehingga dapat dilakukan intervensi akademik yang tepat sasaran. Metode yang digunakan adalah algoritma *K-Nearest Neighbors* (KNN) dengan parameter $K=5$, yang memanfaatkan data akademik mahasiswa angkatan 2016–2020, meliputi jumlah SKS kumulatif, IPK, IPS, semester, dan periode kelulusan. Data diproses melalui tahapan *pre-processing* berupa pembersihan data, pelabelan, normalisasi, serta pembagian data latih dan data uji dengan rasio 80:20. Evaluasi model dilakukan menggunakan metrik akurasi, *precision*, *recall*, dan *f1-score*. Hasil implementasi menunjukkan bahwa model KNN mampu mencapai akurasi sebesar 81,18%, dengan performa terbaik pada kelas Tepat Waktu (*precision* 0,81; *recall* 0,90; *f1-score* 0,85) dan performa yang masih perlu ditingkatkan pada kelas Tidak Tepat Waktu (*precision* 0,81; *recall* 0,68; *f1-score* 0,74). Temuan ini menunjukkan bahwa KNN dapat digunakan sebagai alat bantu prediksi kelulusan yang cukup andal, sekaligus memberikan peluang bagi perguruan tinggi untuk melakukan intervensi dini dalam upaya menekan angka keterlambatan kelulusan.

Kata kunci: Prediksi Kelulusan, K-Nearest Neighbor, Data Akademik, Teknik Informatika, Google Colab.

PENDAHULUAN

Pendidikan tinggi memainkan peran penting dalam mengembangkan sumber daya manusia yang terampil dan mudah beradaptasi, sehingga meningkatkan daya saing individu di pasar tenaga kerja yang semakin mengglobal. Universitas memainkan peran penting dalam memberikan pendidikan berkualitas yang membekali siswa dengan pengetahuan penting, keterampilan berpikir kritis, dan kompetensi praktis yang diperlukan untuk keberhasilan profesional dan kewarganegaraan yang bertanggung jawab dalam masyarakat global yang semakin kompleks. Namun, dalam prakteknya, banyak perguruan tinggi di Indonesia yang menghadapi tantangan dalam memastikan bahwa mahasiswa mereka dapat menyelesaikan studi tepat waktu.

Di Indonesia, meskipun jumlah mahasiswa yang diterima terus meningkat setiap tahunnya, tingkat kelulusan mahasiswa di berbagai perguruan tinggi masih jauh dari harapan. Berdasarkan data dari Direktorat Jenderal Pendidikan Tinggi (2024), pada tahun 2024 terdapat ketidakselarasan yang signifikan antara jumlah mahasiswa yang diterima dan mahasiswa yang lulus tepat waktu (Iddrus, 2024). Misalnya, pada tahun 2016, terdapat 113 mahasiswa yang diterima di Program Studi Teknik Informatika di Universitas Kristen Wira Wacana Sumba, namun hanya 67 mahasiswa yang lulus tepat waktu. Data ini menggambarkan adanya kesenjangan yang jelas antara jumlah mahasiswa yang diterima dan yang berhasil menyelesaikan studi mereka tepat waktu. Fenomena ini mencerminkan adanya permasalahan mendasar dalam sistem pendidikan yang membutuhkan perhatian lebih lanjut. Adapun data mahasiswa angkatan 2016 – 2021 yang dapat dilihat pada tabel berikut.

Tabel 1. Data Mahasiswa Angkatan 2016-2021

| Angkatan | Masuk | Lulus | Aktif | Non Aktif |
|----------|-------|-------|-------|-----------|
| 2016 | 113 | 67 | - | - |
| 2017 | 116 | 71 | - | - |
| 2018 | 143 | 87 | 9 | 6 |
| 2019 | 152 | 70 | 30 | 38 |
| 2020 | 165 | 80 | 38 | 40 |
| 2021 | 153 | - | 125 | 16 |

Data yang tersedia menunjukkan adanya kesenjangan signifikan antara jumlah mahasiswa yang diterima dan yang lulus tepat waktu di Program Studi Teknik Informatika. Sebagai contoh, pada tahun 2016, sebanyak 113 mahasiswa diterima, namun hanya 67 yang lulus tepat waktu. Pada tahun 2019, dari 152 mahasiswa yang diterima, hanya 70 mahasiswa yang berhasil lulus tepat waktu. Angka ketidaklulusan yang tinggi ini mengindikasikan bahwa banyak mahasiswa yang tidak dapat menyelesaikan studi mereka sesuai jadwal, meskipun jumlah penerimaan cukup besar. Hal ini menunjukkan bahwa ketidaklulusan tidak hanya dipengaruhi oleh faktor individu, tetapi juga mencerminkan adanya kekurangan dalam proses pendidikan yang perlu segera diperbaiki. Fenomena ini tidak hanya terjadi di Program Studi Teknik Informatika, tetapi juga menjadi masalah umum di banyak perguruan tinggi di Indonesia.

