

**PERBANDINGAN ALGORITMA K-NEAREST NEIGHBOR (K-NN) DAN
NAIVE BAYES PADA DATA PENERIMA BEASISWA KARTU
INDONESIA PINTAR (KIP) KULIAH**

THESIS

**Oleh:
MANZILUR RAHMAN ROMADHON
NIM. 19841010**



**PROGRAM STUDI MAGISTER INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2023**

**PERBANDINGAN ALGORITMA K-NEAREST NEIGHBOR (K-NN) DAN
NAIVE BAYES PADA DATA PENERIMA BEASISWA KARTU
INDONESIA PINTAR (KIP) KULIAH**

THESIS

**Diajukan kepada:
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang
Untuk memenuhi Salah Satu Persyaratan dalam
Memperoleh Gelar Magister Komputer (M.Kom)**

**Oleh:
MANZILUR RAHMAN ROMADHON
NIM. 19841010**

**PROGRAM STUDI MAGISTER INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2023**


**PERBANDINGAN ALGORITMA K-NEAREST NEIGHBOR (K-NN) DAN
NAIVE BAYES PADA DATA PENERIMA BEASISWA KARTU
INDONESIA PINTAR (KIP) KULIAH**

THESIS


**Oleh:
MANZILUR RAHMAN ROMADHON
NIM. 19841010**

Telah diperiksa dan disetujui untuk diuji:
Tanggal: 26 September 2023

Pembimbing I,

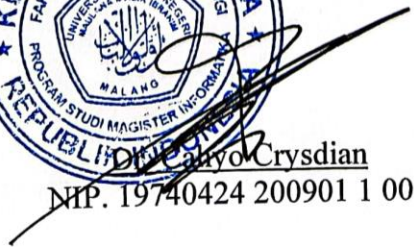

Dr. Muhammad Faisal, M.T
NIP. 19740510 200501 1 007

Pembimbing II,


Dr. M. Imamudin, Lc., MA
NIP. 19740602 200901 1 010

Mengetahui,
Ketua Program Studi Magister Informatika
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang




Ahyo Crysdiyan
NIP. 19740424 200901 1 008

**PERBANDINGAN ALGORITMA K-NEAREST NEIGHBOR (K-NN) DAN
NAIVE BAYES PADA DATA PENERIMA BEASISWA KARTU
INDONESIA PINTAR (KIP) KULIAH**

THESIS

**Oleh:
MANZILUR RAHMAN ROMADHON
NIM. 19841010**

Telah Dipertahankan di Depan Dewan Penguji Thesis
dan Dinyatakan Diterima Sebagai Salah Satu Persyaratan
Untuk Memperoleh Gelar Magister Komputer (M.Kom)
Tanggal: 08.November.2023.

Susunan Dewan Penguji

Penguji Utama : Dr. Usman Pagalay, M.Si
NIP. 19650414 200312 1 001

Ketua Penguji : Dr. Totok Chamidy, M. Kom
NIP. 19691222 200604 1 001

Sekretaris Penguji : Dr. Muhammad Faisal, M.T
NIP. 19740510 200501 1 007

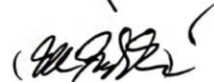
Anggota Penguji : Dr. M.Imamudin, Lc., MA
NIP. 19740602 200901 1 010

Tanda Tangan

()


()

()

()

Mengetahui dan Mengesahkan
Ketua Program Studi Magister Informatika
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang




Crysdiyan
NIP. 19740424 200901 1 008

HALAMAN PERSEMBAHAN

Thesis ini saya persembahkan kepada semua dosen Magister Informatika yang telah membimbing dan mengarahkan saya dalam menyelesaikan thesis ini, mudah-mudahan menjadi ilmu yang bermanfaat bagi saya.

Terima kasih saya haturkan kepada Ummi, Abah, Ibuk, (Alm) Abi, yang terus memberi support motivasi dan doanya sehingga saya dapat menyelesaikan amanah ini, mudah-mudahan bisa menjadi tanda bakti saya kepada mereka.

Saya persembahkan thesis ini juga kepada keluarga kecil saya, istri saya Afifatun Naila Rahmatika, Anak kembar saya Athif dan Abshir, adik Fariha, maafkan saya jika selama ini banyak waktu yang terlewatkan bersama kalian.

Saya persembahkan thesis ini untuk guru, sahabat dan kolega saya di Skwad HTQ, di Kunil café, yang telah mendukung terselesaikannya karya ini. Mudah-mudahan tetap guyub selawase.

PERNYATAAN KEASLIAN TULISAN

Saya yang bertanda tangan dibawah ini:

Nama : Manzilur Rahman Romadhon
NIM : 19841010
Program Studi : Magister Informatika
Fakultas : Sains dan Teknologi

Menyatakan dengan sebenarnya bahwa Thesis yang saya tulis ini benar-banar merupakan hasil karya saya sendiri, bukan merupakan pengambilalihan data, tulisan atau pikiran orang lain yang saya akui sebagai hasil tulisan atau pikiran saya sendiri, kecuali dengan mencantumkan sumber cuplikan pada daftar pustaka.

Apabila dikemudian hari terbukti atau dapat dibuktikan Thesis ini hasil jiplakan, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Malang, 09 November 2023

Yang membuat pernyataan,



Manzilur Rahman Romadhon
NIM. 19841010

KATA PENGANTAR

Assalamu'alaikum Wr. Wb.

Syukur alhamdulillah penulis haturkan kehadiran Allah SWT yang telah melimpahkan Rahmat dan Hidayah-Nya, sehingga penulis dapat menyelesaikan studi di Program Studi Magister Informatika Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang sekaligus menyelesaikan Thesis ini dengan baik.

Selanjutnya penulis haturkan ucapan terima kasih seiring do'a dan harapan jazakumullah ahsanal jaza' kepada semua pihak yang telah membantu terselesaikannya Thesis ini. Ucapan terima kasih ini penulis sampaikan kepada:

1. Bapak Dr. Muhammad Faisal, M.T dan Bapak Dr. Imamudin, Lc., M.A selaku dosen pembimbing Thesis, yang telah banyak memberikan pengarahan dan supportnya yang berharga.
2. Segenap sivitas akademika Program Studi Magister Informatika, terutama seluruh Bapak/ Ibu dosen, terima kasih atas segenap ilmu dan bimbingannya.
3. Afifatun Naila Rahmatika, S.Pd selaku istri penulis yang selalu setia bersama penulis, terima kasih atas pengertiannya selama ini
4. Kedua orang tua dan mertua yang senantiasa memberikan doa dan restunya kepada penulis dalam menuntut ilmu.
5. Semua pihak yang ikut membantu dalam menyelesaikan Thesis ini baik berupa materiil maupun moril.

Penulis menyadari bahwa dalam penyusunan Thesis ini masih terdapat kekurangan dan penulis berharap semoga Thesis ini bias memberikan manfaat kepada para pembaca khususnya bagi penulis secara pribadi. *Amin Ya Rabbal Alamin.*

Wassalamu'alaikum Wr. Wb.

