



Analisis Metode **KNN** K-Nearest Neighbor

dalam Mempredikasi Kelulusan Mahasiswa

Fratika Rahmaliana
Budianto bangun, S.Sos., M.Kom
Sudi Suryadi, S.Kom., M.Kom
Masrizal, S.Kom., M.Kom



Analisis Metode K-Nearest Neighbor

dalam Mempredikasi Kelulusan Mahasiswa

Fratika Rahmaliana
Budianto bangun, S.Sos., M.Kom
Sudi Suryadi, S.Kom., M.Kom
Masrizal, S.Kom., M.Kom

 Penerbit

**ANALISIS METODE K-NEAREST NEIGHBOR (KNN) DALAM
MEMPREDIKASI KELULUSAN MAHASISWA**

Ditulis oleh:

Fratika Rahmaliana
Budianto Bangun, S.Sos., M.Kom.
Sudi Suryadi, S.Kom., M.Kom.
Masrizal, S.Kom., M.Kom.

Diterbitkan, dicetak, dan didistribusikan oleh
PT. Literasi Nusantara Abadi Grup
Perumahan Puncak Joyo Agung Residence Kav. B11 Merjosari
Kecamatan Lowokwaru Kota Malang 65144
Telp : +6285887254603, +6285841411519
Email: literasinusantaraofficial@gmail.com
Web: www.penerbitlitnus.co.id
Anggota IKAPI No. 340/JTI/2022



Hak Cipta dilindungi oleh undang-undang. Dilarang mengutip
atau memperbanyak baik sebagian ataupun keseluruhan isi buku
dengan cara apa pun tanpa izin tertulis dari penerbit.

Cetakan I, April 2025

Perancang sampul: Rosyiful Aqli
Penata letak: Muhammad Ridho Naufal

ISBN : 978-623-127-391-8

viii + 58 hlm. ; 15,5x23 cm.

©April 2025



Prakata

Puji syukur saya panjatkan kepada Allah SWT berkat Rahmat, Hidayah, dan Karunia-Nya kepada kita sehingga saya dapat menyelesaikan proposal skripsi dengan judul “Analisis Metode KNN dalam Memprediksi Kelulusan Mahasiswa Universitas Labuhanbatu Fakultas Sains dan Teknologi Prodi Sistem Informasi”. Laporan proposal skripsi ini disusun sebagai salah satu syarat untuk mengerjakan tugas akhir pada program Studi Sistem Informasi Fakultas Sains Dan Teknologi universitas labuhanbatu.

Saya sebagai Penulis menyadari dalam penyusunan proposal skripsi ini tidak akan dapat terselesaikan tanpa bantuan dari berbagai pihak. Karena itu pada kesempatan ini saya ingin mengucapkan terima kasih kepada:

1. Bapak Rektor Universitas Labuhanbatu, Bapak Assoc. Prof. Ade Parlaungan Nasution, Ph.D
2. Bapak Dr. Iwan Purnama, S.Kom., M.Kom selaku Dekan Fakultas Sains Dan Teknologi Universitas Labuhanbatu.
3. Bapak Budianto Bangun S.Sos., M.Kom, selaku Kepala Program Studi Sistem Informasi.
4. Bapak Budianto Bangun S.Sos., M.Kom selaku Dosen Pembimbing 1.
5. Bapak Sudi Suryadi, S.Kom., M.Kom selaku Dosen Pembimbing 2.

Saya ucapkan terimakasih kepada orang tua saya yang selalu mendukung kuliah saya sampai menyanggah gelar Sarjana Komputer (S.Kom) dan

kepada teman-teman seperjuangan dengan saya, terimakasih telah berjuang Bersama dalam penelitian dan pengerjaan proposal skripsi ini. Dan teman seperjuangan kelas sistem informasi. saya menyadari proposal penelitian ini tidak luput dari berbagai kekurangan. Penulisan mengharapkan saran dan kritik untuk perbaikannya sehingga plaporan proposal penelitian ini dapat memberikan manfaat bagi banyak orang-orang dalam bidang komputer.

Rantauprapat, 2025

Penulis

Fratika Rahmaliana



Daftar Isi

Prakata	iii
Daftar Isi	v

BAB I

PENDAHULUAN—1

Latar Belakang.....	1
Tantangan institusi dalam mengelola kelulusan tepat waktu	3
Dampak kelulusan tepat waktu terhadap kualitas pendidikan dan sumber daya	4
Masalah yang dihadapi oleh institusi dalam melakukan prediksi kelulusan mahasiswa	5
Manfaat Praktis bagi Universitas dalam Merancang Kebijakan Pendidikan.....	7

BAB II

TEORI DASAR—9

Data Mining dan Penerapannya dalam Pendidikan	9
<i>Knowledge Discovery in Database</i> (KDD)	13
Klasifikasi dalam Machine Learning	15
Algoritma K-Nearest Neighbor (KNN)	16
Microsoft Excel	18
Rapidminer	19

BAB III

METODOLOGI PENELITIAN—23

Desain Penelitian	23
Pengumpulan Data	24
Proses <i>Knowledge Discovery in Database</i> (KDD)	26

BAB IV

HASIL PENGUJIAN DAN PEMBAHASAN—37

Penerapan Algoritma KNN	37
Perhitungan jarak Euclidean dan pemilihan K	42
Implementasi dengan Rapidminer	43
Pengujian Model	46
Evaluasi Kinerja Model	47
Faktor-faktor yang mempengaruhi akurasi dan efisiensi model KNN	48

BAB V

PENUTUP—51

Kesimpulan..... 51

Rekomendasi..... 52

Daftar Pustaka..... 55

Tentang Penulis..... 57



BAB I

PENDAHULUAN

Latar Belakang

Kelulusan mahasiswa tepat waktu merupakan salah satu indikator penting dalam menilai efektivitas sistem pendidikan tinggi. Di banyak universitas, termasuk Universitas Labuhanbatu, tantangan utama yang dihadapi adalah ketidakseimbangan antara jumlah mahasiswa yang diterima dan jumlah mahasiswa yang berhasil menyelesaikan studi sesuai dengan waktu yang ditetapkan. Permasalahan ini tidak hanya berdampak pada kualitas pendidikan, tetapi juga pada efisiensi operasional institusi pendidikan. Keterlambatan kelulusan mahasiswa dapat mengganggu perencanaan akademik, pengelolaan kapasitas kelas, dan penggunaan sumber daya yang tersedia, yang pada akhirnya dapat mempengaruhi reputasi dan akreditasi universitas. Selain itu, bagi mahasiswa, kelulusan yang tepat waktu sangat penting untuk meningkatkan daya saing mereka di pasar kerja dan meraih peluang karier yang lebih baik.

Untuk mengatasi masalah ini, banyak universitas mulai memanfaatkan teknologi untuk menganalisis pola-pola yang ada dalam data akademik mahasiswa guna memprediksi kelulusan mereka. Salah satu pendekatan yang digunakan adalah data mining, yang memungkinkan pengolahan data

dalam jumlah besar untuk menemukan pola atau hubungan yang relevan dalam proses kelulusan mahasiswa. Metode K-Nearest Neighbors (KNN) adalah salah satu algoritma yang digunakan dalam data mining untuk memprediksi hasil berdasarkan kemiripan data yang ada. KNN bekerja dengan menghitung jarak antara data yang ingin diprediksi dengan data lainnya yang telah dilabeli, dan kemudian menentukan kelas berdasarkan mayoritas tetangga terdekatnya. Kelebihan utama dari metode ini adalah kemampuannya dalam menangani data yang memiliki pola non-linear, serta keakuratan yang tinggi jika diterapkan dengan benar.

Prediksi kelulusan mahasiswa yang akurat dapat memberikan manfaat yang signifikan bagi pengelolaan akademik di universitas. Dengan informasi prediktif yang tepat, universitas dapat merancang kebijakan pendidikan yang lebih efisien, mengalokasikan sumber daya dengan lebih baik, serta memfokuskan perhatian pada mahasiswa yang berisiko tidak lulus tepat waktu. Di sisi lain, bagi mahasiswa, prediksi ini dapat memberikan gambaran yang jelas mengenai kemungkinan kelulusan mereka, memungkinkan mereka untuk mengambil tindakan preventif jika diperlukan, seperti mengikuti bimbingan akademik atau memperbaiki IPK mereka. Oleh karena itu, penelitian yang menggunakan metode KNN untuk memprediksi kelulusan mahasiswa di Universitas Labuhanbatu ini bertujuan untuk memberikan solusi berbasis data yang dapat meningkatkan kelulusan tepat waktu dan membantu universitas dalam merencanakan langkah-langkah strategis yang lebih efektif.

Penelitian ini juga bertujuan untuk mengisi kesenjangan yang ada dalam literatur terkait dengan perbandingan metode KNN dengan metode data mining lainnya dalam memprediksi kelulusan mahasiswa. Beberapa penelitian sebelumnya menunjukkan bahwa KNN memiliki keunggulan dalam hal akurasi prediksi, namun terdapat variasi hasil yang tergantung pada data dan konteks yang digunakan. Oleh karena itu, penelitian ini penting dilakukan untuk menguji keakuratan metode KNN dalam konteks Universitas Labuhanbatu dan mengevaluasi bagaimana metode ini dapat membantu meningkatkan efektivitas sistem pendidikan di sana.

Melalui penerapan metode KNN, diharapkan penelitian ini dapat memberikan kontribusi bagi pengelolaan kelulusan mahasiswa yang lebih tepat waktu, serta memberikan wawasan yang lebih mendalam mengenai faktor-faktor yang mempengaruhi kelulusan mahasiswa di Universitas Labuhanbatu, khususnya dalam Program Studi Sistem Informasi. Dengan demikian, hasil penelitian ini dapat menjadi referensi bagi universitas lain yang ingin menerapkan teknologi serupa dalam meningkatkan efisiensi akademik mereka.

Tantangan institusi dalam mengelola kelulusan tepat waktu

Dalam dunia pendidikan tinggi, kelulusan mahasiswa tepat waktu menjadi salah satu indikator utama yang digunakan untuk menilai kualitas sebuah institusi pendidikan. Universitas, termasuk Universitas Labuhanbatu, menghadapi tantangan besar dalam mengelola kelulusan mahasiswa, terutama dengan semakin meningkatnya jumlah mahasiswa yang diterima setiap tahunnya. Meskipun jumlah mahasiswa yang diterima mengalami peningkatan, tidak semua mahasiswa dapat menyelesaikan studi mereka tepat waktu sesuai dengan kurikulum yang ditetapkan. Fenomena ini menyebabkan ketidakseimbangan antara jumlah mahasiswa yang masuk dan jumlah mahasiswa yang berhasil menyelesaikan studi tepat waktu. Ketidaktepatan waktu kelulusan dapat memberikan dampak negatif yang cukup besar, baik bagi mahasiswa itu sendiri, institusi pendidikan, maupun bagi dunia kerja.

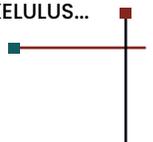
Bagi mahasiswa, kelulusan tepat waktu berperan penting dalam mempersiapkan diri untuk memasuki dunia kerja dengan kompetensi yang sudah teruji dan meningkatkan daya saing mereka. Mahasiswa yang lulus tepat waktu memiliki kesempatan lebih besar untuk mendapatkan pekerjaan sesuai dengan bidangnya, serta membangun jaringan profesional yang lebih cepat. Di sisi lain, bagi institusi pendidikan, ketidaktepatan waktu kelulusan mahasiswa berpotensi merugikan reputasi akademik dan mempengaruhi peringkat institusi tersebut dalam penilaian akreditasi. Universitas yang mampu mempertahankan angka kelulusan tepat waktu yang tinggi menunjukkan kualitas manajerial yang baik, serta pengelolaan

sumber daya yang efisien, yang dapat meningkatkan daya tarik bagi calon mahasiswa baru.

Selain itu, ketidaktepatan waktu kelulusan juga berdampak pada kapasitas kelas, rasio dosen terhadap mahasiswa, dan perencanaan kurikulum. Jumlah mahasiswa yang tertunda kelulusannya dapat meningkatkan beban administrasi dan akademik, serta memerlukan penyesuaian sumber daya yang lebih besar, seperti kapasitas ruang kelas, waktu dosen, dan fasilitas pendukung lainnya. Oleh karena itu, penting bagi institusi untuk memiliki sistem yang efektif dalam memprediksi kelulusan mahasiswa, sehingga dapat mengambil tindakan preventif yang diperlukan. Implementasi teknologi, seperti metode data mining dan algoritma K-Nearest Neighbors (KNN), dapat membantu institusi untuk menganalisis data akademik dan memprediksi kelulusan mahasiswa secara akurat. Dengan demikian, institusi pendidikan dapat merencanakan strategi yang lebih baik dalam mengelola mahasiswa dan memastikan bahwa lebih banyak mahasiswa lulus tepat waktu.

Dampak kelulusan tepat waktu terhadap kualitas pendidikan dan sumber daya

Kelulusan tepat waktu merupakan indikator penting dalam sistem pendidikan tinggi, khususnya bagi perguruan tinggi yang memiliki tujuan untuk menghasilkan lulusan yang kompeten dan siap menghadapi dunia kerja. Dalam konteks Universitas Labuhanbatu, memastikan mahasiswa lulus tepat waktu memiliki dampak yang signifikan terhadap kualitas pendidikan dan pengelolaan sumber daya. Ketidaktepatan waktu kelulusan dapat menimbulkan berbagai masalah, baik bagi mahasiswa itu sendiri, institusi pendidikan, maupun masyarakat luas. Bagi mahasiswa, kelulusan tepat waktu adalah kunci untuk meningkatkan daya saing di pasar kerja. Sebaliknya, keterlambatan kelulusan dapat mengurangi peluang mahasiswa untuk memasuki dunia kerja pada waktu yang tepat, mempengaruhi penghasilan yang dapat diperoleh, dan berpotensi menyebabkan mahasiswa kesulitan dalam mengikuti perkembangan industri.