Menanggapi fenomena ketidaklulusan yang terus berlanjut ini, diperlukan upaya yang lebih efektif dan berbasis data untuk memprediksi kelulusan mahasiswa. Salah satu solusi yang dapat diterapkan adalah dengan memanfaatkan teknologi analisis data, khususnya prediksi kelulusan menggunakan metode data *mining*. Dengan menggunakan analisis data, Universitas dapat mengidentifikasi mahasiswa yang berisiko tidak lulus tepat waktu berdasarkan karakteristik akademik dan non-akademik. Teknologi penambangan data secara sistematis mengintegrasikan faktor akademis dan non-akademis, memungkinkan analisis komprehensif dan pengambilan keputusan yang tepat di berbagai bidang.

Menggunakan metode analisis data seperti *K-Nearest Neighbors* (KNN) memungkinkan peneliti untuk mengklasifikasikan dan memprediksi hasil secara efektif dengan memeriksa kedekatan titik data dalam ruang fitur. Melalui pengelompokan data akademik seperti Indeks Prestasi Kumulatif (IPK), nilai mata kuliah, kehadiran, serta data non-akademik seperti keterlibatan dalam organisasi dan aktivitas ekstrakurikuler, prodi dapat memetakan kelompok mahasiswa dengan potensi kelulusan yang rendah. Hasil dari prediksi ini akan memberikan dasar bagi pengambilan keputusan akademik untuk melakukan intervensi yang lebih awal, seperti pemberian bimbingan, pembinaan, atau penyesuaian kurikulum yang lebih sesuai dengan kebutuhan mahasiswa.

Dengan penerapan prediksi kelulusan ini, dapat mengurangi angka ketidakkelulusan mahasiswa, meningkatkan keberhasilan akademik, dan akhirnya meningkatkan kualitas pendidikan secara keseluruhan. Selain itu, pendekatan ini dapat mempercepat proses kelulusan mahasiswa, mengurangi tingkat putus sekolah, serta menyesuaikan strategi pembelajaran agar lebih efisien dan tepat sasaran.

MATERI DAN METODE

Prediksi Kelulusan

Prediksi kelulusan mahasiswa adalah metode untuk memperkirakan kemungkinan seorang mahasiswa dapat menyelesaikan program studi dalam waktu yang telah ditentukan berdasarkan data historis. prediksi kelulusan sebagai penggunaan data akademik mahasiswa, seperti nilai mata kuliah dan absensi, untuk memodelkan kemungkinan keberhasilan akademik mahasiswa. Prediksi kelulusan mahasiswa berhubungan erat dengan teknik statistik dan algoritma pembelajaran mesin yang digunakan untuk menganalisis data (Iddrus, 2024).

Indeks Prestasi Kumulatif

Rata-rata Nilai Kumulatif (IPK) menggambarkan kinerja akademik keseluruhan siswa pada semua mata kuliah yang diselesaikan. Nilai Rata-Rata (IPK) merupakan agregasi numerik dari semua nilai mata kuliah, yang berfungsi sebagai metrik standar untuk menilai dan membandingkan kinerja akademis dan pencapaian ilmiah siswa secara keseluruhan sepanjang masa pendidikan mereka. Mahasiswa yang memiliki IPK lebih besar dari 3,50 dan mematuhi etika akademik dianggap memiliki prestasi akademik yang tinggi (Mustafid et al., 2020).

Indeks Prestasi Semester

Indeks Prestasi Semester (IPS) merupakan metrik kuantitatif yang digunakan oleh lembaga akademik untuk mengevaluasi dan mengukur kinerja dan kemajuan akademik siswa selama semester tertentu. Setelah ujian akhir semester, perhitungan IPS dilakukan untuk mengetahui nilai indeks yang diperoleh mahasiswa. Nilai IPS yang dianggap baik adalah 3.00, yang merupakan standar indeks yang telah ditetapkan (Septiani & Nunik, 2022).

Satuan Kredit Semester

Sistem Kredit Semester (SKS) adalah sistem yang digunakan dalam penyelenggaraan pendidikan, di mana beban studi mahasiswa, tenaga pengajar, dan penyelenggara lembaga pendidikan diukur menggunakan satuan Kredit Semester. Semester berfungsi sebagai kerangka administratif utama dalam institusi pendidikan, menyediakan periode terstruktur untuk kegiatan akademik. Satuan Kredit Semester (SKS) merupakan bagian integral dari sistem ini, yang memungkinkan mahasiswa untuk merencanakan dan mengelola perkuliahan mereka secara efektif. Dengan menyesuaikan beban kuliah berdasarkan persyaratan SKS, mahasiswa dapat mengoptimalkan kemajuan akademik mereka dan memastikan penyelesaian program gelar mereka tepat waktu (Kurnia, 2015).