Malang, November 2023
Penulis

DAFTAR ISI

HALAMAN SAMPUL	i
HALAMAN JUDUL	ii
HALAMAN PERSETUJUAN	iii
HALAMAN PENGESAHAN	iv
HALAMAN PERSEMBAHAN	v
HALAMAN PERNYATAAN KEASLIAN TULISAN	vi
KATA PENGANTAR	vii
DAFTAR ISI	viii
DAFTAR GAMBAR	x
DAFTAR TABEL	xi
ABSTRAK	xiii
ABSTRACT	xiv
المخلص	xv
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Pernyataan Masalah.....	4
1.3 Tujuan Penelitian.....	4
1.4 Manfaat Penelitian.....	4
1.5 Batasan Masalah.....	5
BAB II KAJIAN PUSTAKA	6
2.1 Penelitian Terdahulu.....	6
2.2 Perbandingan Penelitian Terdahulu.....	10
2.3 Landasan Teori	12
2.3.1 Kartu Indonesia Pintar Kuliah.....	12
2.3.2 Algoritma Naive Bayes	14
2.3.3 Algoritma K-Nearest Neighbor.....	15
2.3.4 Kajian Integrasi Sains dan Islam.....	16
BAB III METODOLOGI PENELITIAN	21
3.1 Desain Penelitian.....	21
3.2 Instrumen Penelitian.....	23
3.3 Pengumpulan Data	23
3.4 Analisis <i>Fishbone</i>	24

3.5 Analisis Data	25
3.5.1 Data Selection	26
3.5.2 Data preprocessing.....	28
3.5.3 Data transformation	28
3.5.4 Data Mining	29
3.5.5 Evaluasi.....	36
BAB IV METODE NAIVE BAYES	37
4.1 Desain Metode Naive Bayes	37
4.2 Pengujian Naive Bayes.....	40
4.2.1. Skenario pengujian.....	40
4.2.2. Hasil pengujian.....	42
BAB V METODE K-NEAREST NEIGHBOR (k-NN)	50
5.1 Desain Metode k-NN	50
5.2 Skenario pengujian.....	54
5.3 Hasil pengujian k-NN.....	57
5.3.1. Matrik Evaluasi	57
5.3.2. Confusion Matrix	63
BAB VI PEMBAHASAN.....	69
6.1 Perbandingan k-NN dan Naive Bayes.....	69
6.2 Analisis Matrik Evaluasi	69
BAB VII PENUTUP.....	73
5.1 Kesimpulan.....	73
5.2 Saran.....	74
DAFTAR PUSTAKA	75

DAFTAR GAMBAR

Gambar 2.1 Alur Pendaftaran KIP Kuliah UIN Maulana Malik Ibrahim Malang.....	14
Gambar 3.1 Diagram Alur Penelitian.....	21
Gambar 3.2 diagram <i>fishbone</i>	24
Gambar 3.3 Langkah-langkah dalam proses KDD (Fayyad et al., 1996).....	25
Gambar 4.1 Desain Metode Naive Bayes	37
Gambar 4.2 Proses import dataset.....	37
Gambar 4.3 Proses memisahkan feature dan target	38
Gambar 4.4 Pemisahan data training dan data uji.....	38
Gambar 4.5 Implementasi model Naive Bayes.....	39
Gambar 4.6 Proses prediksi dengan Naive Bayes.....	39
Gambar 4.7 Melakukan evaluasi Model Naive Bayes	40
Gambar 4.8 Hasil pengujian naive Bayes	49
Gambar 5.1 Desain Metode k-Nearest Neighbor	50
Gambar 5.2 Import dataset untuk proses k-NN	50
Gambar 5.3 proses mengganti data yang tidak sesuai.....	51
Gambar 5.4 seleksi kolom sebelum proses encode	51
Gambar 5.5 Proses encoding data	52
Gambar 5.6 proses split data untuk k-NN.....	52
Gambar 5.7 Proses pemisahan feature dan target	52
Gambar 5.8 Proses pemisahan feature dan target	53
Gambar 5.9 Proses Prediksi pada k-NN.....	53
Gambar 5.10 Menampilkan confusion matrix k-NN	54
Gambar 5.11 Menampilkan matrik evaluasi k-NN	54
Gambar 5.12 Grafik hasil pengujian k-NN dengan variasi k.....	59
Gambar 5.13 Hasil pengujian nilai presisi k-NN dengan variasi k.....	60
Gambar 5.14 Hasil pengujian nilai recall k-NN dengan variasi k	61
Gambar 5.15 Rata-rata hasil pengujian k-NN.....	63
Gambar 6.1 Perbandingan Akurasi k-NN dan Naive Bayes	70
Gambar 6.2 Perbandingan Presisi k-NN dan Naive Bayes	71
Gambar 6.3 Perbandingan Recall k-NN dan Naive Bayes	72

DAFTAR TABEL

Tabel 2.1 Penelitian terdahulu yang telah dilakukan	9
Tabel 2.2 Perbandingan penelitian terdahulu dengan penelitian sekarang	11
Tabel 2.3 Kelebihan dan kekurangan algoritma K-NN dan Naive Bayes	16
Tabel 3.1 Identifikasi akar permasalahan berdasarkan diagram fishbone	25
Tabel 3.2 data jumlah pendaftar KIP Kuliah	26
Tabel 3.3 Atribut dan nilai data pendaftar KIP Kuliah	26
Tabel 3.4 Proses transformasi data beasiswa KIP Kuliah	28
Tabel 3.5 Data Pendaftar KIP	31
Tabel 3.6 Hasil Encode Data Pendaftar KIP	32
Tabel 3.7 encoding data beasiswa untuk perhitungan dalam K-NN	35
Tabel 3.8 Menghitung jarak euclidian	36
Tabel 3.9 Mengurutkan hasil perhitungan jarak	36
Gambar 4.3 Pemisahan data training dan data uji	38
Tabel 4.1 Skenario pengujian naive Bayes	42
Tabel 4.2 confusion matrix naive Bayes pengujian 1	43
Tabel 4.3 confusion matrix naive Bayes pengujian 2	43
Tabel 4.4 confusion matrix naive Bayes pengujian 3	44
Tabel 4.5 confusion matrix naive Bayes pengujian 4	45
Tabel 4.6 confusion matrix naive Bayes pengujian 5	45
Tabel 4.7 confusion matrix naive Bayes pengujian 6	46
Tabel 4.8 confusion matrix naive Bayes pengujian 7	46
Tabel 4.9 confusion matrix naive Bayes pengujian 8	47
Tabel 4.10 hasil uji coba metode naive Bayes	47
Tabel 5.1 Skenario pengujian k-NN	55
Tabel 5.2 skenario pengujian k-NN dengan variasi k	55
Tabel 5.3 Hasil pengujian akurasi k-NN dengan variasi k	58
Tabel 5.4 Hasil pengujian nilai presisi k-NN	59
Tabel 5.5 Hasil Pengujian nilai recall k-NN	61
Tabel 5.6 Rata-rata hasil pengujian k-NN dengan variasi k	62
Tabel 5.7 Confusion matrix k-NN pengujian ke-1	63
Tabel 5.8 Confusion matrix k-NN pengujian ke-2	64
Tabel 5.9 Confusion matrix k-NN pengujian ke-3	65

Tabel 5.10 Confusion matrix k-NN pengujian ke-4.....	65
Tabel 5.11 Confusion matrix k-NN pengujian ke-5.....	66
Tabel 5.12 Confusion matrix k-NN pengujian ke-6.....	66
Tabel 5.13 Confusion matrix k-NN pengujian ke-7.....	67
Tabel 5.14 Confusion matrix k-NN pengujian ke-8.....	68
Tabel 6.1 perbandingan hasil pengujian k-NN dan Naive Bayes	69

ABSTRAK

Romadhon, Manzilur Rahman. 2023. **Perbandingan Algoritma K-Nearest Neighbor (K-NN) dan Naive Bayes Pada Data Penerima Beasiswa Kartu Indonesia Pintar (KIP) Kuliah**. Thesis Program Studi Magister Informatika Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang. (I) Dr. Muhammad Faisal, M.T. (II) Dr. M. Imamudin, Lc., MA

Kata kunci: KIP, K-NN, Naive Bayes, Klasifikasi

UU No. 12 tahun 2012 memberikan mandat untuk mewujudkan pemerataan dan keterjangkauan akses pendidikan tinggi yang dapat dilakukan melalui pemberian beasiswa. Beasiswa Kartu Indonesia Pintar (KIP) Kuliah merupakan salah satu usaha pemerintah untuk mewujudkan akses pendidikan tinggi bagi mahasiswa yang berprestasi dan kurang mampu. Namun, pemberian beasiswa harus tepat sasaran untuk menghindari pemborosan anggaran. Metode klasifikasi seperti k-NN dan Naive Bayes dapat digunakan untuk melakukan klasifikasi kelayakan penerima beasiswa KIP Kuliah. Performa dari kedua metode tersebut dianalisis performanya menggunakan metrik evaluasi. Pengujian yang dilakukan pada metode Naive Bayes menghasilkan akurasi 62,75%, presisi 65,88%, dan recall 88%. Sementara hasil pengujian k-NN untuk k=11 adalah akurasi 63,88%, presisi 66,13%, dan recall 90,88%. k-NN ternyata memiliki performa lebih unggul daripada naive Bayes dalam melakukan prediksi kelayakan penerima beasiswa KIP Kuliah berdasarkan hasil metrik evaluasi yang diperoleh.