Bagi institusi pendidikan, kelulusan tepat waktu berpengaruh langsung pada reputasi dan kualitas akademik. Universitas yang mampu menjaga tingkat kelulusan tepat waktu yang tinggi cenderung memiliki citra yang baik di mata calon mahasiswa dan masyarakat. Selain itu, pengelolaan sumber daya, seperti kapasitas kelas, jumlah dosen yang diperlukan, dan fasilitas pendukung lainnya, juga sangat dipengaruhi oleh kelulusan tepat waktu. Penumpukan mahasiswa yang belum lulus tepat waktu dapat menyebabkan penurunan kualitas pendidikan karena keterbatasan ruang kelas, akses ke dosen, serta fasilitas lainnya. Dengan memastikan mahasiswa lulus tepat waktu, universitas dapat mengalokasikan sumber daya dengan lebih efisien, memberikan kesempatan bagi mahasiswa baru untuk belajar, dan menjaga kualitas layanan pendidikan yang ditawarkan.

Selain itu, kelulusan tepat waktu dapat membantu dalam perencanaan jangka panjang universitas, baik dalam hal pengembangan kurikulum maupun alokasi anggaran. Misalnya, pengurangan jumlah mahasiswa yang tertunda kelulusan dapat mengurangi biaya operasional yang dikeluarkan oleh universitas, seperti biaya untuk pengajaran tambahan, administrasi, dan fasilitas yang harus dipertahankan lebih lama. Oleh karena itu, prediksi kelulusan yang akurat dengan menggunakan metode seperti KNN menjadi sangat penting, karena memberikan informasi yang berguna untuk meningkatkan efisiensi pengelolaan akademik dan kualitas pendidikan.

Masalah yang dihadapi oleh institusi dalam melakukan prediksi kelulusan mahasiswa

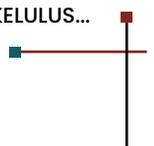
Institusi pendidikan tinggi, seperti Universitas Labuhanbatu, menghadapi berbagai tantangan signifikan dalam upaya memprediksi kelulusan mahasiswa secara tepat waktu. Salah satu masalah utama yang dihadapi adalah ketidakseimbangan antara jumlah mahasiswa yang diterima dan jumlah mahasiswa yang berhasil lulus sesuai dengan waktu yang ditentukan. Ketidaktepatan waktu kelulusan dapat disebabkan oleh berbagai faktor, termasuk kemampuan akademik mahasiswa, kurangnya dukungan institusional, atau faktor eksternal lainnya. Hal ini mengarah pada meningkatnya jumlah mahasiswa yang terlambat lulus, yang tidak hanya berdampak pada

reputasi institusi, tetapi juga pada kualitas pendidikan dan pengelolaan sumber daya.

Salah satu masalah utama dalam prediksi kelulusan adalah kurangnya alat yang akurat untuk menganalisis berbagai faktor yang mempengaruhi keberhasilan mahasiswa. Banyak universitas belum memanfaatkan data yang ada secara optimal untuk menganalisis pola kelulusan. Data akademik, seperti Indeks Prestasi Kumulatif (IPK) dan data demografis lainnya, sering kali dikelola dalam sistem yang terpisah dan tidak dimanfaatkan secara maksimal untuk memprediksi kelulusan. Dengan menggunakan metode tradisional, prediksi kelulusan masih bergantung pada pendekatan konvensional yang kurang efektif dalam mengidentifikasi pola dan tren yang lebih kompleks.

Metode seperti K-Nearest Neighbors (KNN) dapat membantu mengatasi masalah ini dengan memanfaatkan data historis mahasiswa dan mencari pola atau hubungan yang mungkin tidak terlihat dengan jelas. Namun, penerapan metode ini memerlukan pemahaman yang mendalam tentang teknik data mining dan tantangan dalam mengolah data yang besar dan kompleks. Salah satu masalah yang sering muncul adalah kualitas data yang tidak konsisten atau tidak lengkap. Data yang hilang, data yang duplikat, atau data yang tidak relevan dapat mempengaruhi hasil prediksi secara signifikan. Oleh karena itu, institusi pendidikan perlu memastikan bahwa data yang digunakan dalam proses prediksi sudah melalui tahap pembersihan dan pra-pemrosesan yang tepat.

Selain itu, pemilihan metode yang tepat juga menjadi tantangan. Meskipun KNN memiliki keunggulan dalam memprediksi kelulusan berdasarkan kedekatan data, model ini membutuhkan perhitungan jarak antar data yang cukup intensif, yang dapat mengurangi efisiensi saat menangani dataset besar. Oleh karena itu, meskipun KNN menunjukkan potensi besar, implementasinya membutuhkan perhatian khusus terhadap pemilihan parameter yang tepat, seperti nilai K dan teknik evaluasi akurasi yang digunakan, untuk memastikan bahwa hasil prediksi dapat diandalkan dan efektif dalam membantu pengambilan keputusan.



Manfaat Praktis bagi Universitas dalam Merancang Kebijakan Pendidikan

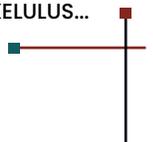
Prediksi kelulusan mahasiswa tidak hanya bermanfaat bagi mahasiswa itu sendiri, tetapi juga memiliki dampak yang signifikan dalam merancang kebijakan pendidikan di tingkat universitas. Sebagai contoh, melalui penerapan algoritma K-Nearest Neighbors (KNN) untuk memprediksi kelulusan mahasiswa, universitas dapat memperoleh data yang akurat mengenai potensi kelulusan mahasiswa berdasarkan indikator akademik yang telah ditetapkan, seperti IPK dan tahun masuk. Informasi ini memberikan gambaran yang lebih jelas kepada pihak universitas terkait dengan efisiensi akademik yang ada, dan dapat dijadikan dasar dalam perencanaan strategi pendidikan.

Dengan memahami tren kelulusan mahasiswa, universitas dapat merancang kebijakan yang lebih terfokus pada peningkatan kualitas pendidikan dan pengelolaan sumber daya. Misalnya, dengan mengetahui mahasiswa yang berisiko tidak lulus tepat waktu, universitas bisa menyediakan program bimbingan atau pelatihan tambahan untuk mendukung mahasiswa tersebut. Selain itu, kebijakan tentang alokasi dosen, kapasitas kelas, dan pengaturan kurikulum juga dapat lebih diprioritaskan berdasarkan hasil prediksi yang lebih terstruktur, untuk memastikan kelancaran proses pembelajaran.

Penerapan teknologi untuk prediksi kelulusan ini juga membantu universitas dalam perencanaan jangka panjang. Universitas dapat merencanakan anggaran dan infrastruktur dengan lebih efisien berdasarkan jumlah mahasiswa yang diprediksi akan lulus tepat waktu atau terlambat. Selain itu, kebijakan rekrutmen dosen dapat disesuaikan dengan kebutuhan akademik yang teridentifikasi melalui hasil prediksi ini. Kebijakan pengelolaan fakultas dan program studi juga dapat disesuaikan, memastikan mahasiswa menerima pendidikan yang lebih baik dan dapat menyelesaikan studi sesuai dengan waktu yang ditentukan.

Dari perspektif institusi, kebijakan pendidikan yang berbasis data ini akan memperkuat daya saing universitas, meningkatkan reputasi institusi di kalangan masyarakat, dan memungkinkan universitas beradaptasi dengan perkembangan dunia pendidikan yang semakin kompleks. Dengan

merancang kebijakan yang lebih tepat sasaran, universitas dapat mengoptimalkan sumber daya dan meningkatkan kualitas pendidikan secara keseluruhan, sambil menciptakan lingkungan akademik yang lebih kondusif bagi perkembangan mahasiswa.





BAB II

TEORI DASAR

Data Mining dan Penerapannya dalam Pendidikan

Data mining merupakan proses untuk menganalisis data skala besar guna menemukan pola, hubungan, atau informasi yang bermakna untuk mendukung pengambilan keputusan yang lebih efektif. Teknik ini digunakan di berbagai bidang, termasuk pendidikan, untuk memahami pola data akademik mahasiswa. Dengan memanfaatkan metode statistik, algoritma pembelajaran mesin, dan teknik analitik lainnya, data mining memungkinkan pengolahan data dalam skala besar guna memprediksi hasil yang diinginkan [1]. Analisis Data Mining adalah menganalisis dari pengolahan data latih yang ada untuk menemukan pola atau aturan pada sistem KNN [2]. Dalam konteks pendidikan, data mining telah diterapkan untuk memprediksi kinerja akademik, menganalisis kelulusan mahasiswa, dan mengukur kemampuan mahasiswa.

Dalam penelitian ini, data mining difokuskan untuk memprediksi kelulusan mahasiswa dengan memanfaatkan data akademik historis mereka. Hasil dari proses ini tidak hanya bertujuan untuk memberikan informasi prediktif, tetapi juga untuk membantu institusi pendidikan dalam

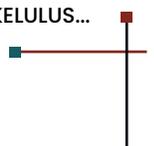
merancang kebijakan akademik yang lebih efektif. Mahasiswa sering disebut sebagai kelompok masyarakat yang memiliki ciri intelektualitas yang lebih luas dibandingkan dengan kelompok seusia mereka yang bukan mahasiswa ataupun kelompok usia lain yang dibawah mereka. Dengan intelektualitasnya, mahasiswa akan mampu menghadapi dan mencari permasalahan secara sistematis yang nantinya diterapkan dalam kehidupan sehari-hari agar bisa bersaing dalam dunia kerja[3]. Perguruan tinggi dituntut untuk menyelenggarakan pendidikan yang berkualitas bagi mahasiswa sehingga menghasilkan sumber daya manusia yang berilmu, kreatif dan bersaing. Dalam sistem pendidikan, mahasiswa adalah aset penting bagi sebuah institusi pendidikan dan untuk itu perlu diperhatikan tingkat kelulusan mahasiswa tepat pada waktunya[4]. Presentasi naik turunnya kemampuan mahasiswa untuk menyelesaikan studi tepat waktu merupakan salah satu elemen penilaian akreditasi universitas[5]. Oleh karena itu, perlu adanya pemantuan dan evaluasi terhadap kecenderungan kelulusan mahasiswa lulus tepat waktu atau tidak.

Teknik Data Mining

Teknik data mining meliputi beberapa pendekatan utama, seperti:

1. **Klasifikasi:** Teknik supervised learning yang digunakan untuk memprediksi kategori atau kelas tertentu berdasarkan data historis.
2. **Regresi:** Metode untuk memprediksi nilai numerik berdasarkan hubungan antar variabel.
3. **Clustering:** Teknik unsupervised learning yang digunakan untuk mengelompokkan data berdasarkan kemiripan.
4. **Analisis Asosiatif:** Teknik untuk menemukan hubungan antar atribut dalam dataset.

Penelitian ini menggunakan pendekatan klasifikasi, yang memungkinkan identifikasi pola dalam data akademik untuk memprediksi kelulusan mahasiswa. Penelitian sebelumnya menunjukkan bahwa algoritma seperti K-Nearest Neighbor (KNN) sering digunakan dalam pengolahan data pendidikan karena efisiensinya dalam menangani data akademik([6]. KNN efektif untuk data dengan pola non-linear.



Implementasi Data Mining dalam Pendidikan

Implementasi data mining dalam pendidikan melibatkan tiga tahapan utama:

1. Pengumpulan Data: Data akademik seperti NIM, IPK, dan atribut lainnya dikumpulkan untuk dianalisis.
2. Pra-Pemrosesan Data: Proses ini mencakup pembersihan data (data cleansing), normalisasi, dan penanganan nilai kosong untuk memastikan kualitas dataset yang optimal. Penelitian oleh [7] menunjukkan bahwa pra-pemrosesan data memiliki dampak signifikan terhadap akurasi model prediksi.
3. Analisis dan Model Prediktif: Data yang telah diproses dianalisis menggunakan algoritma klasifikasi seperti KNN untuk memprediksi kelulusan mahasiswa.

Hasil implementasi data mining tidak hanya mampu memprediksi kelulusan mahasiswa tetapi juga memberikan wawasan bagi institusi untuk meningkatkan kualitas pendidikan. Selain itu, teknik ini membantu mengidentifikasi mahasiswa yang berisiko tidak lulus tepat waktu, sehingga institusi dapat mengambil langkah preventif [8].

Dalam konteks pendidikan, data mining diterapkan untuk menganalisis data akademik mahasiswa, yang kemudian digunakan untuk membuat prediksi dan meningkatkan kualitas sistem pendidikan. Penerapan data mining dalam pendidikan memungkinkan institusi untuk memahami berbagai faktor yang mempengaruhi kinerja akademik mahasiswa, yang selanjutnya dapat dijadikan dasar dalam merancang kebijakan pendidikan yang lebih efisien dan terarah.

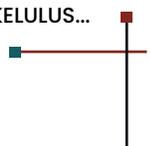
Secara khusus, dalam penerapan data mining untuk prediksi kelulusan mahasiswa, teknik ini memungkinkan universitas untuk memprediksi kemungkinan kelulusan mahasiswa tepat waktu berdasarkan data historis yang tersedia. Data mining dapat mengidentifikasi pola-pola tertentu, seperti hubungan antara Indeks Prestasi Kumulatif (IPK), tahun masuk, atau atribut lainnya dengan status kelulusan mahasiswa. Proses ini sangat penting karena dapat membantu universitas dalam membuat kebijakan yang

lebih tepat sasaran, misalnya dalam menentukan siapa saja mahasiswa yang membutuhkan perhatian khusus agar bisa lulus tepat waktu.

Penerapan data mining dalam pendidikan tidak hanya terbatas pada prediksi kelulusan mahasiswa, tetapi juga mencakup berbagai aspek lain seperti evaluasi kualitas pengajaran, pengelolaan sumber daya manusia (dosen), serta pengelolaan fasilitas pendidikan. Dengan menganalisis data yang ada, seperti nilai ujian, tingkat kehadiran, dan faktor demografis mahasiswa, institusi pendidikan dapat memperoleh wawasan yang lebih mendalam tentang faktor-faktor yang berpengaruh terhadap hasil belajar mahasiswa. Pengetahuan ini memungkinkan universitas untuk merancang strategi yang lebih efektif dalam meningkatkan kualitas pembelajaran dan meningkatkan tingkat kelulusan.

Salah satu teknik yang paling sering digunakan dalam data mining adalah klasifikasi. Klasifikasi adalah proses untuk mengelompokkan data ke dalam kategori tertentu berdasarkan atribut atau karakteristik yang dimilikinya. Dalam konteks pendidikan, klasifikasi dapat digunakan untuk mengelompokkan mahasiswa ke dalam kategori yang berbeda, misalnya mahasiswa yang berpotensi lulus tepat waktu dan mahasiswa yang berpotensi terlambat lulus. Metode klasifikasi yang paling umum digunakan adalah algoritma pembelajaran mesin, seperti K-Nearest Neighbors (KNN), Naïve Bayes, dan Decision Trees, yang memiliki keunggulan dalam memproses data dalam jumlah besar dan menghasilkan model prediksi yang akurat.