Lama Studi

Menurut Medika & Tomi (2020), Lama studi mengacu pada periode yang telah ditentukan yang dialokasikan bagi mahasiswa untuk memenuhi persyaratan program akademik mereka. Durasi ini berfungsi sebagai tolak ukur kemajuan mahasiswa dalam mencapai tujuan

pendidikan mereka dan memastikan bahwa mereka memperoleh pengetahuan dan keterampilan yang dibutuhkan dalam jangka waktu yang terstruktur. Lebih lanjut, lama studi yang ditentukan memiliki implikasi signifikan terhadap kualitas program secara keseluruhan, karena memengaruhi desain kurikulum, alokasi sumber daya, dan kinerja mahasiswa. Durasi ini juga memainkan peran penting dalam menjaga standar akreditasi program dan kredibilitas institusional (Medika & Tomi, 2020)

Metode *K-Nearest Neighbors* (KNN)

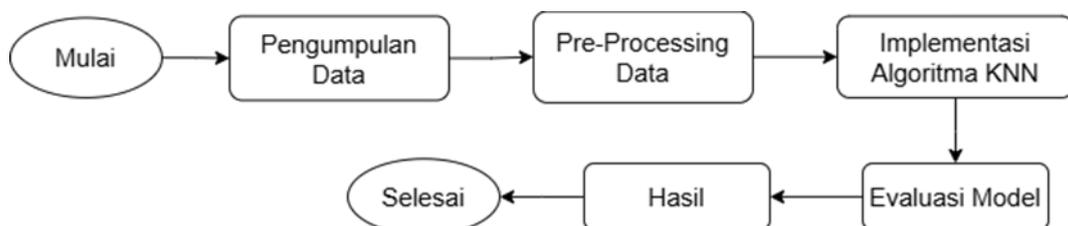
K-Nearest Neighbors (KNN) adalah algoritma pembelajaran mesin terawasi yang banyak digunakan. Algoritma ini melakukan tugas klasifikasi dan regresi dengan menganalisis kedekatan titik data dalam ruang fitur. Metode ini melibatkan perhitungan jarak umumnya *Euclidean* untuk mengidentifikasi 'k' tetangga terdekat dari suatu observasi. Algoritma kemudian menetapkan label kelas atau memprediksi nilai berdasarkan kelas mayoritas atau rata-rata titik data tetangga tersebut, menjadikannya pendekatan yang sederhana namun efektif untuk berbagai masalah pengenalan pola (Kordos & Blachnik, 2014).

Python

Python, bahasa pemrograman tingkat tinggi yang dicetuskan oleh Guido van Rossum pada tahun 1991, terkenal karena kesederhanaan dan keterbacaannya, sehingga mudah diakses oleh programmer pemula maupun berpengalaman. Dirancang untuk mendukung berbagai paradigma pemrograman termasuk pendekatan prosedural, berorientasi objek, dan fungsional bahasa ini mendukung pengembangan kode yang bersih dan efisien. Penekanannya pada kejelasan dan fleksibilitas telah berkontribusi signifikan terhadap adopsinya yang luas di berbagai domain seperti ilmu data, pengembangan *web*, dan otomatisasi (Dyer & Chauhan, 2022).

Google Colab

Google Colab (*Google Collaboratory*) adalah Lingkungan berbasis cloud Google Cloud Platform menyediakan kerangka kerja tangguh yang memungkinkan pengguna mengembangkan, mengeksekusi, dan mengelola kode Python dari jarak jauh. Infrastruktur ini secara signifikan meningkatkan kapabilitas dalam pemrosesan data, pengembangan kecerdasan buatan, dan aplikasi pembelajaran mesin dengan menghilangkan kebutuhan akan sumber daya perangkat keras lokal. Akibatnya, infrastruktur ini mendorong alur kerja yang skalabel, efisien, dan kolaboratif, memberdayakan peneliti dan pengembang untuk berinovasi secara mulus dalam ekosistem cloud yang aman dan fleksibel (Andarsyah & Yanuar, 2024).