ABSTRACT

Romadhon, Manzilur Rahman. 2023. **Comparison of the K-Nearest Neighbor (K-NN) and Naive Bayes Algorithms for Indonesia Smart Card (KIP) College Scholarship Recipients**. Thesis Master of Informatics Study Program, Faculty of Science and Technology, Maulana Malik Ibrahim State Islamic University Malang. (I) Dr. Muhammad Faisal, M.T. (II) Dr. M. Imamudin, Lc., MA

Keywords: KIP, K-NN, Naive Bayes, Classification

Law No. 12/2012 provides a mandate to realize equitable and affordable access to higher education which can be done through scholarships. The Indonesia Smart Card (KIP) Lecture Scholarship is one of the government's efforts to realize access to higher education for outstanding and underprivileged students. However, the scholarship must be right on target to avoid wasting the budget. Classification methods such as k-NN and Naive Bayes can be used to classify the eligibility of KIP College scholarship recipients. The performance of the two methods is analyzed using evaluation metrics. Tests conducted on the Naive Bayes method resulted in an accuracy of 62.75%, precision of 65.88%, and recall of 88%. While the k-NN test results for k=11 are 63.88% accuracy, 66.13% precision, and 90.88% recall. k-NN turns out to have superior performance than naive Bayes in predicting the eligibility of KIP Lecture scholarship recipients based on the results of the evaluation metrics obtained.

الملخص

رمضان، منزل الرحمن. ٢٠٢٣. التنبؤ بأهلية المستفيدين المحتملين من المنح الدراسية لبطاقة إندونيسيا الذكية (KIP) للكلية باستخدام أقرب جار (K-NN) وخوارزميات Naive Bayes رسالة ماجستير في برنامج دراسة المعلوماتية، كلية العلوم والتكنولوجيا، جامعة مولانا مالك إبراهيم الإسلامية الحكومية، مالانج. (أنا) د. محمد فيصل، الماجستير. (الثاني) د. محمد إمام الدين، الماجستير

الكلمات المفتاحية: KIP، K-NN، Naive Bayes، التنبؤ

القانون رقم ١٢ لعام ٢٠١٢ على تفويض لتحقيق الوصول المتساوي والميسور التكلفة إلى التعليم العالي والذي يمكن تحقيقه من خلال توفير المنح الدراسية. تعد منحة الكلية "بطاقة إندونيسيا الذكية (KIP)" إحدى الجهود التي تبذلها الحكومة لتوفير الوصول إلى التعليم العالي للطلاب المتفوقين والمحرومين. ومع ذلك، يجب أن تكون جوائز المنح الدراسية صحيحة في الهدف لتجنب إهدار الميزانية. يمكن استخدام أساليب التصنيف مثل k-NN و Naive Bayes للتنبؤ بأهلية المستفيدين من منح KIP الدراسية للجامعات. تم تحليل أداء الطريقتين باستخدام مقاييس التقييم. أنتجت الاختبارات التي تم إجراؤها باستخدام طريقة Naive Bayes دقة بلغت ٦٢,٧٥٪، و ٦٥,٨٨٪ دقة، و ٨٨٪ استرجاعاً. وفي الوقت نفسه، كانت نتائج اختبار k-NN لـ k=11 دقة ٦٣,٨٨٪ ودقة ٦٦,١٣٪ واستدعاء ٩٠,٨٨٪. يبدو أن أداء k-NN متفوق على أداء Bayes الساذج في التنبؤ بأهلية المستفيدين من منحة KIP الجامعية بناءً على نتائج مقاييس التقييم التي تم الحصول عليها.

BAB I

PENDAHULUAN

1.1 Latar Belakang

Undang-Undang Nomor 12 Tahun 2012 tentang Pendidikan Tinggi telah memberikan mandat kepada pemerintah untuk mewujudkan keterjangkauan dan pemerataan yang berkeadilan dalam mendapatkan akses pendidikan tinggi yang bermutu dan relevan dengan kepentingan masyarakat bagi kemajuan, kemandirian, dan kesejahteraan. Pemerintah berkewajiban meningkatkan akses dan kesempatan belajar serta menyiapkan manusia Indonesia yang cerdas dan kompetitif. (Undang-Undang (UU) No 12 Tahun 2012 Tentang Pendidikan Tinggi, 2012). Salah satu usaha pemerintah untuk meningkatkan akses belajar kepada masyarakat ialah melalui pemberian beasiswa.

Beasiswa Kartu Indonesia Pintar (KIP) Kuliah adalah salah satu jalur beasiswa yang ditawarkan pemerintah untuk meningkatkan akses pendidikan tinggi bagi masyarakat yang berprestasi dan kurang mampu secara ekonomi. (K. Agama, 2020) Beasiswa KIP Kuliah dulunya bernama beasiswa bidikmisi yang kemudian berganti nama pada tahun 2020. Sejak tahun 2015 sampai 2019, Ditjen Pendidikan Islam melalui Direktorat Pendidikan Tinggi Keagamaan Islam telah memberikan beasiswa Bidikmisi kepada 37.850 mahasiswa. Setelah bertransformasi menjadi KIP Kuliah, kuotanya mengalami kenaikan cukup signifikan. Jika pada 2019 kuota Bidikmisi hanya 11.000 mahasiswa, maka pada tahun 2020 menjadi 17.565 orang. (Basori, 2020). Dengan meningkatnya kuota

penerima KIP Kuliah, perlu dilakukan langkah yang akurat dalam menentukan kelayakan penerima KIP Kuliah.

Pemberian beasiswa juga merupakan salah satu bentuk sikap tolong-menolong dalam kebaikan. Tolong menolong terutama dalam hal kebaikan merupakan ajaran Al-Qur'an Surat Al-Maidah ayat 2. Menurut ayat tersebut Allah mengajak orang-orang mukmin untuk saling membantu dalam kebaikan dan takwa, seperti pemberian beasiswa kepada penerima yang layak. Allah melarang orang-orang mukmin saling membantu atau justru bersekongkol dalam hal kejelekan, seperti misalnya kecurangan yang dilakukan dalam proses pembagian beasiswa.

Salah satu cara yang dapat dilakukan dalam proses seleksi penerimaan Beasiswa KIP Kuliah adalah dengan melakukan klasifikasi calon penerima beasiswa, karena hasil klasifikasi yang tepat sangat penting untuk menentukan kelayakan penerima beasiswa. Beberapa metode untuk melakukan klasifikasi kelayakan penerima beasiswa telah diusulkan oleh banyak peneliti. Seperti penelitian M. Kholil yang melakukan Penerapan Metode k-Nearest Neighbor (k-NN) dalam Proses Seleksi Penerima Beasiswa.(Kholil, 2018) A. Mulyani dkk meneliti The Prediction of PPA and KIP Kuliah Scholarship Recipients Using Naive Bayes Algorithm.(Mulyani et al., 2022) G. Suganda dkk melakukan penelitian Penentuan Penerima Bantuan Beasiswa Kartu Indonesia Pintar (KIP Kuliah) Menggunakan Naive Bayes Classifier.(Suganda et al., 2022) P. A. Ani dan A. Andri meneliti Penerapan Algoritma Naive Bayes Untuk Klasifikasi Mahasiswa Penerima KIP Pada Universitas Bina Darma (Ani & Andri, 2022) A.