Di antara berbagai metode klasifikasi, K-Nearest Neighbors (KNN) sering dipilih untuk aplikasi pendidikan, karena kemampuannya dalam menangani data yang tidak terstruktur dan memiliki pola non-linear. KNN bekerja dengan cara membandingkan data yang akan diprediksi dengan data latih yang telah tersedia, kemudian menentukan kelas atau kategori berdasarkan mayoritas kelas dari tetangga terdekat. Kelebihan KNN adalah kesederhanaannya, fleksibilitasnya, dan kemampuannya untuk bekerja dengan data yang relatif kecil. Oleh karena itu, KNN banyak digunakan dalam penelitian pendidikan untuk memprediksi kelulusan mahasiswa, analisis perilaku mahasiswa, serta dalam pengembangan sistem rekomendasi untuk perbaikan sistem pembelajaran.



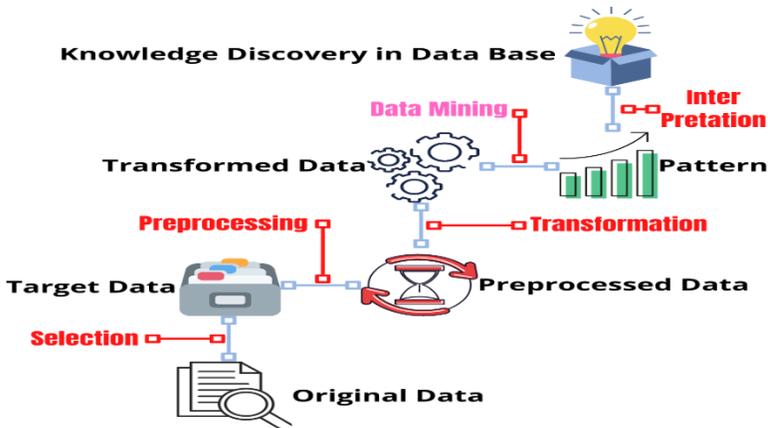
Selain itu, penerapan data mining juga memungkinkan pengembangan sistem yang dapat mendeteksi mahasiswa yang berisiko gagal atau terlambat lulus. Dengan menggunakan algoritma yang tepat, data mining dapat mengidentifikasi mahasiswa yang memiliki nilai rendah, tingkat kehadiran buruk, atau faktor-faktor lain yang dapat mempengaruhi kelulusan mereka. Hal ini memberikan kesempatan bagi universitas untuk melakukan intervensi lebih awal, seperti memberikan bimbingan atau pelatihan tambahan, sehingga mahasiswa memiliki peluang lebih besar untuk lulus tepat waktu.

Secara keseluruhan, penerapan data mining dalam pendidikan menawarkan potensi besar untuk meningkatkan efisiensi pengelolaan akademik, mengoptimalkan penggunaan sumber daya, dan meningkatkan kualitas pendidikan. Dengan memanfaatkan teknik-teknik data mining, universitas dapat merancang kebijakan yang lebih berbasis data, mengidentifikasi masalah lebih dini, serta meningkatkan pengalaman belajar mahasiswa. Dalam konteks penelitian ini, penggunaan KNN sebagai metode untuk memprediksi kelulusan mahasiswa di Universitas Labuhanbatu dapat menjadi contoh bagaimana teknologi dapat digunakan untuk mendukung pengambilan keputusan yang lebih baik dan lebih terinformasi dalam bidang pendidikan.

Knowledge Discovery in Database (KDD)

Proses Knowledge Discovery in Database (KDD) adalah proses sistematis untuk menemukan informasi baru dari data dalam jumlah besar dengan pendekatan bertahap. Knowledge Discovery in Database (KDD) adalah proses sistematis untuk menemukan informasi atau pola yang bermanfaat dari data dalam jumlah besar. KDD mencakup serangkaian langkah terstruktur yang dimulai dari pengumpulan data mentah hingga interpretasi hasil yang dapat digunakan untuk pengambilan keputusan. Fokus KDD adalah mengekstraksi pengetahuan dari data dengan cara yang efisien, bermakna, dan relevan. Tujuan KDD adalah untuk membantu pengguna dalam mengungkap pola tersembunyi, tren, atau hubungan dalam data yang tidak langsung terlihat. Pengetahuan ini digunakan untuk mendukung pengambilan keputusan yang lebih baik dalam berbagai domain, termasuk

pendidikan, kesehatan, bisnis, dan lain-lain. Berikut tahapan utama KDD secara lebih rinci:



Gambar 1. Tahapan Knowledge Discovery in Database (KDD)

1. Pengumpulan Data: Tahap ini mencakup pengumpulan data dari berbagai sumber. Data yang dikumpulkan harus relevan dengan tujuan penelitian untuk memastikan analisis yang tepat.
2. Pra-pemrosesan Data: Transformasi Data, data yang telah diproses diubah menjadi format yang sesuai untuk digunakan dalam analisis lanjutan. Teknik encoding dapat diterapkan untuk data kategorikal (contoh: one-hot encoding), sementara fitur baru dapat dihasilkan melalui rekayasa fitur (feature engineering).
3. Data Mining: Proses inti KDD di mana algoritma tertentu, seperti KNN, digunakan untuk menemukan pola, hubungan, atau klasifikasi berdasarkan data yang telah diproses. Analisis dapat berupa klasifikasi, clustering, atau prediksi, tergantung pada tujuan penelitian.
4. Evaluasi dan Interpretasi: Hasil yang diperoleh dari proses data mining dievaluasi menggunakan metrik seperti akurasi, precision, recall, dan F1-score.

Hasil ini kemudian diinterpretasikan untuk memastikan bahwa temuan sesuai dengan tujuan penelitian dan memberikan wawasan yang dapat diterapkan dalam konteks nyata.

Klasifikasi dalam Machine Learning

Klasifikasi merupakan salah satu teknik supervised learning dalam data mining yang digunakan untuk mengelompokkan data ke dalam kategori atau kelas tertentu berdasarkan pola yang ditemukan dalam data historis. Teknik ini bekerja dengan memanfaatkan data latih yang telah diberi label untuk membangun model dan kemudian menggunakan data uji untuk mengukur akurasi prediksi model tersebut. Algoritma klasifikasi memiliki aplikasi luas, terutama dalam pendidikan, untuk memprediksi hasil kelulusan mahasiswa berdasarkan data akademik historis [9].

Dalam konteks prediksi kelulusan mahasiswa di Universitas Labuhanbatu menggunakan metode K-Nearest Neighbors (KNN), klasifikasi memungkinkan untuk memprediksi apakah seorang mahasiswa akan lulus tepat waktu atau terlambat berdasarkan atribut-atribut yang tersedia, seperti Indeks Prestasi Kumulatif (IPK), tahun masuk, dan data demografis lainnya.

Klasifikasi bekerja dengan menggunakan algoritma tertentu untuk memetakan data ke dalam kelas yang sesuai. Model klasifikasi yang telah dilatih menggunakan data yang berlabel akan memetakan setiap data uji ke dalam salah satu dari kelas yang ada, misalnya “Tepat Waktu” atau “Terlambat” dalam kasus kelulusan mahasiswa. Proses ini dilakukan dengan cara menghitung kesamaan antara data uji dengan data pelatihan yang telah ada, menggunakan metrik jarak seperti Euclidean Distance pada KNN. Teknik klasifikasi juga dapat digunakan untuk memetakan data ke dalam beberapa kelas, yang dikenal dengan klasifikasi multikelas, di mana model akan memprediksi lebih dari dua kategori berdasarkan data input.

Salah satu keuntungan utama dari klasifikasi dalam machine learning adalah kemampuannya untuk memberikan prediksi yang dapat digunakan dalam pengambilan keputusan berbasis data. Sebagai contoh, dalam penelitian ini, metode KNN digunakan untuk memprediksi kelulusan mahasiswa. Dengan menganalisis data akademik historis dan atribut lainnya, model klasifikasi dapat memberikan gambaran yang jelas tentang kemungkinan kelulusan mahasiswa, yang membantu universitas dalam merencanakan

kebijakan pendidikan lebih baik, termasuk intervensi bagi mahasiswa yang berisiko terlambat lulus.

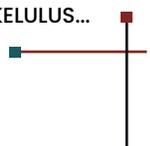
Ada berbagai jenis algoritma klasifikasi yang dapat digunakan dalam machine learning, di antaranya adalah K-Nearest Neighbors (KNN), Decision Trees, Naive Bayes, dan Support Vector Machines (SVM). Setiap algoritma memiliki karakteristik dan kelebihan masing-masing tergantung pada jenis data yang digunakan dan tujuan analisis. KNN, sebagai salah satu algoritma yang digunakan dalam penelitian ini, bekerja dengan cara mencari sejumlah tetangga terdekat dari data uji berdasarkan jarak tertentu. Setelah itu, kelas yang paling banyak ditemukan di antara tetangga terdekat akan dipilih sebagai hasil klasifikasi. Kelebihan KNN adalah kesederhanaannya, kemampuannya dalam menangani data non-linear, dan efektivitasnya dalam dataset yang relatif kecil dengan struktur data yang jelas.

Namun, KNN juga memiliki kelemahan, terutama terkait dengan waktu komputasi yang tinggi jika jumlah data pelatihan sangat besar, karena harus menghitung jarak antara data uji dengan semua data pelatihan. Untuk mengatasi hal ini, teknik seperti normalisasi data dan pemilihan nilai K yang tepat dapat digunakan untuk meningkatkan efisiensi dan akurasi model KNN. Oleh karena itu, pemilihan algoritma klasifikasi yang tepat sangat bergantung pada karakteristik data yang dimiliki dan kebutuhan spesifik dalam penelitian.

Penerapan klasifikasi dalam pendidikan, terutama untuk memprediksi kelulusan mahasiswa, memberikan manfaat yang signifikan dalam perencanaan pendidikan yang lebih terstruktur. Dengan mengetahui kemungkinan kelulusan mahasiswa, pihak universitas dapat merencanakan strategi untuk meningkatkan efisiensi sistem akademik, serta memberikan intervensi yang tepat untuk membantu mahasiswa yang berisiko gagal atau terlambat lulus.

Algoritma K-Nearest Neighbor (KNN)

K-Nearest Neighbor (KNN) adalah algoritma klasifikasi yang sederhana namun efektif. Algoritma ini bekerja dengan menghitung jarak antara data baru dan data latih, kemudian menentukan kelas berdasarkan mayoritas tetangga terdekatnya (k nearest neighbors). Keunggulan utama KNN adalah



kemampuannya untuk menangani data dengan pola non-linear, serta memberikan akurasi tinggi pada dataset tertentu [10]. Namun, algoritma ini memerlukan waktu komputasi yang lebih tinggi untuk dataset besar karena harus menghitung jarak ke semua titik data dalam setiap prediksi [11]. Kelebihan utama dari KNN adalah kesederhanaannya dan kemampuannya untuk menangani data yang tidak memiliki distribusi yang jelas atau data dengan pola non-linear. Proses utama dalam KNN adalah mengukur jarak antara data baru dengan data yang ada dalam dataset menggunakan metrik tertentu, seperti jarak Euclidean, Manhattan, atau Minkowski, untuk menentukan kelas data yang relevan.

$$d(p, q) = \sqrt{\sum_{i=1}^n (p_i - q_i)^2}$$

Gambar Rumus jarak Euclidean

Dalam konteks prediksi kelulusan mahasiswa, KNN digunakan untuk memprediksi status kelulusan berdasarkan variabel yang ada, seperti IPK, tahun masuk, dan jenis kelamin mahasiswa. Misalnya, dalam sebuah dataset yang berisi informasi tentang mahasiswa dan status kelulusan mereka (tepat waktu atau terlambat), KNN akan mencari tetangga terdekat dari data baru dan mengklasifikasikannya berdasarkan mayoritas kelas dari tetangga tersebut. Pemilihan nilai K yang tepat sangat penting dalam KNN, karena K adalah jumlah tetangga terdekat yang akan dihitung dalam proses klasifikasi. Nilai K yang terlalu kecil dapat menyebabkan model terlalu sensitif terhadap noise dalam data, sedangkan nilai K yang terlalu besar dapat membuat model menjadi lebih stabil namun kurang memperhatikan pola lokal dalam data.

Salah satu keunggulan dari algoritma KNN adalah bahwa ia tidak memerlukan fase pelatihan eksplisit seperti algoritma lain, misalnya Decision Trees atau Support Vector Machines. Hal ini menjadikan KNN lebih fleksibel dan mudah diterapkan dalam berbagai situasi. Namun, meskipun sederhana, KNN juga memiliki kelemahan, di antaranya adalah

kebutuhan akan memori yang besar untuk dataset yang besar, serta waktu komputasi yang lebih lama karena setiap kali prediksi dilakukan, algoritma harus menghitung jarak terhadap seluruh dataset pelatihan.

Dalam penelitian ini, penerapan algoritma KNN untuk memprediksi kelulusan mahasiswa sangat relevan, karena memungkinkan universitas untuk melakukan klasifikasi mahasiswa yang berisiko terlambat lulus atau yang berpotensi lulus tepat waktu, berdasarkan data akademik mereka. Hal ini bisa membantu universitas untuk merancang kebijakan akademik yang lebih efektif, seperti pengalokasian dosen yang lebih tepat, pengelolaan kapasitas kelas, serta program bimbingan bagi mahasiswa yang membutuhkan bantuan tambahan.

Microsoft Excel

Microsoft Excel adalah salah satu perangkat lunak yang paling banyak digunakan di dunia dalam bidang pengolahan data. Sebagai alat bantu yang sederhana namun kuat, Excel telah lama digunakan dalam berbagai jenis analisis, termasuk di bidang pendidikan, untuk memproses data dan melakukan analisis statistik. Dalam konteks prediksi kelulusan mahasiswa, Excel menawarkan sejumlah fitur yang memungkinkan pengguna untuk melakukan manipulasi data secara efisien, tanpa memerlukan keterampilan pemrograman yang mendalam. Penggunaannya sangat relevan dalam penelitian ini, terutama pada tahap pra-pemrosesan data yang penting untuk membangun model prediksi yang akurat.