Gambar 1. Alur Penelitian

HASIL DAN PEMBAHASAN

Pengumpulan Data

Tahap pertama dalam penelitian ini adalah pengumpulan data, yang melibatkan data mahasiswa Program Studi Teknik Informatika angkatan 2016 hingga 2020 yang telah menyelesaikan studi. Data tersebut mencakup informasi akademik seperti jumlah SKS kumulatif, IPK, IPS, semester, dan periode selesai studi. Berdasarkan hasil rekapitulasi, jumlah

mahasiswa per angkatan sekaligus pembagian berdasarkan status kelulusan yaitu angkatan 2016 terdiri dari 49 mahasiswa, dengan 1 mahasiswa lulus tepat waktu dan 48 mahasiswa tidak tepat waktu; angkatan 2017 berjumlah 62 mahasiswa, terdiri dari 12 mahasiswa tepat waktu dan 50 mahasiswa tidak tepat waktu; angkatan 2018 sebanyak 89 mahasiswa, dengan 57 mahasiswa tepat waktu dan 32 mahasiswa tidak tepat waktu; angkatan 2019 berjumlah 75 mahasiswa, terdiri dari 52 mahasiswa tepat waktu dan 23 mahasiswa tidak tepat waktu; sedangkan angkatan 2020 berjumlah 88 mahasiswa yang seluruhnya lulus tepat waktu. Secara keseluruhan, jumlah data yang digunakan dalam penelitian ini adalah 363 mahasiswa. Data ini menunjukkan variasi yang signifikan pada distribusi kelulusan antar angkatan, yang menjadi dasar penting dalam proses analisis prediksi kelulusan menggunakan algoritma K-Nearest Neighbor (KNN).

| 1 | NIM | Nama Mahasiswa | Angkatan | Status | IPS | IPK | Semester | SKS Kumulatif | Periode Selesai |
|----|---------|-------------------------------------|----------|--------|------|-----------|----------|---------------|-----------------|
| 2 | 2116001 | NI PUTU FRALINCHIA CITRA ARYA PUTRI | 2016 | A | | 0 3,24 | 20211 | 150 | 20202 |
| 3 | 2116001 | NI PUTU FRALINCHIA CITRA ARYA PUTRI | 2016 | A | | 4 3,4 | 20212 | 150 | 20202 |
| 4 | 2116002 | ELMIANTI KALE HERE | 2016 | A | | 1,77 2,66 | 20211 | 135 | 20222 |
| 5 | 2116002 | ELMIANTI KALE HERE | 2016 | A | | 0 2,49 | 20212 | 144 | 20222 |
| 6 | 2116002 | ELMIANTI KALE HERE | 2016 | A | | 0 2,58 | 20221 | 144 | 20222 |
| 7 | 2116002 | ELMIANTI KALE HERE | 2016 | A | | 3,33 2,78 | 20222 | 144 | NULL |
| 8 | 2116003 | SISILIA SRIAYU LESTARI | 2016 | A | | 3,5 3,2 | 20211 | 150 | NULL |
| 9 | 2116004 | YESYURUN ROBI HUMA DOKO | 2016 | A | | 1 3,23 | 20211 | 137 | 20221 |
| 10 | 2116004 | YESYURUN ROBI HUMA DOKO | 2016 | A | | 0,83 3,22 | 20212 | 140 | 20221 |
| 11 | 2116004 | YESYURUN ROBI HUMA DOKO | 2016 | A | | 3,5 3,21 | 20221 | 146 | NULL |
| 12 | 2116005 | WILLIAM CHRISTIANSEN | 2016 | A | | 4 3,24 | 20211 | 150 | NULL |
| 13 | 2116006 | EUNIKE VIA WARTONO | 2016 | NULL | NULL | NULL NULL | NULL | NULL | NULL |
| 14 | 2116008 | DEVIYANA TATU WADANG | 2016 | A | | 0 3,13 | 20211 | 138 | 20221 |
| 15 | 2116008 | DEVIYANA TATU WADANG | 2016 | A | | 0 3,13 | 20212 | 138 | 20221 |
| 16 | 2116008 | DEVIYANA TATU WADANG | 2016 | A | | 4 3,18 | 20221 | 144 | NULL |
| 17 | 2116009 | MESAK KATANGA MULUNG | 2016 | A | | 0 2,83 | 20211 | 144 | 20222 |
| 18 | 2116009 | MESAK KATANGA MULUNG | 2016 | A | | 3,5 2,97 | 20212 | 144 | 20222 |

Gambar 2. Pengumpulan Data

Pre-Processing Data

Setelah data terkumpul, tahap selanjutnya adalah pre-processing data. Pada tahap ini, data yang diperoleh akan dipersiapkan agar dapat digunakan dalam algoritma KNN. Adapun tahap-tahap dalam pre-processing data yaitu : pembersihan data, pelabelan, Normalisasi data, dan pemisahan data:

1. Pembersihan Data

Pada tahap Pre-Processing Data, dataset yang telah dikumpulkan akan dipersiapkan agar dapat digunakan dalam algoritma K-Nearest Neighbor (KNN). Data mentah yang diperoleh menunjukkan beberapa kolom dengan informasi penting seperti NIM, Nama Mahasiswa, Angkatan, Status, IPS, IPK, Semester, SKS Kumulatif, dan Periode Selesai. Kemudian akan dilakukan pembersihan data, dengan tujuan untuk menghapus data yang tidak lengkap atau tidak relevan, seperti data mahasiswa yang tidak memiliki nilai lengkap untuk variabel-variabel yang digunakan dalam penelitian. Source code untuk melakukan proses pembersihan dapat dilihat pada gambar berikut:

```

1 #proses menghapus data kosong atau null
2 df_cleaned = df.dropna(axis=1)
3 df_cleaned.head()
4 df.info()

<class 'pandas.core.frame.DataFrame'>
Index: 1352 entries, 0 to 1713
Data columns (total 10 columns):