Sumiah dan N. Mirantika juga melakukan perbandingan metode K-Nearest Neighbour dan Naive Bayes untuk rekomendasi penentuan mahasiswa penerima beasiswa.(Sumiah & Mirantika, 2020)

Algoritma Naive Bayes memiliki kelebihan dalam kecepatan, membutuhkan data training sedikit, dan kemudahan menentukan probabilitas prediksi tetapi memiliki kelemahan dalam hal probabilitas hasil prediksi kurang dapat diandalkan. K-NN memiliki kelebihan mudah digunakan dalam klasifikasi biner dan multiclass dengan baik dan metodenya cukup mudah dipahami dan dilaksanakan. Namun k-NN memiliki kelemahan dalam hal kecepatan karena metode ini menghitung jarak antar semua poin sehingga membutuhkan waktu yang lebih lama.(Verdhan, 2020)

Penelitian ini akan melakukan perbandingan algoritma Naive Bayes dan K-NN untuk mengetahui performa kedua algoritma tersebut. Penggunaan metode k-NN dan Naive Bayes terbukti dapat meningkatkan akurasi dan performa dalam klasifikasi pengenalan emosi dibandingkan penggunaan satu metode seperti k-NN atau Naive Bayes.(Lee, 2015). Kombinasi k-NN dan Naive Bayes dengan pembobotan atribut juga mengungguli metode tradisional Naive Bayes dan k-NN(Jiang et al., 2006). Penggunaan Metode berbasis jarak yaitu k-NN dan algoritma berbasis statistik yaitu Naive Bayes dalam dalam sebagian besar kondisi dapat mengungguli algoritma lain dengan menggunakan kelebihanannya dan menghindari kelemahannya.(Ferdousy, Elma Zannatul, Md Mafijul Islam, 2013)

Perbandingan metode k-NN dan Naive Bayes akan digunakan dalam penelitian ini untuk mencari nilai akurasi terbaik dalam melakukan klasifikasi

Kelayakan Penerima Beasiswa KIP Kuliah. Pada penelitian ini penggunaan metode K-Nearest Neighbor (k-NN) dan Naive Bayes diharapkan dapat meningkatkan efisiensi dan efektivitas dalam melakukan klasifikasi kelayakan penerima beasiswa KIP Kuliah. Selain itu dengan peningkatan akurasi metode diharapkan dapat mempermudah proses pengambilan keputusan penerima beasiswa KIP Kuliah yang layak.

1.2 Pernyataan Masalah

Berdasarkan latar belakang masalah di atas, masalah yang akan dijadikan pokok pembahasan dalam penelitian ini sebagai berikut:

- a. Bagaimana perbandingan algoritma k-NN dan Naive Bayes untuk melakukan klasifikasi kelayakan calon penerima beasiswa KIP Kuliah?
- b. Bagaimana performa dan akurasi algoritma k-NN dan Naive Bayes pada data calon penerima beasiswa KIP Kuliah?

1.3 Tujuan Penelitian

- a. Melakukan analisis perbandingan metode Naive Bayes dan k-NN pada data calon penerima beasiswa KIP Kuliah.
- b. Melakukan analisis performa dan akurasi algoritma Naive Bayes dan k-NN untuk klasifikasi kelayakan calon penerima beasiswa KIP Kuliah.

1.4 Manfaat Penelitian

- a. Membantu dalam proses seleksi karena dapat melakukan klasifikasi kelayakan calon penerima beasiswa KIP Kuliah.
- b. Meningkatkan efektivitas dalam pengelolaan beasiswa dengan melakukan klasifikasi kelayakan calon penerima beasiswa KIP Kuliah.

- c. Membantu dalam pengambilan keputusan yang lebih baik dan tepat
- d. Membantu mengidentifikasi faktor-faktor yang mempengaruhi kelayakan calon penerima beasiswa KIP Kuliah.
- e. Mengetahui performa algoritma Naive Bayes dan k-NN dalam menyelesaikan masalah kelayakan calon penerima beasiswa KIP Kuliah

1.5 Batasan Masalah

Adapun Batasan masalah dalam penelitian ini sebagai berikut:

- a. Penelitian ini hanya digunakan untuk kelayakan calon penerima beasiswa yang mendaftar melalui jalur KIP Kuliah. Dataset yang digunakan dalam penelitian ini adalah data pendaftar KIP Kuliah tahun 2022 di Universitas Islam Negeri Maulana Malik Ibrahim Malang dari berbagai jalur masuk.
- b. Penelitian ini hanya menggunakan algoritma Naive Bayes dan k-NN dalam proses prediksi kelayakan calon penerima beasiswa KIP Kuliah.
- c. Penelitian ini dilakukan pada periode tertentu, sehingga hasil yang diperoleh hanya berlaku untuk periode tersebut saja.

BAB II

KAJIAN PUSTAKA

2.1 Penelitian Terdahulu

Penelitian yang diusulkan oleh penulis dengan judul “Perbandingan Algoritma K-Nearest Neighbor (K-NN) dan Naive Bayes Pada Data Penerima Beasiswa Kartu Indonesia Pintar (KIP) Kuliah” merupakan penelitian lanjutan yang diambil berdasarkan beberapa penelitian terdahulu yang telah dilakukan oleh peneliti-peneliti sebelumnya dengan topik serupa.

Penelitian oleh Mulyani et al yang berjudul *The Prediction Of PPA And KIP Kuliah Scholarship Recipients Using Naive Bayes Algorithm* melakukan prediksi diterima atau ditolaknya calon pendaftar beasiswa PPA dan KIP-K di Universitas Teknologi Garut. Peneliti menggunakan 145 data mahasiswa dan menggunakan metode Naive Bayes sebagai pengklasifikasi. Hasilnya nilai akurasi sebesar 80% untuk PPA dan 91% untuk beasiswa KIP-K.

Suganda et al juga meneliti *Penentuan Penerima Bantuan Beasiswa Kartu Indonesia Pintar (KIP) Kuliah Menggunakan Naive Bayes Classifier* yang menghasilkan nilai akurasi 88.21%. dataset yang diolah ialah 100 data calon penerima bantuan KIP Kuliah di Universitas Catur Insan Cendekia periode 2019/2020. Atribut data yang digunakan diantaranya penghasilan orang tua, rekening listrik, luas tanah bangunan, dan jumlah tanggungan orang tua.

Penelitian Ani & Andri melakukan *Penerapan Algoritma Naive Bayes Untuk Klasifikasi Mahasiswa Penerima KIP Pada Universitas Bina Darma* menggunakan metode Naive Bayes dengan teknik CRISP-DM. data yang diolah

adalah 342 data mahasiswa penerima KIP dan diolah dengan bantuan tool Rapidminer. Nilai akurasi yang diperoleh adalah 49,15%.

Perbandingan Metode K-Nearest Neighbor dan Naive Bayes untuk Rekomendasi Penentuan Mahasiswa Penerima Beasiswa pada Universitas Kuningan adalah penelitian yang telah dilakukan oleh Sumiah & Mirantika untuk menghasilkan sebuah sistem informasi berbasis visual basic.net dan sql server. Data yang digunakan untuk penelitian adalah dari database penerimaan mahasiswa baru, data riwayat akademik mahasiswa 2010 dan file mentah penerimaan beasiswa tahun 2012. Selain itu juga data nilai IPK mahasiswa semester 4 digunakan sebagai sumber data yang akan diolah. Hasil penelitian menunjukkan Nilai akurasi K-NN 100% dan Naive Bayes 99,89%

Metode Naive Bayes Untuk Penentuan Penerima Beasiswa Bidikmisi Universitas Mulawarman merupakan judul yang diambil Dahri, D untuk melakukan analisis tentang Penentuan Penerima Bidikmisi menggunakan metode Naive Bayes. Penelitian ini menghasilkan sebuah aplikasi sistem pendukung keputusan untuk membantu bagian proses seleksi. Variabel yang digunakan antara lain pekerjaan orang tua, penghasilan orang tua, dan jumlah tanggungan. Namun dalam penelitian ini masih menggunakan data penerima dari jalur SNMPTN saja. Hasil akurasi yang diperoleh ialah 85,56%.