Dalam penelitian mengenai prediksi kelulusan mahasiswa, Excel dapat digunakan untuk mengimpor data yang berasal dari berbagai sumber, seperti sistem akademik mahasiswa, untuk kemudian dibersihkan dan dianalisis lebih lanjut. Excel mendukung berbagai operasi dasar yang diperlukan dalam proses pembersihan data, seperti penghapusan data duplikat, pengisian nilai yang hilang, serta penyusunan data dalam format yang mudah dianalisis. Dengan menggunakan berbagai formula dan fungsi statistik, seperti rata-rata, median, atau standar deviasi, Excel memungkinkan peneliti untuk memperoleh gambaran umum mengenai distribusi data akademik mahasiswa yang digunakan untuk prediksi kelulusan.



Gambar 2. Aplikasi *Excel*

Selain itu, Excel memungkinkan visualisasi data dengan menggunakan berbagai jenis grafik, seperti histogram, pie chart, atau scatter plot. Visualisasi ini penting untuk mendeteksi pola-pola atau hubungan antar variabel yang ada, seperti hubungan antara Indeks Prestasi Kumulatif (IPK) dan kelulusan tepat waktu. Fitur pivot table dalam Excel juga sangat berguna untuk mengelompokkan dan merangkum data berdasarkan kriteria tertentu, seperti jurusan atau tahun masuk mahasiswa. Hal ini memungkinkan peneliti untuk melihat tren kelulusan dengan cara yang lebih terstruktur dan membantu dalam membuat keputusan yang lebih tepat.

Walaupun Excel tidak dilengkapi dengan algoritma machine learning yang kompleks seperti perangkat lunak analitik lainnya, kelebihanannya terletak pada kemudahan penggunaannya dan kemampuan untuk melakukan analisis dasar dengan cepat. Excel sangat berguna pada tahap awal pengolahan data dan eksperimen prediksi, sebelum akhirnya menggunakan perangkat lunak analitik yang lebih canggih, seperti RapidMiner atau Python, untuk membangun model prediksi yang lebih kompleks. Dengan demikian, Excel tetap menjadi alat bantu yang esensial dalam mempersiapkan dan mengelola data yang diperlukan untuk penelitian prediksi kelulusan mahasiswa, memberikan fondasi yang solid sebelum langkah-langkah analitis lebih lanjut diambil.

Rapidminer

RapidMiner merupakan salah satu perangkat lunak yang banyak digunakan dalam bidang data mining, yang menyediakan antarmuka grafis untuk memudahkan pengguna dalam menganalisis data tanpa memerlukan pemrograman yang kompleks. Dalam konteks penelitian ini, RapidMiner

berfungsi sebagai alat bantu untuk menerapkan algoritma K-Nearest Neighbors (KNN) dalam memprediksi kelulusan mahasiswa. Keunggulan utama dari RapidMiner adalah kemampuannya untuk mengintegrasikan berbagai teknik analisis data seperti klasifikasi, clustering, dan regresi, yang memungkinkan pengguna untuk melakukan eksplorasi data secara mendalam dengan cara yang efisien.

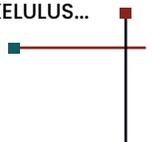


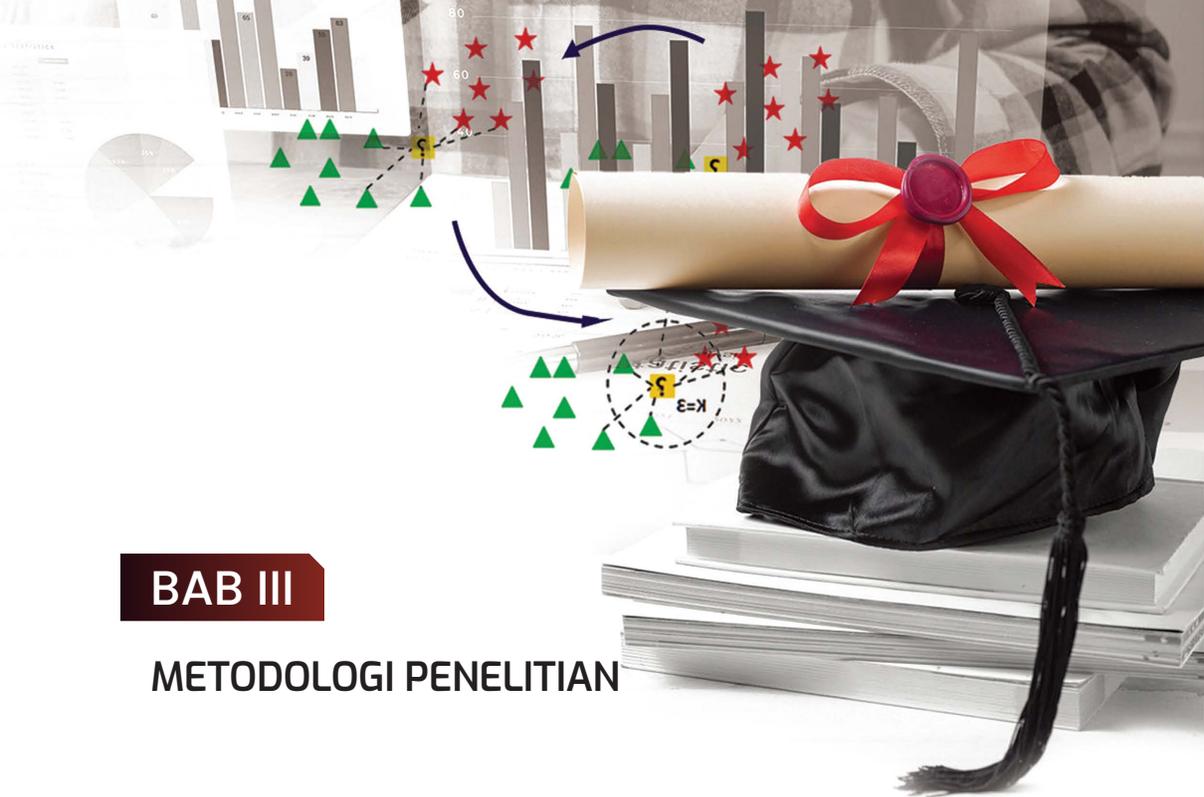
Gambar 2. Aplikasi *RapidMiner*

RapidMiner mendukung berbagai fungsi yang sangat penting dalam proses data mining, mulai dari pengumpulan data, pra-pemrosesan, hingga penerapan model prediktif. Salah satu fitur unggulannya adalah kemudahan dalam melakukan pembersihan data, normalisasi, serta pengisian nilai yang hilang (missing values). Proses ini sangat krusial dalam memastikan bahwa data yang digunakan dalam model prediksi adalah data yang bersih dan relevan, sehingga meningkatkan akurasi prediksi yang dihasilkan oleh algoritma KNN. Selain itu, RapidMiner juga memungkinkan visualisasi data yang mempermudah pengguna dalam mengidentifikasi pola-pola penting dalam data yang dianalisis, seperti hubungan antara IPK mahasiswa dengan status kelulusan mereka.

Dalam penelitian ini, RapidMiner digunakan untuk mempermudah penerapan KNN dalam memprediksi kelulusan mahasiswa di Universitas Labuhanbatu. Dengan menggunakan antarmuka grafis, pengguna dapat memilih dan menggabungkan berbagai modul untuk menyusun alur kerja (workflow) yang sesuai dengan tujuan analisis. Misalnya, modul untuk memasukkan data, modul untuk pemrosesan data, serta modul untuk aplikasi algoritma KNN dapat digabungkan dalam satu workflow yang terintegrasi dengan mudah. Penggunaan RapidMiner sebagai alat bantu memberikan keuntungan besar bagi peneliti yang mungkin tidak memiliki latar belakang pemrograman yang mendalam, namun tetap ingin melakukan analisis data yang kompleks dengan efisien dan akurat.

RapidMiner juga mendukung pengujian model yang telah dibangun, dengan menyediakan berbagai metrik evaluasi, seperti akurasi, precision, recall, dan F1-score, yang memungkinkan peneliti untuk menilai performa dari model KNN yang diterapkan. Dengan demikian, RapidMiner tidak hanya memfasilitasi proses pembuatan model prediktif tetapi juga memberikan gambaran yang jelas tentang kualitas dan efektivitas model tersebut dalam memprediksi kelulusan mahasiswa. Ini menjadi salah satu alasan mengapa RapidMiner dipilih sebagai alat bantu dalam penelitian ini, karena kemudahan penggunaan serta kelengkapan fitur yang ditawarkannya dalam mendukung analisis data berbasis machine learning.





BAB III

METODOLOGI PENELITIAN

Desain Penelitian

Desain penelitian yang digunakan dalam studi ini bertujuan untuk memprediksi kelulusan mahasiswa di Universitas Labuhanbatu Fakultas Sains dan Teknologi Prodi Sistem Informasi, dengan menggunakan algoritma K-Nearest Neighbors (KNN). Pendekatan yang diambil berfokus pada penerapan metode data mining, khususnya KNN, yang merupakan teknik klasifikasi dalam machine learning. Penelitian ini dirancang untuk menganalisis faktor-faktor yang mempengaruhi kelulusan mahasiswa, seperti Indeks Prestasi Kumulatif (IPK), jenis kelamin, dan tahun masuk, serta bagaimana hubungan antara faktor-faktor tersebut dapat memprediksi kelulusan tepat waktu atau terlambat.

Dalam penelitian ini, desain penelitian menggunakan data historis mahasiswa yang mencakup atribut-atribut penting yang dianggap relevan untuk memprediksi kelulusan, seperti IPK dan tahun masuk. Data yang digunakan bersumber dari sistem akademik Universitas Labuhanbatu, yang telah diproses untuk memastikan kualitas dan kelengkapan data sebelum diterapkan dalam model. Data tersebut kemudian dibagi menjadi dua

kelompok, yaitu data latih dan data uji. Data latih digunakan untuk melatih model KNN, sedangkan data uji digunakan untuk menguji akurasi dan performa model dalam memprediksi kelulusan mahasiswa.

Pendekatan yang digunakan dalam penelitian ini adalah pendekatan berbasis algoritma KNN, yang bekerja dengan prinsip dasar mencari tetangga terdekat untuk menentukan kelas suatu data. Dalam konteks ini, kelas yang dimaksud adalah status kelulusan mahasiswa, yang dibagi menjadi dua kategori: “Tepat Waktu” dan “Terlambat.” KNN menghitung jarak antara data uji dengan data latih menggunakan metrik seperti Euclidean Distance. Setelah jarak dihitung, data uji akan diklasifikasikan berdasarkan mayoritas kelas dari tetangga terdekatnya. Pemilihan parameter K (jumlah tetangga terdekat) sangat krusial dalam algoritma ini, dan biasanya dipilih melalui metode cross-validation untuk memastikan akurasi model yang optimal.

Penelitian ini menggunakan KNN karena kemampuannya yang baik dalam menangani data yang tidak linier dan mengklasifikasikan data ke dalam dua atau lebih kelas berdasarkan atribut yang ada. Selain itu, KNN tidak memerlukan asumsi distribusi data, yang menjadikannya metode yang fleksibel untuk digunakan dalam berbagai konteks, termasuk dalam prediksi kelulusan mahasiswa.

Dengan menggunakan KNN, penelitian ini bertujuan untuk memberikan wawasan yang lebih baik kepada pihak universitas dalam merancang kebijakan pendidikan yang lebih berbasis data, seperti alokasi sumber daya, pengaturan kurikulum, serta penyusunan strategi untuk meningkatkan kelulusan tepat waktu mahasiswa.

Pengumpulan Data

Pengumpulan data merupakan langkah penting dalam penelitian ini untuk memastikan kualitas dan akurasi prediksi kelulusan mahasiswa menggunakan metode K -Nearest Neighbors (KNN). Sumber data yang digunakan dalam penelitian ini terdiri dari data akademik mahasiswa, data demografis, dan informasi terkait kelulusan mahasiswa yang telah ada. Data akademik, seperti Indeks Prestasi Kumulatif (IPK), tahun masuk, dan jumlah SKS

yang diambil, merupakan indikator utama dalam memprediksi kelulusan tepat waktu atau terlambat. Selain itu, data demografis yang mencakup jenis kelamin, usia, dan faktor sosial ekonomi juga turut dipertimbangkan sebagai variabel yang dapat mempengaruhi hasil kelulusan mahasiswa. Penggunaan data yang luas dan komprehensif ini diharapkan dapat memberikan hasil prediksi yang lebih akurat dan relevan dalam konteks institusi pendidikan.

Metode pengumpulan data yang digunakan dalam penelitian ini melibatkan dua pendekatan utama, yaitu observasi dan kuesioner. Observasi dilakukan dengan mengamati data yang ada pada sistem akademik Universitas Labuhanbatu, yang mencakup data mahasiswa yang telah terdaftar dan tercatat dalam sistem akademik universitas. Pengamatan ini meliputi pemantauan data IPK, data kehadiran, serta data kelulusan mahasiswa yang telah selesai atau yang sedang berjalan. Selain itu, kuesioner juga disebarluaskan kepada mahasiswa untuk mengumpulkan informasi yang lebih mendalam mengenai latar belakang demografis, motivasi belajar, dan faktor-faktor eksternal yang dapat mempengaruhi keberhasilan akademik mereka. Kuesioner ini dirancang untuk mengidentifikasi variabel-variabel lain yang tidak tercakup dalam data akademik, seperti tingkat dukungan keluarga, kondisi sosial ekonomi, dan akses ke fasilitas pembelajaran.

Kriteria pemilihan data yang relevan untuk penelitian ini sangat penting untuk memastikan bahwa data yang digunakan dapat memberikan gambaran yang representatif mengenai kelulusan mahasiswa. Kriteria pemilihan ini mencakup mahasiswa yang terdaftar di Program Studi Sistem Informasi Fakultas Sains dan Teknologi Universitas Labuhanbatu, dengan rentang waktu studi yang mencakup mahasiswa yang baru masuk hingga mereka yang telah berada pada semester akhir. Hanya data mahasiswa yang memiliki rekam jejak akademik lengkap yang akan dipilih untuk memastikan bahwa model prediksi yang dibangun dapat bekerja dengan optimal. Selain itu, mahasiswa yang terlibat dalam penelitian ini juga harus memiliki data kelulusan yang tercatat secara akurat, agar hasil analisis yang dilakukan dapat memprediksi kelulusan dengan tingkat keandalan yang tinggi. Data yang dipilih juga harus mencakup variabel yang cukup untuk

mewakili faktor-faktor yang mempengaruhi kelulusan, seperti IPK, tahun masuk, dan demografi mahasiswa.