```

Gambar 3. Proses Pembersihan Data

Pada tahap pembersihan data, kolom-kolom yang mengandung nilai kosong atau null dihapus. Hasilnya, dataset yang sebelumnya terdiri dari 1.715 entri, kini memiliki 1.352 entri yang lengkap dan bebas dari nilai kosong.

Adapun hasil dari proses pembersihan data yang dapat dilihat pada gambar ini

| | | | | | | | | | |
|----|---------|------------------------------------|------|---|------|------|---------|-------|---------|
| 0 | 2116001 | NI PUTU FRALINCHIA CITRAARYA PUTRI | 2016 | A | 0.0 | 3.24 | 20211.0 | 150.0 | 20202.0 |
| 1 | 2116001 | NI PUTU FRALINCHIA CITRAARYA PUTRI | 2016 | A | 4.0 | 3.4 | 20212.0 | 150.0 | 20202.0 |
| 2 | 2116002 | ELMIANTI KALE HERE | 2016 | A | 1.77 | 2.66 | 20211.0 | 135.0 | 20222.0 |
| 3 | 2116002 | ELMIANTI KALE HERE | 2016 | A | 0.0 | 2.49 | 20212.0 | 144.0 | 20222.0 |
| 4 | 2116002 | ELMIANTI KALE HERE | 2016 | A | 0.0 | 2.58 | 20221.0 | 144.0 | 20222.0 |
| 7 | 2116004 | YESYURUN ROBI HUMA DOKO | 2016 | A | 1.0 | 3.23 | 20211.0 | 137.0 | 20221.0 |
| 8 | 2116004 | YESYURUN ROBI HUMA DOKO | 2016 | A | 0.83 | 3.22 | 20212.0 | 140.0 | 20221.0 |
| 12 | 2116008 | DEVYANA TATU WADANG | 2016 | A | 0.0 | 3.13 | 20211.0 | 138.0 | 20221.0 |
| 13 | 2116008 | DEVYANA TATU WADANG | 2016 | A | 0.0 | 3.13 | 20212.0 | 138.0 | 20221.0 |
| 15 | 2116009 | MESAK KATANGA MULUNG | 2016 | A | 0.0 | 2.83 | 20211.0 | 144.0 | 20222.0 |
| 16 | 2116009 | MESAK KATANGA MULUNG | 2016 | A | 3.5 | 2.97 | 20212.0 | 144.0 | 20222.0 |
| 19 | 2116012 | PENI GARAMESA | 2016 | A | 0.0 | 3.46 | 20211.0 | 138.0 | 20212.0 |
| 21 | 2116013 | SHAHNAZ PRATIWI SIU | 2016 | A | 0.0 | 2.9 | 20211.0 | 138.0 | 20231.0 |
| 22 | 2116013 | SHAHNAZ PRATIWI SIU | 2016 | A | 0.0 | 2.9 | 20212.0 | 138.0 | 20231.0 |
| 23 | 2116013 | SHAHNAZ PRATIWI SIU | 2016 | A | 3.5 | 2.87 | 20221.0 | 144.0 | 20231.0 |
| 25 | 2116016 | IRMA TINGGI ATAHAU | 2016 | A | 0.0 | 3.05 | 20211.0 | 144.0 | 20212.0 |
| 27 | 2116018 | HERDINANT KALAMBAR LAMIK | 2016 | A | 0.0 | 2.76 | 20211.0 | 144.0 | 20222.0 |

Gambar 4. Hasil Pembersihan Data

2. Pelabelan

Selanjutnya akan dilakukan pelabelan, dimana Dalam penelitian ini, pelabelan dilakukan berdasarkan lama studi mahasiswa. Mahasiswa yang berhasil menyelesaikan studi dalam waktu 8 semester diberi label “Tepat Waktu (TW)”, sedangkan yang menyelesaikan lebih dari 8 semester diberi label “Tidak Tepat Waktu (TTW)”. *Source code* untuk melakukan pelabelan dapat dilihat pada gambar berikut:

```
1 # proses pelabelan
2 df['Jumlah Semester'] = df['Periode Selesai'] - df['Angkatan']
3 df['Label'] = df['Jumlah Semester'].apply(lambda x: 'TW' if x <= 4 else 'TTW')
4 df
```

Gambar 5. Proses pelabelan

Dari source di atas, diperoleh hasil pelabelan. Hasil tersebut dapat dilihat pada tabel berikut:

Tabel 2. Hasil Pelabelan

| Nama | Angkatan | IPK | Periode Selesai | Label |
|--------------------|----------|------|-----------------|-------|
| NI PUTU | 2016 | 3.24 | 2020 | TW |
| ELMIANTI KALE HERE | 2016 | 2.78 | 2022 | TTW |

3. Normalisasi Data

Dalam penelitian ini, metode yang digunakan untuk normalisasi adalah *Min-Max Normalization*, yang berfungsi untuk mengubah nilai data sehingga berada dalam rentang antara 0 dan 1. Proses ini dilakukan dengan cara mengurangi nilai minimum data dan membaginya dengan rentang antara nilai maksimum dan minimum. *Source code* untuk melakukan proses normalisasi dapat dilihat pada gambar berikut:

```

1 # Normalisasi data untuk fitur SKS Kumulatif, IPK, dan IPS
2 features = ['SKS Kumulatif', 'IPK', 'IPS']
3
4 for feature in features:
5     x_min = df[feature].min()
6     x_max = df[feature].max()
7     df[feature] = (df[feature] - x_min) / (x_max - x_min)
8
9 # Menampilkan hasil setelah normalisasi
10 display(df)

```

Gambar 6. Proses Normalisasi

Hasil dari proses normalisasi dapat dilihat pada tabel berikut:

Tabel 3. Hasil Normalisasi

| Nama | SKS Kumulatif | IPK | IPS |
|--------------------|---------------|------|------|
| ELMIANTI KALE HERE | 0.88 | 0.29 | 0.00 |
| NI PUTU | 0.93 | 0.72 | 1.0 |

4. Pembagian Data Latih Dan Data Uji

Selama tahap Pembagian Data Pelatihan dan Uji, set data yang telah diproses dibagi secara sistematis menjadi dua subset untuk memfasilitasi pengembangan dan evaluasi model. Biasanya, rasio 80:20 digunakan, dengan delapan puluh persen data dialokasikan untuk pelatihan algoritma pembelajaran mesin, yang memungkinkannya mempelajari pola-pola dasar, sementara dua puluh persen sisanya dialokasikan untuk pengujian, yang memungkinkan penilaian kinerja prediktif model secara objektif. Source code untuk melakukan pembagian data dapat dilihat pada gambar berikut:

```

1 #pembagian data
2 from sklearn.model_selection import train_test_split
3 X = df[['SKS Kumulatif', 'IPK', 'IPS']]
4 y = df['Label']
5 # Membagi data menjadi 80% untuk pelatihan dan 20% untuk pengujian
6 X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
7 # Menampilkan ukuran data latih dan data uji
8 print(f"Data latih: {X_train.shape[0]} baris")
9 print(f"Data uji: {X_test.shape[0]} baris")

```

Gambar 7. Proses Pembagian Data

Implementasi Algoritma KNN

Pada tahap ini, algoritma K-Nearest Neighbors (KNN) diterapkan untuk memprediksi kelulusan mahasiswa berdasarkan data yang telah diproses. KNN bekerja dengan cara mengidentifikasi lima tetangga terdekat (K=5) dari setiap data mahasiswa yang ingin diprediksi kelulusannya, menggunakan pengukuran jarak Euclidean antar fitur seperti IPK, jumlah SKS, dan fitur akademik lainnya. Prediksi kelulusan ditentukan berdasarkan mayoritas kelas kelulusan yang terdapat pada kelima tetangga tersebut, apakah mahasiswa tersebut lulus tepat waktu atau tidak tepat waktu. Pemilihan nilai K=5 didasarkan pada hasil uji coba beberapa nilai K yang berbeda, dimana nilai ini memberikan akurasi terbaik dalam memisahkan kelas kelulusan secara konsisten. Nilai ini juga cukup representatif karena menghindari sensitivitas terhadap outlier dan dominasi kelas mayoritas.

Hasil akhir dari implementasi KNN dalam penelitian ini berupa prediksi kelas kelulusan mahasiswa pada data uji yang telah dipisahkan dari data latih. Evaluasi performa model dilakukan menggunakan metrik seperti akurasi, precision, recall, dan F1-score. Model KNN dengan K=5 berhasil mencapai tingkat akurasi sebesar [masukkan nilai akurasi]% pada data uji, yang menunjukkan kemampuan model dalam memprediksi kelulusan mahasiswa secara