Risman et al melakukan penelitian dengan judul Penerapan Metode K-Nearest Neighbor Pada Aplikasi Penentu Penerima Beasiswa Mahasiswa Di STMIK Sinar Nusantara Surakarta. Penelitian ini bertujuan untuk membangun sebuah aplikasi penentuan calon penerima beasiswa yang daapt mempermudah

tim seleksi dalam menentukan calon penerima. Metode pengumpulan data dilakukan dengan wawancara dan observasi dan implementasi sistem menggunakan bahasa pemrograman PHP dan database mysql. Pengujian dilakukan dengan mengambil 22 data sampel sehingga diperoleh nilai akurasi dengan metode k-nearest neighbor sebesar 90,90%

Noviana et al melakukan penelitian Analisis Rekomendasi Penerima Beasiswa Menggunakan Algoritma K-Nearest Neighbor (K-NN) dan Algoritma C4. 5. Peneliti mengambil sampel 1000 data calon penerima beasiswa Peningkatan Prestasi Akademik (PPA). 907 mahasiswa diklasifikasikan dengan benar menggunakan metode K-NN dan 883 mahasiswa diklasifikasikan dengan benar menggunakan metode C4.5. Hasil penelitian menunjukkan akurasi algoritma K-NN sebesar 90,7% sedangkan pada algoritma C4.5 memiliki akurasi sebesar 88,3%. Penelitian dilakukan untuk mengklasifikasi dan membantu menentukan yang layak mendapatkan beasiswa PPA dari Universitas Sebelas Maret.

Sumarlin S meneliti tentang Implementasi Algoritma K-Nearest Neighbor Sebagai Pendukung Keputusan Klasifikasi Penerima Beasiswa PPA dan BBM. Penelitian ini menghasilkan sebuah aplikasi Sistem Pendukung Keputusan untuk menentukan penerima beasiswa PPA dan BBM. Data yang digunakan adalah data penerima beasiswa PPA dan BBM tahun 2012/2013. Penerima PPA sebanyak 177 mahasiswa sedangkan penerima BBM sebanyak 133 mahasiswa dan sebanyak 60 mahasiswa tidak lolos seleksi keduanya. Pengujian menggunakan metode K-NN dengan cross validasi, confusion matrix dan kurva ROC. Akurasi yang diperoleh

untuk beasiswa PPA sebesar 88,33% dengan 227 dataset record, sedangkan untuk beasiswa BBM diperoleh akurasi sebesar 90% dengan 183 dataset record.

Penelitian M. Riyyan dan Firdaus melakukan Perbandingan Algoritme Naive Bayes Dan K-NN Terhadap Data Penerimaan Beasiswa (Studi Kasus Lembaga Beasiswa Baznas Jabar). Model klasifikasi yang digunakan akan diuji menggunakan confusion matrix. Hasil penelitian menunjukkan nilai yang berbeda pada setiap metode yang digunakan. Metode Naive Bayes memperoleh nilai akurasi sebesar 80%, presisi sebesar 82,82% dan recall sebesar 92,47%. Sedangkan metode K-NN memperoleh nilai lebih kecil dibandingkan Naive Bayes yaitu akurasi 78,97%, presisi 82,61% dan recall 91,10% dimana nilai K yang digunakan K=7.

Tabel 2.1 Penelitian terdahulu yang telah dilakukan

No	Peneliti	Judul	Metode	Hasil
1	(Mulyani et al., 2022)	The Prediction Of PPA And KIP Kuliah Scholarship Recipients Using Naive Bayes Algorithm	Naive Bayes	Nilai akurasi 80% untuk PPA dan 91% untuk KIP-K
2	(Suganda et al., 2022)	Penentuan Penerima Bantuan Beasiswa Kartu Indonesia Pintar (KIP) Kuliah Menggunakan Naive Bayes Classifier		Nilai akurasi 88,21%
3	(Ani & Andri, 2022)	Penerapan Algoritma Naive Bayes Untuk Klasifikasi Mahasiswa Penerima KIP Pada Universitas Bina Darma		Nilai akurasi 49,15%
4	(Sumiah & Mirantika, 2020)	Perbandingan Metode K-Nearest Neighbor dan Naive Bayes untuk Rekomendasi Penentuan Mahasiswa Penerima Beasiswa pada Universitas Kuningan	K-NN dan Naive Bayes	Nilai akurasi K-NN 100% dan Naive Bayes 99,89%
5	(Dahri et al., 2017)	Metode Naive Bayes Untuk Penentuan Penerima Beasiswa Bidikmisi Universitas Mulawarman	Naive Bayes	Nilai akurasi 85,56%
6	(Risman et al., 2015)	Penerapan Metode K-Nearest Neighbor Pada Aplikasi Penentu Penerima Beasiswa Mahasiswa Di STMIK Sinar Nusantara Surakarta	K-NN	Nilai akurasi 90,90%
7	(Noviana et al., 2019)	Analisis Rekomendasi Penerima Beasiswa Menggunakan Algoritma K-Nearest Neighbor (K-NN) dan	K-NN dan C4.5	Nilai akurasi K-NN 90,7% dan C4.5 88,3%

		Algoritma C4. 5		
8	(Sumarlin, 2015)	Implementasi Algoritma K-Nearest Neighbor Sebagai Pendukung Keputusan Klasifikasi Penerima Beasiswa PPA dan BBM	K-NN	Nilai akurasi PPA 88,33% dan BBM 90%
9	(Riyyan & Firdaus, 2022)	Perbandingan Algoritme Naive Bayes Dan K-NN Terhadap Data Penerimaan Beasiswa (Studi Kasus Lembaga Beasiswa Baznas Jabar)	Naive Bayes dan K-NN	Nilai akurasi Naive Bayes 80% dan K-NN 78,97%

2.2 Perbandingan Penelitian Terdahulu

Penelitian yang diusulkan oleh penulis dengan judul “Perbandingan Algoritma K-Nearest Neighbor (K-NN) dan Naive Bayes Pada Data Penerima Beasiswa Kartu Indonesia Pintar (KIP) Kuliah” memiliki beberapa kesamaan dan juga perbedaan dengan beberapa penelitian terdahulu yang telah dilakukan. Berdasarkan tabel di atas, beberapa kesamaan yang ditemukan antara penelitian terdahulu dengan penelitian yang penulis lakukan adalah:

- Metode yang sama
- Melakukan perbandingan metode yang sama
- Menggunakan data yang serupa yaitu penerima KIP Kuliah

Sedangkan perbedaan penelitian terdahulu dengan penelitian yang penulis lakukan ialah:

- Penelitian ini membandingkan dua metode
- Penelitian ini menggunakan dataset penerima KIP Kuliah di UIN Maulana Malik Ibrahim

Berikut ini tabel perbandingan penelitian terdahulu dengan penelitian yang sedang dilakukan oleh penulis berdasarkan Judul penelitian:

Tabel 2.2 Perbandingan penelitian terdahulu dengan penelitian sekarang

No	Peneliti terdahulu	Persamaan	Perbedaan
1	(Mulyani et al., 2022)	<ul style="list-style-type: none"> ▪ Metode Naive Bayes ▪ Data KIP Kuliah 	<ul style="list-style-type: none"> ▪ Perbandingan K-NN dan Naive Bayes ▪ KIP di UIN Malang
2	(Suganda et al., 2022)	<ul style="list-style-type: none"> ▪ Metode Naive Bayes ▪ Data KIP Kuliah 	Perbandingan metode
3	(Ani & Andri, 2022)	<ul style="list-style-type: none"> ▪ Metode Naive Bayes ▪ Beasiswa KIP di UBD 	<ul style="list-style-type: none"> ▪ Perbandingan metode ▪ Beasiswa KIP di UIN Malang
4	(Sumiah & Mirantika, 2020)	<ul style="list-style-type: none"> ▪ Perbandingan K-NN dan Naive Bayes ▪ Data beasiswa 	Data KIP
5	(Dahri et al., 2017)	<ul style="list-style-type: none"> ▪ Metode Naive Bayes 	<ul style="list-style-type: none"> ▪ Perbandingan metode ▪ Data KIP
6	(Risman et al., 2015)	<ul style="list-style-type: none"> ▪ Metode K-NN ▪ Data beasiswa 	<ul style="list-style-type: none"> ▪ Perbandingan metode ▪ Data KIP
7	(Noviana et al., 2019)	<ul style="list-style-type: none"> ▪ Metode K-NN ▪ Data beasiswa 	<ul style="list-style-type: none"> ▪ Perbandingan metode
8	(Sumarlin, 2015)	<ul style="list-style-type: none"> ▪ Metode K-NN 	<ul style="list-style-type: none"> ▪ Perbandingan metode
9	(Riyyan & Firdaus, 2022)	<ul style="list-style-type: none"> ▪ Perbandingan K-NN dan Naive Bayes ▪ Data Baznas 	Data KIP UIN Malang

Selain beberapa penelitian diatas, beberapa penelitian sebelumnya yang berkaitan dengan topik implementasi Naive Bayes telah banyak dilakukan. Seperti yang dilakukan oleh Budiman dan Umami yang melakukan penelitian di SMK YPM 14 Sumobito Jombang.(Budiman & Umami, 2022) M. R. Romadhon dkk menganalisis performa Naive Bayes, logistic regression, dan K-Nearest Neighbour untuk prediksi kesembuhan pasien covid-19.(Romadhon & Kurniawan, 2021) Sumarni adi mengimplementasikan Naive Bayes untuk klasifikasi penerima beasiswa di AMIKOM yang menghasilkan akurasi sebesar 97,66% dengan sampel 386 data.(Adi, 2018) Heni sulistiani menerapkan algoritma klasifikasi yaitu Naive Bayes dengan menghasilkan nilai akurasi sebesar 80%.(Sulistiani, 2018)

Berdasarkan beberapa penelitian yang telah dilakukan, dapat diketahui bahwa metode naive Bayes dan k-NN dapat digunakan untuk melakukan

klasifikasi kelayakan mahasiswa penerima beasiswa. Namun, beberapa penelitian hanya menggunakan salah satu metode saja dan menggunakan obyek penelitian yang berbeda. Oleh karena itu, dalam penelitian ini dilakukan penelitian untuk mengetahui perbandingan dari kedua metode tersebut dalam memprediksi kelayakan mahasiswa jalur KIP Kuliah.

2.3 Landasan Teori

2.3.1 Kartu Indonesia Pintar Kuliah

Kartu Indonesia Pintar Kuliah atau disingkat menjadi KIP Kuliah adalah bantuan sosial berupa biaya pendidikan yang diberikan oleh pemerintah kepada mahasiswa yang tidak mampu secara ekonomi dan memiliki potensi akademik baik untuk melanjutkan studi pada program sarjana (S1). (Rektor, 2022) KIP Kuliah adalah perluasan dari Program Indonesia Pintar (PIP) dimana PIP adalah bantuan berupa uang tunai, perluasan akses, dan kesempatan belajar dari pemerintah yang diberikan kepada peserta didik yang berasal dari keluarga tidak mampu. PIP semula hanya diberikan kepada anak yang menempuh pendidikan dasar dan menengah. Kemudian pada tahun 2020 PIP juga diberikan kepada mahasiswa dengan nama Kartu Indonesia Pintar Kuliah (KIP Kuliah). (M. Agama, 2020)

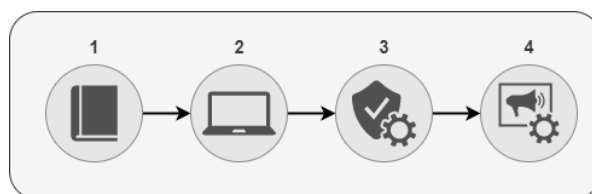
Universitas Islam Negeri Maulana Malik Ibrahim Malang sebagai salah satu PTKIN di Indonesia menjadi salah satu kampus yang menerima calon mahasiswa jalur KIP Kuliah. Sasaran program KIP Kuliah adalah untuk mahasiswa yang diterima di Universitas Islam Negeri Maulana Malik Ibrahim Malang yang

memenuhi persyaratan untuk mendapatkan KIP Kuliah. Sedangkan tujuan KIP Kuliah sebagaimana tertuang dalam Petunjuk Teknis KIP Kuliah adalah untuk:

- a. Meningkatkan akses dan kesempatan belajar bagi peserta didik yang tidak mampu secara ekonomi dan berpotensi akademik yang baik
- b. Meningkatkan prestasi dan motivasi belajar mahasiswa, khususnya bagi mereka yang memiliki keterbatasan ekonomi
- c. Menjamin keberlangsungan studi mahasiswa sampai tuntas dan tepat waktu, serta mampu berprestasi
- d. Melahirkan lulusan yang berkarakter, mandiri, produktif, dan memiliki kepedulian sosial, sehingga mampu memutus mata rantai kemiskinan

Alur pendaftaran KIP Kuliah Universitas Islam Negeri Maulana Malik Ibrahim Malang

1. Siswa membaca dengan seksama juknis pendaftaran kip kuliah uin. Beberapa hal yang perlu diperhatikan antara lain; ketentuan umum dan persyaratan
2. Siswa mendaftar secara online pada laman <http://kip.uin-malang.ac.id>.
Username: nomor peserta (sesuai dengan bukti pendaftaran jalur).
Password: tanggal lahir (thn-bln-tgl). Contoh: 1999-11-01
3. Panitia melakukan seleksi secara online dan juga melakukan verifikasi lapangan terkait kondisi sebenarnya yang sifatnya sampel
4. Siswa memperhatikan jadwal pengumuman kelayakan dan ketentuan lainnya, sebagaimana pada website pmb.uin-malang.ac.id



Gambar 2.1 Alur Pendaftaran KIP Kuliah UIN Maulana Malik Ibrahim Malang

2.3.2 Algoritma Naive Bayes

Algoritma Naive Bayes dimulai dengan teorema Bayes yang ditemukan oleh ahli matematika Inggris, Thomas Bayes, pada tahun 1763. (Zhang, 2018) Penerapan teorema Bayes untuk masalah klasifikasi muncul pada tahun 1950-an, ketika algoritma ini digunakan dalam sistem pengenalan tulisan tangan. Pada 1980-an, algoritma Naive Bayes mulai digunakan dalam pengenalan suara dan teknologi pengolah kata. Pada tahun 1990-an, algoritma ini mulai digunakan dalam machine learning dan digunakan dalam banyak aplikasi, seperti pengenalan wajah, penyaringan spam, dan klasifikasi teks.