Dengan menggunakan data yang terpilih dan relevan ini, penelitian ini bertujuan untuk menghasilkan model prediksi yang dapat diandalkan, yang nantinya dapat digunakan oleh Universitas Labuhanbatu untuk merancang kebijakan pendidikan yang lebih tepat sasaran. Informasi yang diperoleh melalui pengumpulan data yang sistematis ini akan memberikan wawasan penting bagi pihak universitas dalam mengelola dan meningkatkan tingkat kelulusan mahasiswa.

Proses Knowledge Discovery in Database (KDD)

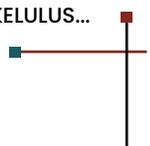
Penjelasan mendalam tentang langkah-langkah analisis yang dilakukan untuk efektivitas algoritma *K-Nearest Neighbors* (KNN) dalam memprediksi kelulusan mahasiswa Universitas Labuhanbatu. Fokus utama adalah pada pengolahan dataset, implementasi algoritma, evaluasi model, dan pembuktian perhitungan matematis masing-masing metode. Hasil dari analisis ini bertujuan memberikan wawasan berbasis data untuk meningkatkan tingkat kelulusan mahasiswa tepat waktu.

Dalam penelitian ini, analisis dilakukan secara sistematis, mulai dari pra-pemrosesan data, implementasi algoritma, hingga evaluasi model menggunakan metrik kinerja seperti *akurasi*, *precision*, *recall*, dan *F1-score*. Pembuktian matematis disajikan untuk mendukung perhitungan setiap langkah, memastikan keabsahan hasil.

Dataset yang digunakan dalam penelitian ini terdiri dari atribut-atribut berikut:

1. Nama Mahasiswa: Nama lengkap mahasiswa disamarkan.
2. Tahun Masuk: Tahun pertama mahasiswa mulai studi.
3. Jenis Kelamin : Menunjukkan *Gender* Mahasiswa
4. IPK: Indeks Prestasi Kumulatif yang mencerminkan performa akademik.

Dataset ini mencakup mahasiswa dari dua tahun masuk utama, yaitu 2019 dan 2020. Data ini diolah untuk menghasilkan model prediksi yang



dapat menentukan kemungkinan kelulusan mahasiswa berdasarkan IPK dan Tahun Masuk.

Nama	Jenis Kelamin	Tahun Masuk	Ipk	Status
MHS1	Laki-Laki	2019	Dengan Pujian	Tepat Waktu
MHS2	Perempuan	2019	Sangat Memuaskan	Tepat Waktu
MHS3	Perempuan	2019	Dengan Pujian	Tepat Waktu
MHS4	Laki-Laki	2019	Sangat Memuaskan	Tepat Waktu
MHS5	Perempuan	2019	Dengan Pujian	Tepat Waktu
MHS6	Laki-Laki	2019	Sangat Memuaskan	Tepat Waktu
MHS7	Perempuan	2019	Dengan Pujian	Tepat Waktu
MHS8	Laki-Laki	2019	Sangat Memuaskan	Tepat Waktu
MHS9	Laki-Laki	2019	Sangat Memuaskan	Tepat Waktu
MHS10	Perempuan	2019	Dengan Pujian	Tepat Waktu
MHS11	Laki-Laki	2019	Dengan Pujian	Tepat Waktu
MHS12	Perempuan	2019	Sangat Memuaskan	Tepat Waktu
MHS13	Perempuan	2019	Dengan Pujian	Tepat Waktu
MHS14	Perempuan	2019	Dengan Pujian	Tepat Waktu
MHS15	Perempuan	2019	Dengan Pujian	Tepat Waktu
MHS16	Laki-Laki	2019	Dengan Pujian	Terlambat
MHS17	Perempuan	2019	Dengan Pujian	Terlambat
MHS18	Perempuan	2019	Sangat Memuaskan	Tepat Waktu
MHS19	Laki-Laki	2019	Dengan Pujian	Tepat Waktu
MHS20	Perempuan	2019	Sangat Memuaskan	Tepat Waktu
MHS21	Perempuan	2019	Sangat Memuaskan	Tepat Waktu
MHS22	Laki-Laki	2019	Sangat Memuaskan	Tepat Waktu

Nama	Jenis Kelamin	Tahun Masuk	Ipk	Status
MHS23	Perempuan	2019	Dengan Pujian	Tepat Waktu
MHS24	Laki-Laki	2019	Sangat Memuaskan	Tepat Waktu
MHS25	Perempuan	2019	Sangat Memuaskan	Tepat Waktu
MHS26	Perempuan	2019	Sangat Memuaskan	Tepat Waktu
MHS27	Perempuan	2019	Sangat Memuaskan	Tepat Waktu
MHS28	Perempuan	2020	Sangat Memuaskan	Terlambat
MHS29	Perempuan	2019	Dengan Pujian	Tepat Waktu
MHS30	Laki-Laki	2019	Sangat Memuaskan	Tepat Waktu
MHS31	Perempuan	2019	Dengan Pujian	Tepat Waktu
MHS32	Perempuan	2019	Sangat Memuaskan	Tepat Waktu
MHS33	Perempuan	2019	Dengan Pujian	Tepat Waktu
MHS34	Laki-Laki	2019	Sangat Memuaskan	Tepat Waktu
MHS35	Laki-Laki	2019	Sangat Memuaskan	Tepat Waktu
MHS36	Laki-Laki	2019	Sangat Memuaskan	Tepat Waktu
MHS37	Laki-Laki	2019	Sangat Memuaskan	Tepat Waktu
MHS38	Perempuan	2019	Dengan Pujian	Tepat Waktu
MHS39	Perempuan	2019	Sangat Memuaskan	Tepat Waktu
MHS40	Laki-Laki	2019	Dengan Pujian	Tepat Waktu
MHS41	Laki-Laki	2019	Sangat Memuaskan	Tepat Waktu
MHS42	Laki-Laki	2020	Dengan Pujian	Tepat Waktu
MHS43	Laki-Laki	2020	Dengan Pujian	Tepat Waktu

Nama	Jenis Kelamin	Tahun Masuk	Ipk	Status
MHS44	Perempuan	2019	Sangat Memuaskan	Terlambat
MHS45	Perempuan	2020	Dengan Pujian	Terlambat
MHS46	Perempuan	2020	Dengan Pujian	Tepat Waktu
MHS47	Perempuan	2020	Dengan Pujian	Tepat Waktu
MHS48	Laki-Laki	2020	Dengan Pujian	Tepat Waktu
MHS49	Perempuan	2020	Dengan Pujian	Tepat Waktu
MHS50	Perempuan	2020	Dengan Pujian	Tepat Waktu
MHS51	Perempuan	2020	Dengan Pujian	Tepat Waktu
MHS52	Perempuan	2020	Dengan Pujian	Tepat Waktu
MHS53	Perempuan	2020	Dengan Pujian	Tepat Waktu
MHS54	Laki-Laki	2020	Dengan Pujian	Tepat Waktu
MHS55	Laki-Laki	2020	Dengan Pujian	Tepat Waktu
MHS56	Laki-Laki	2020	Sangat Memuaskan	Terlambat
MHS57	Laki-Laki	2020	Dengan Pujian	Tepat Waktu
MHS58	Perempuan	2020	Dengan Pujian	Tepat Waktu
MHS59	Laki-Laki	2020	Sangat Memuaskan	Tepat Waktu
MHS60	Perempuan	2020	Dengan Pujian	Tepat Waktu
MHS61	Perempuan	2020	Dengan Pujian	Tepat Waktu
MHS62	Perempuan	2020	Dengan Pujian	Tepat Waktu
MHS63	Laki-Laki	2020	Sangat Memuaskan	Tepat Waktu
MHS64	Laki-Laki	2020	Sangat Memuaskan	Tepat Waktu
MHS65	Perempuan	2020	Dengan Pujian	Tepat Waktu
MHS66	Laki-Laki	2019	Sangat Memuaskan	Terlambat
MHS67	Laki-Laki	2020	Sangat Memuaskan	Tepat Waktu
MHS68	Perempuan	2020	Sangat Memuaskan	Tepat Waktu

Nama	Jenis Kelamin	Tahun Masuk	Ipk	Status
MHS69	Perempuan	2020	Dengan Pujian	Tepat Waktu
MHS70	Perempuan	2020	Dengan Pujian	Tepat Waktu
MHS71	Perempuan	2020	Sangat Memuaskan	Tepat Waktu
MHS72	Perempuan	2020	Sangat Memuaskan	Tepat Waktu
MHS73	Laki-Laki	2020	Sangat Memuaskan	Tepat Waktu
MHS74	Laki-Laki	2020	Sangat Memuaskan	Tepat Waktu
MHS75	Laki-Laki	2020	Sangat Memuaskan	Tepat Waktu
MHS76	Perempuan	2020	Dengan Pujian	Tepat Waktu
MHS77	Perempuan	2020	Sangat Memuaskan	Tepat Waktu
MHS78	Perempuan	2020	Dengan Pujian	Tepat Waktu
MHS79	Laki-Laki	2020	Dengan Pujian	Tepat Waktu
MHS80	Perempuan	2020	Dengan Pujian	Tepat Waktu
MHS81	Laki-Laki	2020	Dengan Pujian	Tepat Waktu
MHS82	Perempuan	2020	Dengan Pujian	Tepat Waktu
MHS83	Perempuan	2020	Dengan Pujian	Tepat Waktu
MHS84	Perempuan	2020	Dengan Pujian	Tepat Waktu
MHS85	Perempuan	2020	Dengan Pujian	Tepat Waktu
MHS86	Perempuan	2020	Sangat Memuaskan	Tepat Waktu
MHS87	Perempuan	2020	Dengan Pujian	Tepat Waktu
MHS88	Perempuan	2020	Dengan Pujian	Tepat Waktu
MHS89	Perempuan	2020	Sangat Memuaskan	Tepat Waktu
MHS90	Perempuan	2020	Sangat Memuaskan	Terlambat
MHS91	Perempuan	2020	Dengan Pujian	Tepat Waktu
MHS92	Perempuan	2020	Dengan Pujian	Tepat Waktu

Nama	Jenis Kelamin	Tahun Masuk	Ipk	Status
MHS93	Perempuan	2020	Dengan Pujian	Tepat Waktu
MHS94	Perempuan	2020	Dengan Pujian	Tepat Waktu
MHS95	Laki-Laki	2020	Dengan Pujian	Tepat Waktu
MHS96	Perempuan	2020	Dengan Pujian	Tepat Waktu
MHS97	Perempuan	2020	Dengan Pujian	Tepat Waktu
MHS98	Perempuan	2020	Dengan Pujian	Tepat Waktu
MHS99	Perempuan	2020	Dengan Pujian	Tepat Waktu
MHS100	Perempuan	2020	Dengan Pujian	Tepat Waktu
MHS101	Perempuan	2020	Dengan Pujian	Tepat Waktu
MHS102	Perempuan	2020	Sangat Memuaskan	Tepat Waktu
MHS103	Perempuan	2020	Dengan Pujian	Tepat Waktu
MHS104	Perempuan	2020	Dengan Pujian	Tepat Waktu
MHS105	Perempuan	2020	Dengan Pujian	Tepat Waktu
MHS106	Perempuan	2020	Dengan Pujian	Tepat Waktu
MHS107	Perempuan	2020	Dengan Pujian	Tepat Waktu
MHS108	Perempuan	2020	Sangat Memuaskan	Tepat Waktu
MHS109	Laki-Laki	2020	Sangat Memuaskan	Tepat Waktu
MHS110	Laki-Laki	2020	Dengan Pujian	Tepat Waktu
MHS111	Laki-Laki	2020	Dengan Pujian	Tepat Waktu
MHS112	Laki-Laki	2020	Sangat Memuaskan	Tepat Waktu
MHS113	Laki-Laki	2020	Dengan Pujian	Tepat Waktu
MHS114	Perempuan	2020	Dengan Pujian	Tepat Waktu
MHS115	Laki-Laki	2020	Sangat Memuaskan	Tepat Waktu
MHS116	Laki-Laki	2020	Sangat Memuaskan	Tepat Waktu
MHS117	Perempuan	2020	Dengan Pujian	Tepat Waktu
MHS118	Perempuan	2020	Dengan Pujian	Tepat Waktu
MHS119	Perempuan	2020	Dengan Pujian	Tepat Waktu

Nama	Jenis Kelamin	Tahun Masuk	Ipk	Status
MHS120	Perempuan	2020	Sangat Memuaskan	Tepat Waktu
MHS121	Perempuan	2020	Dengan Pujian	Tepat Waktu
MHS122	Laki-Laki	2020	Dengan Pujian	Tepat Waktu
MHS123	Laki-Laki	2020	Dengan Pujian	Tepat Waktu
MHS124	Laki-Laki	2020	Sangat Memuaskan	Tepat Waktu
MHS125	Laki-Laki	2020	Sangat Memuaskan	Tepat Waktu
MHS126	Laki-Laki	2020	Dengan Pujian	Tepat Waktu
MHS127	Perempuan	2020	Sangat Memuaskan	Tepat Waktu
MHS128	Perempuan	2020	Dengan Pujian	Tepat Waktu
MHS129	Laki-Laki	2020	Dengan Pujian	Tepat Waktu
MHS130	Perempuan	2020	Sangat Memuaskan	Tepat Waktu
MHS131	Perempuan	2020	Dengan Pujian	Tepat Waktu
MHS132	Laki-Laki	2020	Sangat Memuaskan	Tepat Waktu
MHS133	Perempuan	2020	Dengan Pujian	Tepat Waktu
MHS134	Perempuan	2020	Dengan Pujian	Tepat Waktu
MHS135	Laki-Laki	2020	Dengan Pujian	Tepat Waktu
MHS136	Laki-Laki	2020	Sangat Memuaskan	Tepat Waktu
MHS137	Laki-Laki	2020	Dengan Pujian	Tepat Waktu
MHS138	Laki-Laki	2020	Dengan Pujian	Tepat Waktu
MHS139	Perempuan	2020	Sangat Memuaskan	Tepat Waktu
MHS140	Laki-Laki	2020	Sangat Memuaskan	Tepat Waktu
MHS141	Perempuan	2020	Dengan Pujian	Tepat Waktu

Dalam penelitian ini, pemilihan data dilakukan secara sistematis untuk memastikan keakuratan prediksi kelulusan mahasiswa menggunakan

metode K-Nearest Neighbors (KNN). Data yang digunakan berasal dari Universitas Labuhanbatu, Fakultas Sains dan Teknologi, Program Studi Sistem Informasi, yang mencakup 141 mahasiswa dari berbagai angkatan. Data ini dikumpulkan melalui sistem akademik dan mencakup variabel yang relevan dengan kelulusan mahasiswa, seperti NIM, jenis kelamin, tahun masuk, dan Indeks Prestasi Kumulatif (IPK).