tepat. Selain itu, hasil klasifikasi divisualisasikan melalui confusion matrix yang memperlihatkan jumlah mahasiswa yang diprediksi lulus tepat waktu dan tidak tepat waktu dengan benar maupun yang salah klasifikasi. Dengan adanya hasil ini, dapat diketahui sejauh mana model KNN mampu membantu dalam memetakan potensi kelulusan mahasiswa, sehingga dapat menjadi alat bantu dalam pengambilan keputusan akademik seperti pemberian intervensi dini kepada mahasiswa yang berpotensi tidak lulus tepat waktu.. Source code untuk penerapan algoritma KNN dapat dilihat pada gambar ini:

```
1 # implementasi KNN
2 from sklearn.neighbors import KNeighborsClassifier
3 knn = KNeighborsClassifier(n_neighbors=5)
4 knn.fit(X_train, y_train)
5 # evaluasi model
6 from sklearn.metrics import classification_report, accuracy_score
7 y_pred = knn.predict(X_test)
8 accuracy = accuracy_score(y_test, y_pred)
9 report = classification_report(y_test, y_pred)
```

Gambar 8. Implementasi Algoritma KNN

Evaluasi Model

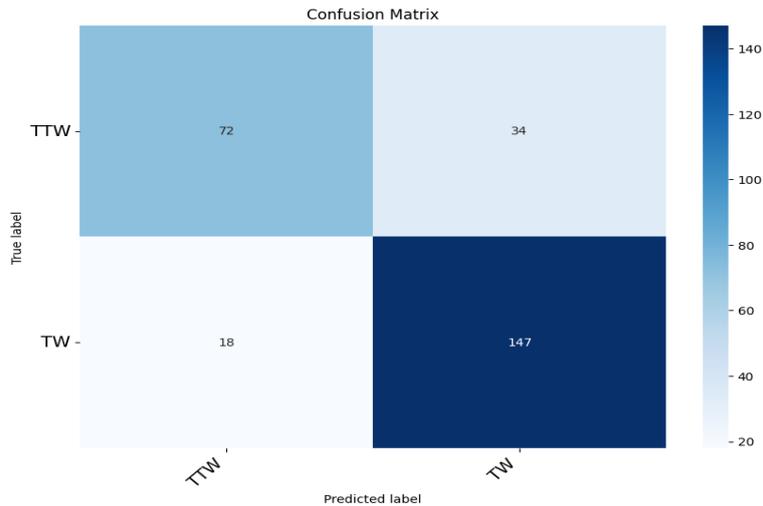
Tahap terakhir dalam proses ini adalah evaluasi model. Setelah model KNN menghasilkan prediksi, kinerjanya diuji dengan menggunakan data uji yang belum pernah dilihat oleh model selama pelatihan. Evaluasi dilakukan untuk mengukur seberapa baik model dalam memprediksi kelulusan mahasiswa dengan menggunakan beberapa metrik evaluasi, yaitu akurasi, precision, recall, dan F1-Score.

Akurasi memberikan ukuran umum ketepatan prediksi model terhadap seluruh data uji. Precision mengukur proporsi prediksi positif yang benar di antara semua prediksi positif yang dibuat oleh model, yang mencerminkan ketepatan model dalam memprediksi mahasiswa yang lulus tepat waktu. Recall mengukur kemampuan model dalam mengidentifikasi seluruh mahasiswa yang sebenarnya lulus tepat waktu, yaitu sensitivitas model terhadap kelas positif. Sementara itu, F1-Score merupakan rata-rata harmonis dari precision dan recall, yang memberikan penilaian seimbang antara ketepatan dan kemampuan deteksi model.. Source code untuk melakukan evaluasi model dapat dilihat pada gambar berikut:

```
10
11 # Menampilkan hasil evaluasi
12 print(f"Akurasi model KNN: {accuracy * 100:.2f}%")
13 print(report)
```

Gambar 9. Evaluasi Model

Setelah proses evaluasi ini dilakukan, hasil yang diperoleh dari performa model KNN dalam memprediksi kelulusan mahasiswa Universitas Kristen Wira Wacana Sumba dapat dilihat pada gambar berikut:



Gambar 10. *Confusion Matriks*

```

Akurasi model KNN: 81.18%
precision    recall  f1-score   support

   TTW      0.81    0.68    0.74     106
   TW       0.81    0.90    0.85     165

 accuracy          0.81     271
 macro avg         0.81    0.79    0.80     271
 weighted avg      0.81    0.81    0.81     271
  