Naive Bayes adalah sebuah metode klasifikasi yang berakar pada Teorema Bayes. Karakteristik utama dari Naive Bayes adalah asumsi (naif) yang sangat kuat akan independensi dari setiap kondisi atau peristiwa. Sebelum menjelaskan Naive Bayes, perlu dijelaskan Teorema Bayes terlebih dahulu yang menjadi landasan dari metode Naive Bayes. (Wibawa et al., 2018)

Teorema Bayes adalah rumus matematika yang digunakan untuk menghitung probabilitas posterior suatu peristiwa berdasarkan probabilitas a priori dan probabilitas likelihood. Pada teorema Bayes, jika terdapat dua kejadian yang terpisah, misalkan A dan B, maka teorema Bayes dapat dirumuskan seperti berikut:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Persamaan Teorema
Bayes(Ariffin & Syalwah, 2020)

dimana

$P(A|B)$ = Probabilitas hipotesis A berdasarkan kondisi B (posteriori probability)

$P(B|A)$ = Probabilitas hipotesis B berdasarkan kondisi hipotesis A (likelihood)

$P(A)$ = Probabilitas hipotesis A (prior probability)

$P(B)$ = Probabilitas dari B (evidence)

2.3.3 Algoritma K-Nearest Neighbor

K-Nearest Neighbor (K-NN) adalah salah satu algoritma yang digunakan dalam penyelesaian masalah klasifikasi. Prinsip operasi K-NN adalah mencari jarak terpendek antara data yang akan diuji dengan tetangga terdekat dalam data latih.(Nikmatun & Waspada, 2019) Algoritma K-Nearest Neighbor (K-NN) merupakan salah satu algoritma yang paling sederhana untuk menyelesaikan masalah klasifikasi dan seringkali menghasilkan hasil yang kompetitif dan signifikan.(Adeniyi et al., 2018) Untuk menghitung jarak, dapat menggunakan *Euclidean distance*. Rumus *Euclidean distance* ditentukan dalam persamaan berikut:(Id, 2021)

$$d(p, q) = \sqrt{(p_1 - q_1)^2 + \dots + (p_n - q_n)^2}$$

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

Rumus Euclidian
Distance

dimana

p = data latih

i = variabel data

n = dimensi data

q = data uji

d = jarak

Alur penghitungan k-Nearest Neighbor

- a. Menentukan nilai K. Nilai K sebaiknya bilangan asli ganjil agar tidak terjadi hasil voting seimbang.
- b. Menghitung jarak euclidian antara data uji dan semua data training
- c. Tentukan tetangga terdekat dengan data uji
- d. Pada kasus klasifikasi, dilakukan penghitungan jumlah titik terdekat berdasarkan kelas. Titik yang baru akan masuk ke kategori yang memiliki tetangga terdekat terbanyak.(Id, 2021)

Tabel 2.3 Kelebihan dan kekurangan algoritma K-NN dan Naive Bayes

Kelebihan	Kekurangan
Algoritma Naive Bayes	
<ol style="list-style-type: none"> 1. Sederhana, mudah, cepat, dan sangat kuat 2. Bekerja sangat baik pada data yang clean atau noisy 3. Membutuhkan data training yang sedikit, dengan asumsi data training yang diambil merupakan representasi dari populasi 4. Mudah untuk menentukan probabilitas prediksi 	<ol style="list-style-type: none"> 1. Umumnya tidak cocok untuk dataset dengan atribut numerik besar 2. Probabilitas yang diprediksi kurang dapat diandalkan 3. Dalam beberapa kasus, ada data yang tidak terdapat dalam training tetapi muncul dalam data testing, sehingga probabilitas yang dihasilkan menjadi salah
Algoritma K-Nearest Neighbor	
<ol style="list-style-type: none"> 1. Dapat digunakan untuk klasifikasi biner dan klasifikasi multiclass dengan baik 2. Metodenya cukup mudah untuk dipahami dan dilaksanakan 3. Metodenya kuat dan tidak mempengaruhi outlier ketika nilai K besar 	<ol style="list-style-type: none"> 1. Akurasi bergantung pada nilai K sehingga terkadang kesulitan mencari nilai yang paling optimal 2. Metode ini bergantung distribusi kelas agar tidak overlapping 3. Tidak adanya output spesifik sebagai model dan jika nilai K kecil, itu dipengaruhi secara negatif oleh outlier 4. Metode ini melakukan perhitungan secara intensif karena disebut lazy learner. Jarak dihitung antar semua poin, oleh karenanya metode ini bukan metode yang cepat.(Verdhan, 2020)

2.3.4 Kajian Integrasi Sains dan Islam

Menurut perspektif Al-Qur'an, pemberian beasiswa juga dapat dikategorikan sebagai bentuk syafaat yang dilakukan seseorang atau lembaga kepada orang lain yang membutuhkan. Sebagaimana telah diperingatkan oleh Al-Qur'an dalam Surat An-Nisa' ayat 85

مَنْ يَشْفَعُ شَفَاعَةً حَسَنَةً يَكُنْ لَهُ نَصِيبٌ مِّنْهَا ۚ وَمَنْ يَشْفَعُ شَفَاعَةً سَيِّئَةً يَكُنْ لَهُ كِفْلٌ مِّنْهَا ۗ
وَكَانَ اللَّهُ عَلَىٰ كُلِّ شَيْءٍ مُّقْتَدِرًا

“Siapa yang memberi pertolongan yang baik niscaya akan memperoleh bagian (pahala) darinya. Siapa yang memberi pertolongan yang buruk niscaya akan menanggung bagian (dosa) darinya. Allah Maha Kuasa atas segala sesuatu.”

Ayat di atas berisi penjelasan bahwa orang yang melakukan syafaat berbentuk kebajikan umpamanya menolong atau menganjurkan kepada orang lain melakukan perbuatan baik, orang yang menganjurkan akan mendapat ganjaran dari perbuatan orang yang mengikuti anjurannya tersebut seolah-olah ia sendiri yang berbuat. Demikian juga orang yang melakukan syafaat berbentuk kejahatan. Ia akan mendapat bagian ganjaran dari perbuatan tersebut seolah-olah ia berserikat dalam pekerjaan itu. (D. Agama, 2011) Menurut Wahbah Zuhaili dalam Tafsir Wajiz, orang yang melakukan sesuatu yang mencegah hak orang lain akan menerima bagian dari dosa pertolongan buruk yang dilakukan. (Zuhaili, n.d.)

Pemberian beasiswa juga merupakan salah satu bentuk sikap tolong-menolong dalam kebaikan. Perintah untuk saling tolong menolong dalam kebaikan juga telah dijelaskan oleh Al-Qur'an. Allah memberikan perintah kepada orang-orang mukmin untuk tolong-menolong atau berserikat dalam kebaikan yaitu dalam menjalankan ketaatan kepada Allah. Allah melarang untuk saling tolong-menolong dalam perbuatan dosa yang dapat menjauhkan dari hukum-hukum positif agama. Sebagaimana dijelaskan dalam Al-Qur'an Surat Al-Maidah ayat 2:

وَتَعَاوَنُوا عَلَىٰ الْبِرِّ وَالتَّقْوَىٰ ۖ وَلَا تَعَاوَنُوا عَلَىٰ الْإِثْمِ وَالْعُدْوَانِ ۗ وَاتَّقُوا اللَّهَ ۖ إِنَّ اللَّهَ شَدِيدُ الْعِقَابِ

“Dan tolong-menolonglah kamu dalam (mengerjakan) kebajikan dan takwa, dan jangan tolong-menolong dalam berbuat dosa dan permusuhan. Bertakwalah kepada Allah, sungguh, Allah sangat berat siksaan-Nya.”

Beasiswa merupakan salah satu bentuk dukungan finansial kepada para pencari ilmu. Karena tidak semua pencari ilmu memiliki kecukupan finansial sehingga tidak dapat mengkases ilmu seperti orang yang berkecukupan. Padahal menuntut ilmu merupakan sebuah keharusan bagi setiap muslim. Sebagaimana hadits Nabi SAW: (القزويني, n.d.)

حَدَّثَنَا هِشَامُ بْنُ عَمَّارٍ حَدَّثَنَا حَفْصُ بْنُ سُلَيْمَانَ حَدَّثَنَا كَثِيرُ بْنُ شَنْظِيرٍ عَنْ مُحَمَّدِ بْنِ سِيرِينَ عَنْ أَنَسِ بْنِ مَالِكٍ قَالَ قَالَ رَسُولُ اللَّهِ صَلَّى اللَّهُ عَلَيْهِ وَسَلَّمَ طَلَبُ الْعِلْمِ فَرِيضَةٌ عَلَى كُلِّ مُسْلِمٍ وَوَأَضِعُ الْعِلْمَ عِنْدَ غَيْرِ أَهْلِهِ كَمَقْلَدِ الْخَنَازِيرِ الْجَوْهَرَ وَاللُّؤْلُؤَ وَالذَّهَبَ

Telah menceritakan kepada kami Hisyam bin Ammar, telah menceritakan kepada kami Hafsh bin Sulaiman, telah menceritakan kepada kami Katsir bin Syinzhir dari Muhammad bin Sirin dari Anas bin Malik ia berkata; Rasulullah SAW bersabda: "Menuntut ilmu adalah kewajiban bagi setiap muslim. Dan orang yang meletakkan ilmu bukan pada pada ahlinya, seperti seorang yang mengalungkan mutiara, intan dan emas ke leher babi."