Agar analisis lebih akurat, data yang diperoleh telah melalui tahap pembersihan (data cleaning) untuk menghapus nilai yang kosong atau tidak valid. Data yang telah diproses kemudian diimpor ke dalam RapidMiner untuk dilakukan eksplorasi lebih lanjut. Dengan menggunakan teknik data preprocessing, data yang tersedia dikonversi ke dalam format yang sesuai agar dapat digunakan dalam penerapan algoritma KNN. Pemilihan data yang representatif ini diharapkan dapat meningkatkan performa model dalam memprediksi kelulusan mahasiswa.

Pada sub-bab ini, dilakukan preprocessing data yang sangat penting untuk memastikan kualitas dataset yang akan digunakan dalam model KNN. Salah satu tahapan dalam preprocessing data adalah transformasi data, yang bertujuan untuk mengubah format dan struktur data agar lebih mudah diproses dan dianalisis. Pada tahap ini, data yang telah dikumpulkan sebelumnya akan dibersihkan dan dinormalisasi, serta diubah menjadi format yang sesuai untuk analisis lebih lanjut.

Tabel Transformasi Data

Variabel	Sebelum Transformasi	Setelah Transformasi
Gender MHS	Perempuan	1
	Laki-Laki	2
Tahun Masuk	2020	1
	2019	2
IPK	Dengan Pujian	1
	Sangat Memuaskan	2

Transformasi data menggunakan RapidMiner dilakukan dengan mengaplikasikan beberapa teknik, antara lain normalisasi dan standarisasi. Normalisasi bertujuan untuk menyesuaikan skala antar fitur, sehingga semua atribut memiliki kontribusi yang setara dalam perhitungan jarak, seperti pada algoritma KNN yang bergantung pada jarak antar titik data. Teknik ini memastikan bahwa tidak ada fitur yang mendominasi proses klasifikasi hanya karena memiliki skala yang lebih besar dibandingkan fitur lainnya. Selain itu, pengisian data yang hilang (missing values) juga dilakukan dengan menggunakan metode imputasi, seperti pengisian dengan nilai rata-rata atau median, untuk menghindari bias yang dapat terjadi akibat data yang hilang.

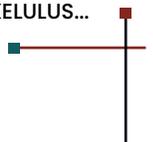
Tabel 4.2 Hasil Transformasi Data

Nama	Jenis Kelamin	Tahun Masuk	IPK	Status
MHS1	2	2	1	Tepat Waktu
MHS2	1	2	2	Tepat Waktu
MHS3	1	2	1	Tepat Waktu
MHS4	2	2	2	Tepat Waktu
MHS5	1	2	1	Tepat Waktu
MHS6	2	2	2	Tepat Waktu
MHS7	1	2	1	Tepat Waktu
MHS8	2	2	2	Tepat Waktu
MHS9	2	2	2	Tepat Waktu
MHS10	1	2	1	Tepat Waktu
.....				
MHS131	1	1	1	Tepat Waktu
MHS132	1	1	1	Tepat Waktu
MHS134	1	1	1	Tepat Waktu
MHS135	1	1	2	Tepat Waktu
MHS136	1	1	1	Tepat Waktu
MHS137	1	1	1	Tepat Waktu
MHS138	1	1	1	Tepat Waktu
MHS139	1	1	1	Tepat Waktu

Nama	Jenis Kelamin	Tahun Masuk	IPK	Status
MHS140	1	1	1	Tepat Waktu
MHS141	1	1	2	Tepat Waktu

Pembagian data latih dan data uji adalah tahap penting dalam penerapan model prediksi, seperti pada algoritma K-Nearest Neighbors (KNN), untuk memastikan keakuratan dan validitas hasil prediksi. Data latih digunakan untuk melatih model, di mana model ini mempelajari pola-pola yang ada pada data untuk membuat prediksi. Biasanya, data latih mencakup sebagian besar dataset dan terdiri dari pasangan data input dan output yang sudah diketahui. Dalam penelitian ini, data latih terdiri dari informasi mahasiswa seperti IPK, tahun masuk, dan status kelulusan yang akan digunakan untuk mempelajari hubungan antar atribut dan kelulusan mahasiswa.

Sementara itu, data uji digunakan untuk menguji kinerja model setelah dilatih menggunakan data latih. Data uji terdiri dari data yang belum pernah digunakan dalam pelatihan dan memiliki hasil output yang tidak diketahui oleh model. Tujuan dari pembagian ini adalah untuk mengevaluasi sejauh mana model dapat memprediksi dengan akurat berdasarkan data yang belum pernah dilihat sebelumnya. Dalam penelitian ini, data uji akan membantu mengukur akurasi model dalam memprediksi kelulusan mahasiswa berdasarkan atribut yang telah dipelajari pada data latih. Pembagian data latih dan uji secara adil dan proporsional sangat penting untuk mendapatkan hasil yang representatif dan dapat diandalkan.





BAB IV

HASIL PENGUJIAN DAN PEMBAHASAN

Penerapan Algoritma KNN

Pada sub-bab ini, dilakukan pembagian data menjadi dua bagian utama, yaitu data training dan data testing, untuk keperluan implementasi model prediksi menggunakan metode K-Nearest Neighbors (KNN) di RapidMiner. Pembagian data dilakukan dengan proporsi 70% untuk data training dan 30% untuk data testing, yang bertujuan untuk memastikan bahwa model dapat dilatih dengan jumlah data yang cukup, sementara data testing digunakan untuk mengukur akurasi model yang telah dibangun.

Tabel Data Latih

Nama	Jenis Kelamin	Tahun Masuk	IPK	Status
MHS1	2	2	1	Tepat Waktu
MHS2	1	2	2	Tepat Waktu
MHS3	1	2	1	Tepat Waktu
MHS4	2	2	2	Tepat Waktu
MHS5	1	2	1	Tepat Waktu

Nama	Jenis Kelamin	Tahun Masuk	IPK	Status
MHS6	2	2	2	Tepat Waktu
MHS7	1	2	1	Tepat Waktu
MHS8	2	2	2	Tepat Waktu
MHS9	2	2	2	Tepat Waktu
MHS10	1	2	1	Tepat Waktu
MHS11	2	2	1	Tepat Waktu
MHS12	1	2	2	Tepat Waktu
MHS13	1	2	1	Tepat Waktu
MHS14	1	2	1	Tepat Waktu
MHS15	1	2	1	Tepat Waktu
MHS16	2	2	1	Terlambat
MHS17	1	2	1	Terlambat
MHS18	1	2	2	Tepat Waktu
MHS19	2	2	1	Tepat Waktu
MHS20	1	2	2	Tepat Waktu
MHS21	1	2	2	Tepat Waktu
MHS22	2	2	2	Tepat Waktu
MHS23	1	2	1	Tepat Waktu
MHS24	2	2	2	Tepat Waktu
MHS25	1	2	2	Tepat Waktu
MHS26	1	2	2	Tepat Waktu
MHS27	1	2	2	Tepat Waktu
MHS28	1	1	2	Terlambat
MHS29	1	2	1	Tepat Waktu
MHS30	2	2	2	Tepat Waktu
MHS31	1	2	1	Tepat Waktu
MHS32	1	2	2	Tepat Waktu
MHS33	1	2	1	Tepat Waktu
MHS34	2	2	2	Tepat Waktu
MHS35	2	2	2	Tepat Waktu
MHS36	2	2	2	Tepat Waktu
MHS37	2	2	2	Tepat Waktu

Nama	Jenis Kelamin	Tahun Masuk	IPK	Status
MHS38	1	2	1	Tepat Waktu
MHS39	1	2	2	Tepat Waktu
MHS40	2	2	1	Tepat Waktu
MHS41	2	2	2	Tepat Waktu
MHS42	2	1	1	Tepat Waktu
MHS43	2	1	1	Tepat Waktu
MHS44	1	2	2	Terlambat
MHS45	1	1	1	Terlambat
MHS46	1	1	1	Tepat Waktu
MHS47	1	1	1	Tepat Waktu
MHS48	2	1	1	Tepat Waktu
MHS49	1	1	1	Tepat Waktu
MHS50	1	1	1	Tepat Waktu
MHS51	1	1	1	Tepat Waktu
MHS52	1	1	1	Tepat Waktu
MHS53	1	1	1	Tepat Waktu
MHS54	2	1	1	Tepat Waktu
MHS55	2	1	1	Tepat Waktu
MHS56	2	1	2	Terlambat
MHS57	2	1	1	Tepat Waktu
MHS58	1	1	1	Tepat Waktu
MHS59	2	1	2	Tepat Waktu
MHS60	1	1	1	Tepat Waktu
MHS61	1	1	1	Tepat Waktu
MHS62	1	1	1	Tepat Waktu
MHS63	2	1	2	Tepat Waktu
MHS64	2	1	2	Tepat Waktu
MHS65	1	1	1	Tepat Waktu
MHS66	2	2	2	Terlambat
MHS67	2	1	2	Tepat Waktu
MHS68	1	1	2	Tepat Waktu
MHS69	1	1	1	Tepat Waktu

Nama	Jenis Kelamin	Tahun Masuk	IPK	Status
MHS70	1	1	1	Tepat Waktu
MHS71	1	1	2	Tepat Waktu
MHS72	1	1	2	Tepat Waktu
MHS73	2	1	2	Tepat Waktu
MHS74	2	1	2	Tepat Waktu
MHS75	2	1	2	Tepat Waktu
MHS76	1	1	1	Tepat Waktu
MHS77	1	1	2	Tepat Waktu
MHS78	1	1	1	Tepat Waktu
MHS79	2	1	1	Tepat Waktu
MHS80	1	1	1	Tepat Waktu
MHS81	2	1	1	Tepat Waktu
MHS82	1	1	1	Tepat Waktu
MHS83	1	1	1	Tepat Waktu
MHS84	1	1	1	Tepat Waktu
MHS85	1	1	1	Tepat Waktu
MHS86	1	1	2	Tepat Waktu
MHS87	1	1	1	Tepat Waktu
MHS88	1	1	1	Tepat Waktu
MHS89	1	1	2	Tepat Waktu
MHS90	1	1	2	Terlambat
MHS91	1	1	1	Tepat Waktu
MHS92	1	1	1	Tepat Waktu
MHS93	1	1	1	Tepat Waktu
MHS94	1	1	1	Tepat Waktu
MHS95	2	1	1	Tepat Waktu
MHS96	1	1	1	Tepat Waktu
MHS97	1	1	1	Tepat Waktu
MHS98	1	1	1	Tepat Waktu

Dari total 141 data mahasiswa, sebanyak 98 data mahasiswa digunakan sebagai data training. Data ini akan digunakan untuk melatih model KNN sehingga algoritma dapat mempelajari pola-pola yang ada, termasuk

atribut-atribut penting seperti IPK dan tahun masuk mahasiswa, serta status kelulusan. Data training ini menjadi dasar bagi model dalam melakukan prediksi kelulusan berdasarkan data historis yang telah ada.

Tabel 4.4 Data Uji K-Nearest Neighbors (KNN)

Nama	Jenis Kelamin	Tahun Masuk	IPK	Status
MHS99	1	1	1	Tepat Waktu
MHS100	1	1	1	Tepat Waktu
MHS101	1	1	1	Tepat Waktu
MHS102	1	1	2	Tepat Waktu
MHS103	1	1	1	Tepat Waktu
MHS104	1	1	1	Tepat Waktu
MHS105	1	1	1	Tepat Waktu
MHS106	1	1	1	Tepat Waktu
MHS107	1	1	1	Tepat Waktu
MHS108	1	1	2	Tepat Waktu
MHS109	2	1	2	Tepat Waktu
MHS110	2	1	1	Tepat Waktu
MHS111	2	1	1	Tepat Waktu
MHS112	2	1	2	Tepat Waktu
MHS113	2	1	1	Tepat Waktu
MHS114	1	1	1	Tepat Waktu
MHS115	2	1	2	Tepat Waktu
MHS116	2	1	2	Tepat Waktu
MHS117	1	1	1	Tepat Waktu
MHS118	1	1	1	Tepat Waktu
MHS119	1	1	1	Tepat Waktu
MHS120	1	1	2	Tepat Waktu
MHS121	1	1	1	Tepat Waktu
MHS122	2	1	1	Tepat Waktu
MHS123	2	1	1	Tepat Waktu
MHS124	2	1	2	Tepat Waktu
MHS125	2	1	2	Tepat Waktu
MHS126	2	1	1	Tepat Waktu

Nama	Jenis Kelamin	Tahun Masuk	IPK	Status
MHS127	1	1	2	Tepat Waktu
MHS128	1	1	1	Tepat Waktu
MHS129	2	1	1	Tepat Waktu
MHS130	1	1	2	Tepat Waktu
MHS131	1	1	1	Tepat Waktu
MHS132	2	1	2	Tepat Waktu
MHS133	1	1	1	Tepat Waktu
MHS134	1	1	1	Tepat Waktu
MHS135	2	1	1	Tepat Waktu
MHS136	2	1	2	Tepat Waktu
MHS137	2	1	1	Tepat Waktu
MHS138	2	1	1	Tepat Waktu
MHS139	1	1	2	Tepat Waktu
MHS140	2	1	2	Tepat Waktu
MHS141	1	1	1	Tepat Waktu

Sedangkan, 43 data mahasiswa lainnya digunakan sebagai data testing. Data testing ini berfungsi untuk menguji kemampuan model dalam memprediksi kelulusan mahasiswa berdasarkan hasil pelatihan yang telah dilakukan pada data training. Pembagian ini memastikan bahwa model KNN diuji secara objektif, dan hasil prediksi dapat dianalisis untuk mengevaluasi akurasi, precision, recall, dan F1-score sebagai parameter utama dalam menilai performa model prediksi yang dibangun.