```

Gambar 11. Hasil Evaluasi Model

Hasil evaluasi model *K-Nearest Neighbors* (KNN) menunjukkan akurasi sebesar 81,18%, yang berarti model mampu melakukan klasifikasi dengan benar pada sekitar 8 dari 10 data uji. Pada kelas TTW, *precision* sebesar 0,81 mengindikasikan bahwa 81% prediksi TTW tepat sasaran, namun *recall* hanya 0,68 yang menunjukkan masih ada sebagian data TTW yang tidak terdeteksi dengan baik, sehingga *f1-score*-nya berada pada angka 0,74. Sebaliknya, kelas TW memiliki *precision* yang sama yaitu 0,81, namun dengan *recall* lebih tinggi sebesar 0,90, yang menandakan kemampuan deteksi kelas ini jauh lebih baik dan menghasilkan *f1-score* sebesar 0,85. Nilai *macro average* untuk *precision* (0,81), *recall* (0,79), dan *f1-score* (0,80) menunjukkan performa rata-rata antar kelas tanpa memperhitungkan jumlah data, sedangkan *weighted average* yang mempertimbangkan distribusi data menghasilkan nilai yang sama dengan akurasi keseluruhan, menandakan performa yang seimbang di kedua kelas. Secara umum, model ini cukup andal, meskipun masih perlu perbaikan pada peningkatan *recall* di kelas TTW agar deteksi kelas tersebut menjadi lebih optimal.

KESIMPULAN

Berdasarkan hasil implementasi dan evaluasi model *K-Nearest Neighbors* (KNN) pada prediksi kelulusan mahasiswa Program Studi Teknik Informatika Universitas Kristen Wira Wacana Sumba, model dengan parameter $K=5$ mampu mencapai tingkat akurasi sebesar 81,18% pada data uji. Hasil ini menunjukkan bahwa model dapat mengklasifikasikan kelulusan mahasiswa dengan cukup baik, di mana pada kelas Tepat Waktu (TW) diperoleh *precision* sebesar 0,81, *recall* sebesar 0,90, dan *f1-score* sebesar 0,85, sedangkan pada kelas Tidak Tepat Waktu (TTW)

diperoleh *precision* sebesar 0,81, *recall* sebesar 0,68, dan *f1-score* sebesar 0,74. Nilai *macro average* untuk *precision*, *recall*, dan *f1-score* masing-masing sebesar 0,81; 0,79; dan 0,80, sementara *weighted average* menunjukkan hasil yang konsisten dengan akurasi keseluruhan, menandakan distribusi performa yang seimbang di kedua kelas. Hasil ini membuktikan bahwa KNN dapat digunakan sebagai alat bantu dalam memetakan potensi kelulusan mahasiswa sehingga memungkinkan intervensi dini bagi mahasiswa yang berisiko tidak lulus tepat waktu. Namun demikian, masih diperlukan perbaikan pada peningkatan recall kelas TTW agar model dapat lebih optimal dalam mendeteksi mahasiswa yang berpotensi mengalami keterlambatan kelulusan.

DAFTAR PUSTAKA

- Andarsyah, R., & Yanuar, A. (2024). Jurnal Teknik Informatika, Vol. 16, No. 2, April 2024. *Jurnal Teknik Informatika*, 16(2), 1–7.
- Dyer, R., & Chauhan, J. (2022). An Exploratory Study on the Predominant Programming Paradigms in Python Code. *Proceedings of the 30th ACM Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering (ESEC/FSE '22)*, 1(1), 14–18. <https://doi.org/10.1145/3540250.3549158>
- Iddrus, A. (2024). Jurnal Advance Research Informatika Prediksi Kelulusan Mahasiswa Tepat Waktu atau Tidak Menggunakan Metode Naïve Bayes Classifier (Studi kasus : Mahasiswa Fakultas Teknik). *Jurnal Advance Research Informatika*, 3(1), 1–10.
- Kordos, M., & Blachnik, M. (2014). Do We Need Whatever More Than k-NN? Do We Need Whatever More than k-NN? *Lecture Notes in Computer Science, January 2010*. <https://doi.org/10.1007/978-3-642-13208-7>
- Kurnia, H. (2015). Academy Of Education Journal. Pendidikan Pancasila dan Kewarganegaraan Vol. 5 No. 2 Juli 2014. *Academy Of Education Journal. Pendidikan Pancasila Dan Kewarganegaraan*, 5(2), 91–103.
- Medika, G. H., & Tomi, Z. B. (2020). Hubungan Lama Studi dengan Nilai Ujian Komprehensif Mahasiswa Program Studi Pendidikan Matematika IAIN Bukittinggi. *JURNAL SAINS DAN TEKNOLOGI*, 12(1), 2018–2021.
- Mustafid, M. F., Wedi, A., & Adi, pranomo E. (2020). BERDASARKAN GAYA BELAJAR PADA MAHASISWA JURUSAN TEKNOLOGI PENDIDIKAN UNIVERSITAS NEGERI MALANG ANGKATAN 2017. *JINOTEP (Jurnal Inovasi Teknologi Pembelajaran)*, 6(2), 119–128. <https://doi.org/10.17977/um031v6i22020p119>
- Septiani, W. P., & Nunik, P. (2022). Analisis Perbandingan Indeks Prestasi (IP) Mahasiswa Sebelum dan Saat Pandemi Menggunakan Metode Naïve Bayes. *Seminar Nasional TEKNOKA*, 7(2502), 100–105.