Perhitungan jarak Euclidean dan pemilihan K

Perhitungan jarak Euclidean digunakan untuk mengukur kedekatan antara data uji dan data latih dalam algoritma K-Nearest Neighbors (KNN). Jarak ini dihitung dengan rumus:

$$d(i, j) = \sqrt{\sum_{k=1}^n (x_{i,k} - x_{j,k})^2}$$

di mana:

$d(i,j)$ Jarak antara titik i dan j .

$x_{i,k}$: Fitur ke- k dari titik i .

$x_{j,k}$: Fitur ke- k dari titik j .

Dalam penelitian ini, nilai K yang dipilih adalah 15, yang berarti prediksi kelulusan mahasiswa didasarkan pada mayoritas hasil dari 15 tetangga terdekat. Pemilihan K ini bertujuan untuk menyeimbangkan akurasi dan stabilitas prediksi.

Tabel 3.3 Data baru yang akan diuji

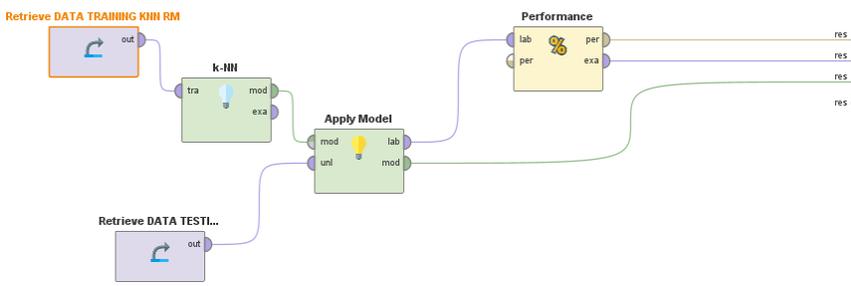
Nama	Jenis Kelamin	Tahun Masuk	IPK	Status
MHS99	1	1	1	?

$$\begin{aligned} \text{Mahasiswa 1} &= \sqrt{(2 - 1)^2 + (2 - 1)^2 + (1 - 1)^2} \\ &= 1,4142 \end{aligned}$$

Cara perhitungan sama sampai data latih mahasiswa ke 98 dengan menyesuaikan nilai dari data yang ingin dihitung jaraknya. Kemudian jarak diurutkan dari jarak yang terkecil ke jarak yang terbesar. Hasil akhir perhitungan jarak dan setelah diurutkan dari data terkecil ke data terbesar seperti tabel berikut.

Implementasi dengan Rapidminer

Tahap implementasi yaitu tahap yang dilakukan untuk merancang model untuk membantuu proses klasifikasi data.



Gambar Rancangan Model *K-Nearest Neighbor (KNN)*

Gambar diatas merupakan rancangan model machine learning dengan menggunakan metode *K-Nearest Neighbor (KNN)*. Implementasi ini dilakukan dengan menggunakan software data mining yaitu RapidMiner Studio. Berikut adalah penjelasan singkat mengenai setiap komponen dan perintah yang ada pada gambar tersebut:

1. Retrieve Data Training KNN ; Komponen ini digunakan untuk mengambil data pelatihan yang digunakan dalam model K-Nearest Neighbors (K-NN). Ini adalah data yang akan digunakan untuk melatih model.
2. Retrieve Data Testing KNN ; Komponen ini digunakan untuk mengambil data pengujian yang digunakan dalam model K-Nearest Neighbors (K-NN). Ini adalah data yang akan digunakan untuk menguji model.
3. K-NN ; Di sini, model K-Nearest Neighbors diterapkan pada data pelatihan. Terdapat beberapa parameter yang dapat dikonfigurasi, seperti tra (training data), mod (model), dan exa (data uji).
4. Apply Model ; Setelah model dilatih menggunakan data pelatihan, komponen ini digunakan untuk menerapkan model yang telah dibuat ke data baru (data uji). Proses ini menguji model dengan menggunakan parameter seperti mod (model) dan unl (unlabeled data).
5. Performance ; Komponen ini digunakan untuk mengevaluasi kinerja model dengan membandingkan prediksi yang dihasilkan oleh model dengan data yang sesungguhnya. Parameter seperti lab (label yang benar) dan per (performance) digunakan untuk mengukur performa model.

Secara keseluruhan, alur ini mengambil data pelatihan, melatih model menggunakan K-Nearest Neighbors, kemudian menguji model tersebut pada data uji dan mengevaluasi kinerjanya.

Tabel 4.5 Hasil Klasifikasi *K-Nearest Neighbor (KNN)*.

Nama	Jenis Kelamin	Tahun Masuk	IPK	Status
MHS99	1	1	1	Tepat Waktu
MHS100	1	1	1	Tepat Waktu
MHS101	1	1	1	Tepat Waktu
MHS102	1	1	2	Tepat Waktu

Nama	Jenis Kelamin	Tahun Masuk	IPK	Status
MHS103	1	1	1	Tepat Waktu
MHS104	1	1	1	Tepat Waktu
MHS105	1	1	1	Tepat Waktu
MHS106	1	1	1	Tepat Waktu
MHS107	1	1	1	Tepat Waktu
MHS108	1	1	2	Tepat Waktu
MHS109	2	1	2	Tepat Waktu
MHS110	2	1	1	Tepat Waktu
MHS111	2	1	1	Tepat Waktu
MHS112	2	1	2	Tepat Waktu
MHS113	2	1	1	Tepat Waktu
MHS114	1	1	1	Tepat Waktu
MHS115	2	1	2	Tepat Waktu
MHS116	2	1	2	Tepat Waktu
MHS117	1	1	1	Tepat Waktu
MHS118	1	1	1	Tepat Waktu
MHS119	1	1	1	Tepat Waktu
MHS120	1	1	2	Tepat Waktu
MHS121	1	1	1	Tepat Waktu
MHS122	2	1	1	Tepat Waktu
MHS123	2	1	1	Tepat Waktu
MHS124	2	1	2	Tepat Waktu
MHS125	2	1	2	Tepat Waktu
MHS126	2	1	1	Tepat Waktu
MHS127	1	1	2	Tepat Waktu
MHS128	1	1	1	Tepat Waktu
MHS129	2	1	1	Tepat Waktu
MHS130	1	1	2	Tepat Waktu
MHS131	1	1	1	Tepat Waktu
MHS132	2	1	2	Tepat Waktu
MHS133	1	1	1	Tepat Waktu
MHS134	1	1	1	Tepat Waktu

Nama	Jenis Kelamin	Tahun Masuk	IPK	Status
MHS135	2	1	1	Tepat Waktu
MHS136	2	1	2	Tepat Waktu
MHS137	2	1	1	Tepat Waktu
MHS138	2	1	1	Tepat Waktu
MHS139	1	1	2	Tepat Waktu
MHS140	2	1	2	Tepat Waktu
MHS141	1	1	1	Tepat Waktu

Tabel diatas merupakan hasil dari klasifikasi menggunakan model metode *K-Nearest Neighbor (KNN)*.

Pengujian Model

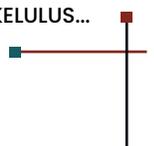
Evaluasi model dengan *confusion matrix* yang akan menampilkan akurasi dari proses klasifikasi ini.

accuracy: 100.00%

	true Tepat Waktu	true Terlambat	class precision
pred. Tepat Waktu	43	0	100.00%
pred. Terlambat	0	0	0.00%
class recall	100.00%	0.00%	

Gambar *Confusion matrix K-Nearest Neighbor (KNN)*

Tabel diatas merupakan hasil *confusion matrix K-Nearest Neighbor (KNN)* dimana menunjukkan akurasi sebesar 100%.





Gambar Visualisasi K-Nearest Neighbor (KNN)

Gambar diatas merupakan visualisasi hasil klasifikasi dari dataset klasifikasi menggunakan model metode *K-Nearest Neighbor (KNN)*.

Evaluasi Kinerja Model

Hasil pengujian model K-Nearest Neighbors (KNN) menunjukkan kinerja yang sangat baik dengan akurasi 100%. Tabel yang diberikan menunjukkan perhitungan dari confusion matrix, di mana model berhasil memprediksi semua mahasiswa yang lulus tepat waktu dengan tepat (43 prediksi benar untuk “Tepat Waktu”) dan tidak ada prediksi yang salah untuk kategori “Terlambat”. Hal ini tercermin dari nilai true Tepat Waktu yang mencapai 43, sementara true Terlambat adalah 0, yang menandakan bahwa tidak ada mahasiswa yang diprediksi terlambat namun pada kenyataannya lulus tepat waktu.

Namun, hasil ini juga menunjukkan kelemahan model dalam hal presisi dan recall untuk kategori “Terlambat”. Presisi untuk kelas “Terlambat” adalah 0%, yang berarti model tidak dapat mengidentifikasi mahasiswa yang terlambat lulus, dan recall untuk kelas “Terlambat” juga 0%, menandakan bahwa semua mahasiswa yang terlambat lulus tidak terdeteksi dalam

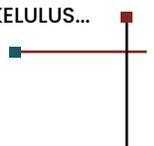
prediksi. Hal ini bisa disebabkan oleh model yang terlalu fokus pada prediksi kelulusan tepat waktu, sementara data untuk mahasiswa yang terlambat lulus sangat terbatas. Meskipun demikian, akurasi keseluruhan model tetap sangat tinggi, dengan recall untuk kelas “Tepat Waktu” mencapai 100%, menunjukkan bahwa model sangat baik dalam memprediksi mahasiswa yang lulus tepat waktu.

Kekuatan utama dari model K-Nearest Neighbors (KNN) dalam memprediksi kelulusan mahasiswa terletak pada akurasinya yang sangat tinggi, mencapai 100%. Hal ini menunjukkan bahwa model berhasil memprediksi semua mahasiswa yang lulus tepat waktu dengan benar, memberikan gambaran yang sangat baik terkait kelulusan yang tepat waktu. Keunggulan lainnya adalah kemampuannya untuk menangani data yang tidak linear, dengan memanfaatkan kedekatan antar data untuk menentukan prediksi, tanpa memerlukan asumsi distribusi data tertentu. Ini membuat model KNN sangat fleksibel dalam berbagai jenis dataset.

Namun, kelemahan utama model ini adalah rendahnya presisi dan recall untuk kategori “Terlambat”. Model gagal mengidentifikasi mahasiswa yang terlambat lulus, dengan presisi dan recall yang mencapai 0% untuk kelas tersebut. Hal ini mengindikasikan bahwa model kurang sensitif terhadap mahasiswa yang berisiko terlambat lulus, sehingga dapat menimbulkan masalah dalam merencanakan kebijakan akademik terkait mahasiswa yang membutuhkan intervensi. Kelemahan ini mungkin disebabkan oleh ketidakseimbangan data, di mana jumlah mahasiswa yang terlambat lulus jauh lebih sedikit dibandingkan yang lulus tepat waktu, membuat model lebih cenderung memprediksi mahasiswa yang lulus tepat waktu. Untuk meningkatkan kinerja model, penyesuaian lebih lanjut terhadap data dan parameter model, seperti pemilihan nilai K dan penanganan data tidak seimbang, diperlukan.

Faktor-faktor yang mempengaruhi akurasi dan efisiensi model KNN

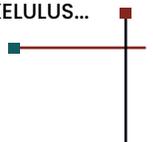
Akurasi dan efisiensi model K-Nearest Neighbors (KNN) dipengaruhi oleh berbagai faktor yang dapat mempengaruhi kualitas prediksi yang dihasilkan.



Salah satu faktor utama adalah pemilihan nilai K . Nilai K yang terlalu kecil dapat menyebabkan model menjadi sangat sensitif terhadap data pelatihan, sehingga meningkatkan risiko overfitting, di mana model terlalu menyesuaikan diri dengan data pelatihan dan gagal melakukan generalisasi dengan baik. Sebaliknya, nilai K yang terlalu besar dapat menyebabkan model kehilangan detail penting dalam data, mengarah pada underfitting.

Jarak antar data juga berperan penting dalam akurasi model. KNN mengandalkan perhitungan jarak, seperti Euclidean, untuk menentukan kedekatan antar data. Oleh karena itu, data yang memiliki skala yang berbeda-beda antara satu atribut dengan atribut lainnya perlu dinormalisasi untuk memastikan bahwa semua atribut memiliki pengaruh yang seimbang dalam perhitungan jarak.

Kualitas data juga memengaruhi efisiensi dan akurasi model. Data yang tidak lengkap, terdapat nilai yang hilang, atau tidak relevan dapat mengurangi kemampuan model untuk membuat prediksi yang akurat. Selain itu, dimensi data atau jumlah fitur juga mempengaruhi kinerja model. Semakin banyak fitur, semakin sulit bagi KNN untuk menghitung jarak dengan efisien, yang dapat mengurangi performa model secara keseluruhan. Oleh karena itu, pemilihan fitur yang relevan dan preprocessing data yang tepat sangat penting untuk mendapatkan hasil yang optimal.





BAB V

PENUTUP

Kesimpulan

Penelitian ini bertujuan untuk menguji efektivitas metode K-Nearest Neighbors (KNN) dalam memprediksi kelulusan mahasiswa di Universitas Labuhanbatu, khususnya pada Program Studi Sistem Informasi. Berdasarkan hasil analisis yang dilakukan, penelitian ini berhasil mencapai tujuan utama, yaitu mengembangkan model prediksi yang dapat memprediksi kelulusan mahasiswa dengan tingkat akurasi yang tinggi. Hasil pengujian model menunjukkan akurasi 100%, dengan model berhasil memprediksi semua mahasiswa yang lulus tepat waktu secara benar. Namun, model ini juga memiliki keterbatasan, terutama dalam memprediksi mahasiswa yang terlambat lulus, yang tercermin dalam nilai presisi dan recall untuk kelas “Terlambat” yang sangat rendah.

Efektivitas metode KNN dalam penelitian ini terbukti tinggi dalam hal prediksi kelulusan tepat waktu. Algoritma KNN dapat mengklasifikasikan mahasiswa ke dalam kategori “Tepat Waktu” dengan sangat akurat, karena model ini menggunakan jarak antar data untuk menentukan tetangga terdekat dan memberikan prediksi berdasarkan mayoritas kelas tetangga.

Namun, prediksi untuk mahasiswa yang terlambat lulus masih memerlukan perbaikan. Hal ini dapat disebabkan oleh ketidakseimbangan data, di mana jumlah mahasiswa yang terlambat lulus jauh lebih sedikit dibandingkan dengan yang lulus tepat waktu, sehingga memengaruhi kemampuan model untuk memprediksi kelas yang minoritas ini.

Dengan demikian, meskipun model ini sangat efektif untuk memprediksi kelulusan tepat waktu, penelitian ini juga mengidentifikasi area yang perlu ditingkatkan, khususnya dalam menangani ketidakseimbangan data dan meningkatkan prediksi untuk mahasiswa yang terlambat lulus. Ke depannya, penelitian lebih lanjut dapat dilakukan dengan mengeksplorasi teknik lain untuk menangani ketidakseimbangan kelas, seperti menggunakan oversampling atau teknik pembobotan kelas, untuk meningkatkan kinerja model dalam memprediksi kelulusan mahasiswa secara lebih menyeluruh. Secara keseluruhan, penelitian ini menunjukkan bahwa KNN adalah metode yang efektif dalam memprediksi kelulusan mahasiswa, terutama ketika diterapkan dalam konteks yang tepat dan dengan data yang relevan.

Rekomendasi

Berdasarkan hasil penelitian ini, terdapat beberapa saran untuk pengembangan lebih lanjut serta rekomendasi yang dapat diberikan kepada institusi pendidikan dalam penerapan metode K-Nearest Neighbors (KNN) untuk pengelolaan kelulusan mahasiswa. Pertama, penelitian ini menunjukkan bahwa model KNN dapat memberikan prediksi yang sangat akurat terkait kelulusan mahasiswa, terutama dalam kategori “Tepat Waktu”. Oleh karena itu, pengembangan model KNN dapat diarahkan pada peningkatan kemampuan model untuk mendeteksi kelulusan yang terlambat. Salah satu cara untuk meningkatkan akurasi dalam kategori “Terlambat” adalah dengan mengoptimalkan pemilihan nilai K dan melakukan penyesuaian terhadap data yang digunakan dalam pelatihan model, seperti menambah data mahasiswa yang terlambat lulus untuk memperkaya prediksi.

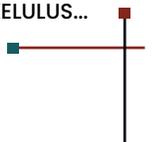
Kedua, penelitian ini mengindikasikan pentingnya penggunaan data yang relevan dan berkualitas tinggi. Oleh karena itu, bagi institusi pendidikan, disarankan untuk meningkatkan sistem pengumpulan dan

manajemen data akademik mahasiswa agar lebih lengkap dan terstruktur. Data yang lebih beragam dan lengkap akan membantu model prediksi dalam mengenali lebih banyak pola yang mempengaruhi kelulusan mahasiswa, sehingga dapat meningkatkan efisiensi dan akurasi dalam perencanaan akademik.

Selain itu, institusi pendidikan dapat menggunakan model KNN ini sebagai alat bantu dalam perencanaan dan pengelolaan kebijakan akademik, seperti penyesuaian alokasi dosen, pengaturan kurikulum, dan penyediaan program pendampingan bagi mahasiswa yang berisiko terlambat lulus. Mengingat pentingnya pengelolaan waktu dan sumber daya, KNN dapat menjadi alat yang sangat berguna dalam meningkatkan efisiensi pengelolaan pendidikan tinggi.

Secara keseluruhan, untuk pengembangan lebih lanjut, penelitian ini dapat diperluas dengan mempertimbangkan metode machine learning lainnya yang lebih canggih, seperti Random Forest atau Support Vector Machine (SVM), yang dapat memberikan hasil prediksi yang lebih kompleks dan mempertimbangkan lebih banyak variabel. Penelitian lanjutan juga dapat menilai pengaruh faktor-faktor eksternal, seperti kondisi sosial ekonomi mahasiswa, terhadap kelulusan.

Dengan demikian, model prediksi berbasis KNN ini dapat lebih dioptimalkan untuk meningkatkan kualitas pendidikan dan pengelolaan akademik di universitas.





Daftar Pustaka

- [1] A. Algarni, "Data Mining in Education," *Int. J. Adv. Comput. Sci. Appl.*, vol. 7, no. 6, 2016, doi: 10.14569/ijacsa.2016.070659.
- [2] Kurniawan, H., Munthe, I. R. . and Bangun, B. . (2022) "ANALYSIS OF NAIVE BAYES METHODS ON SCHOLARSHIP ADMISSIONS AT SMKS AL-WASLIYAH 2 MERBAU ", *Jurnal Mantik*, 6(1), pp. 364-373. doi: 10.35335/mantik.v6i1.2215..
- [3] Azwar S. Penyusunan Skala Psikologi Yogyakarta. Pustaka Pelajar. 2016.
- [4] Buku VI Matriks Penilaian Instrumen Akreditasi Program Studi Badan Akreditasi Nasional Perguruan Tinggi, 2008
- [5] Rohman, Abdul. "Model Algoritma K-Nearest Neighbor (KNN) Untuk Prediksi Kelulusan Mahasiswa." *Jurnal Unpand* 1.1 (2015). Web.
- [6] S. A. A. Kharis, "Prediksi Kelulusan Siswa Pada Mata Pelajaran Matematika Menggunakan Educational Data Mining," *J. Ris. Pembelajaran Mat. Sekol.*, vol. 7, no. 1, pp. 21–29, 2023, doi: 10.21009/jrpms.071.03.
- [7] T. Gori, "Preprocessing Data Dan Klasifikasi Untuk Prediksi Kinerja Akademik Siswa," *J. Teknol. Inf. Dan Ilmu Komput.*, vol. 11, no. 1, pp. 215–224, 2024, doi: 10.25126/jtiik.20241118074.
- [8] A. Desiani, S. Yahdin, and D. Rodiah, "Prediksi Tingkat Indeks Prestasi Kumulatif Akademik Mahasiswa Dengan Menggunakan Teknik Data

- Mining” *J. Teknol. Inf. Dan Ilmu Komput.*, vol. 7, no. 6, pp. 1237–1244, 2020, doi: 10.25126/jtiik.2020722493.
- [9] A. Desiani, “Perbandingan Implementasi Algoritma Naïve Bayes Dan K-Nearest Neighbor Pada Klasifikasi Penyakit Hati,” *Simkom*, vol. 7, no. 2, pp. 104–110, 2022, doi: 10.51717/simkom.v7i2.96.
- [10] Z. Zulkarnain, “Performance Comparison K-Nearest Neighbor, Naive Bayes, and Decision Tree Algorithms for Netflix Rating Classification,” *Ijatis*, vol. 1, no. 1, pp. 16–22, 2024, doi: 10.57152/ijatis.v1i1.1104.
- [11] Y. C. Tapidingan and D. Paseru, “Comparative Analysis of Classification Methods of KNN and Naïve Bayes to Determine Stress Level of Junior High School Students,” *Indones. J. Inf. Syst.*, vol. 2, no. 2, pp. 80–89, 2020, doi: 10.24002/ijis.v2i2.3035.



Tentang Penulis



Fratika Rahmaliana lahir di siluman pada 11 Mei 2002, menempuh pendidikan Sekolah dasar di SD 114368 Pulo Jantan, Kemudian Sekolah Menengah Pertama di SMP N1 NA IX-X Aek Kota Batu, Selanjutnya Sekolah Menengah Atas di SMA N1 Aek Natas, Sebelum akhirnya melanjutkan studi di jenjang perguruan tinggi. pernah menjadi bagian dari mahasiswa S1 Sistem Informasi di Universitas labuhanbatu sejak 2021-2025 dengan menemukan minat dan bakatnya dalam bidang teknologi komputer dan analisis data.



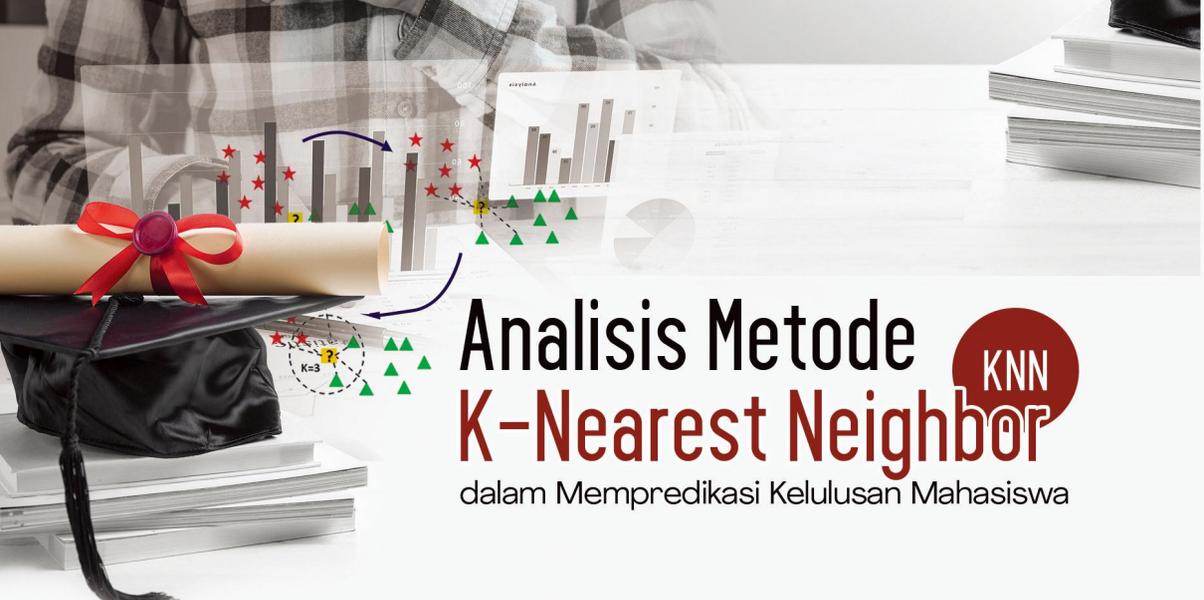
Budiarto Bangun, S.Sos., M.Kom. Lahir Medan 04 April 1970 Pendidikan S1 Dharmawangsa Medan dengan gelar (S.Sos), S2 Universitas Putra Indonesia (YPTK) Padang dengan gelar Magister Komputer (M.Kom). Saat ini sebagai tenaga pengajar Universitas Labuhanbatu dan aktif di kegiatan sosial lainnya.



Sudi Suryadi, S.Kom., M.Kom .Lahir di Sawahlunto 28 Februari 1979 Sumatera Barat. Pendidikan S1 di Universitas Putra Indonesia (YPTK) Padang (S.Kom) dan pendidikan S2 di Universitas Putra Indonesia (YPTK) Padang gelar Magister Komputer (M.Kom) (Saat ini sebagai tenaga pengajar Universitas Labuhanbatu dan aktif di kegiatan sosial lainnya.



Masrizal, S.Kom., M.Kom, NIDN 0105039401. Lahir Di Manggung 05 Maret 1994 Proses Dalam Menempuh Pendidikan dimulai Dari Sekolah Dasar SD.N 13 Manggung, SMP.N 7 Pariaman, SMA.N 4 Pariaman Dan Melanjudi Keperguruan Tinggi Swasta dengan. Jurusan Sistem Infomasi S1 (Sarjana) di Universitas Putra Indonesia “YPTK” Padang Tahun Lulusan 2016 Dan Melanjudi Magister Tahun 2016 Universitas Putra Indonesia”YPTK” Padang. Tahun Lulusan 2018 Dengan Jurusan Sistem Informasi, Fakultas Teknologi Informasi. Pada Tahun 2018. Saya Mengabdikan Diri sebagai Salah Satu Dosen Dibiidang Ilmu Komputer di Fakultas Sain dan Teknologi (FST) Prodi Sistem Informasi di Universitas LabuhanBatu salah satu menjadi Dosen Tetap di Kampus tersebut. Syukur Alhamdulillah sekarang saya mengabdikan diri dibawah Wakil Rektor III di Bidang Pengembangan, Kerjasama dan Alumni sebagai Ka.biro Bagian Pengembangan, Kerjasama dan Alumni serta Devisi Tracer Studi Di Universitas Labuhanbatu–Sumatera Utara.



Analisis Metode K-Nearest Neighbor

dalam Mempredikasi Kelulusan Mahasiswa

Pendidikan tinggi memiliki peran penting dalam membentuk masa depan individu dan masyarakat. Namun, salah satu tantangan terbesar yang dihadapi oleh banyak institusi pendidikan adalah mengelola kelulusan mahasiswa tepat waktu. Ketidaktepatan waktu kelulusan dapat berdampak pada kualitas pendidikan, reputasi institusi, serta kesempatan karier bagi mahasiswa. Oleh karena itu, penting bagi universitas untuk memiliki alat bantu yang efektif dalam memprediksi kelulusan mahasiswa, agar dapat merancang kebijakan pendidikan yang lebih tepat sasaran.

Penelitian ini bertujuan untuk mengeksplorasi penerapan metode K-Nearest Neighbors (KNN) dalam memprediksi kelulusan mahasiswa di Universitas Labuhanbatu, khususnya pada Fakultas Sains dan Teknologi Program Studi Sistem Informasi. Melalui metode KNN, diharapkan dapat diperoleh prediksi yang akurat mengenai status kelulusan mahasiswa, baik itu tepat waktu maupun terlambat. Hasil penelitian ini memberikan wawasan yang berguna bagi pihak universitas dalam mengambil kebijakan untuk mengoptimalkan proses pembelajaran dan pengelolaan mahasiswa.

Sebagai bagian dari upaya meningkatkan kualitas pendidikan, penelitian ini juga mengusulkan beberapa rekomendasi yang dapat diimplementasikan oleh institusi pendidikan. Rekomendasi tersebut mencakup pengembangan lebih lanjut dari model prediksi serta saran bagi universitas dalam memanfaatkan teknologi dan data yang tersedia untuk merancang kebijakan yang lebih efisien dan mendukung kelulusan mahasiswa secara tepat waktu.