Namun penuntut ilmu harus memperhatikan enam syarat sebagaimana dijelaskan oleh Imam Syafi'i:

أَخِي لَنْ تَنَالَ الْعِلْمَ إِلَّا بِسِتَّةٍ * سَأُنْبِيكَ عَنْ تَفْصِيلِهَا بِبَيَانٍ
ذِكَاؤٌ وَحِرْصٌ وَاجْتِهَادٌ وَبُلْغَةٌ * وَصُحْبَةٌ أُسْتَاذٍ وَطَوَّلُ زَمَانٍ

“Saudaraku, ilmu tidak akan diperoleh kecuali dengan enam perkara yang akan saya beritahukan perinciannya yaitu kecerdasan, semangat, sungguh-sungguh, berkecukupan, bersahabat (belajar) dengan ustadz (guru), dan membutuhkan waktu yang lama”.(الشافعي, n.d.)

Salah satu syarat orang mencari ilmu sebagaimana keterangan Imam Syafi'I ialah memiliki kecukupan finansial. Hal ini sangat penting karena adanya biaya dapat

digunakan untuk membeli peralatan belajar atau memberikan upah untuk pengajar.

Penyelesaian masalah dengan Teknik klasifikasi merupakan salah satu metode dalam machine learning yang digunakan untuk menemukan pengetahuan baru berdasarkan kejadian masa lampau. Konsep ini juga dapat ditemukan inspirasinya dalam Al-Qur'an Surat Al-Lail. Allah melakukan klasifikasi golongan manusia menjadi dua. Pertama ialah orang yang memudahkan jalan kebahagiaannya dan kedua orang yang memudahkan jalan kesengsaraannya berdasarkan beberapa ketentuan sebagaimana dijelaskan dalam Surat Al-Lail:

فَأَمَّا مَنْ أَعْطَىٰ وَاتَّقَىٰ ۖ وَصَدَّقَ بِالْحُسْنَىٰ ۖ فَسَنُيَسِّرُهُ لِلْيُسْرَىٰ ۖ وَأَمَّا مَنْ بَخِلَ وَاسْتَغْنَىٰ ۖ ۝٨
وَكَذَّبَ بِالْحُسْنَىٰ ۖ فَسَنُيَسِّرُهُ لِلْعُسْرَىٰ ۖ ۝٩

“5. Maka barangsiapa memberikan (hartanya di jalan Allah) dan bertakwa 6. dan membenarkan (adanya pahala) yang terbaik (surga). 7. maka akan Kami mudahkan baginya jalan menuju kemudahan (kebahagiaan). 8. dan adapun orang yang kikir dan merasa dirinya cukup (tidak perlu pertolongan Allah). 9. serta mendustakan (pahala) yang terbaik. 10. maka akan Kami mudahkan baginya jalan menuju kesukaran (kesengsaraan)” (Q.S. Al-Lail:5-10)

Kata *أَعْطَىٰ* bermakna mendermakan harta. pada Surat Al-Lail ayat 5 – 10 di atas terdapat muqabalah (perbandingan) yaitu pada ayat *فَأَمَّا مَنْ أَعْطَىٰ وَاتَّقَىٰ* dan ayat *وَأَمَّا مَنْ بَخِلَ وَاسْتَغْنَىٰ*. Perbandingan tersebut ditujukan untuk orang-orang yang dermawan dan orang yang pelit. Sebab turunnya ayat 5 surat Al-Lail adalah berkaitan dengan abu bakar yang memerdekakan orang-orang lemah dan wanita jika mereka memeluk islam. (Wahbah, 2013)

Surat Al-Lail ayat 1 – 5 memberikan pesan bahwa orang-orang yang mendermakan hartanya di jalan Allah dan membenarkan adanya balasan

kebaikan, mereka akan mendapatkan kemudahan dalam langkahnya melakukan kebaikan. Adapun orang yang kikir dan tidak membenarkan adanya pahala dari Allah, maka Allah akan menjadikan kemudahan baginya jalan untuk berbuat keburukan.

Apabila dikaitkan dengan pemberian beasiswa, ayat di atas memberikan pesan kepada para pengelola beasiswa untuk melakukan pembagian beasiswa dengan jujur, adil dan tepat sasaran. Kejujuran dan keadilan dalam pendistribusian beasiswa termasuk langkah dalam mendermakan harta dengan benar, sehingga Allah akan memberikan balasan kebaikan yang setimpal dengan kebaikan yang telah dilakukan. Namun jika pemberian beasiswa tidak dilakukan dengan Amanah dan tepat sasaran, maka sama halnya dengan mendustakan adanya balasan pahala dari Allah dan justru akan dipersulit urusannya terutama dalam melakukan kebaikan karena tidak percaya dengan balasan Allah.

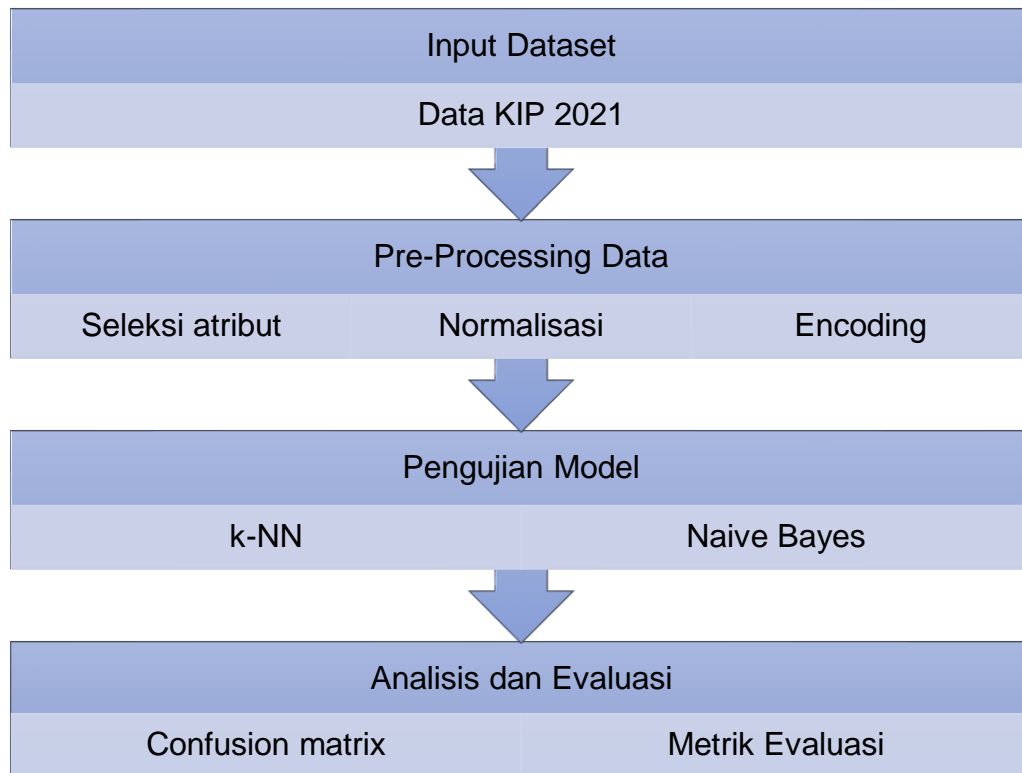
Proses klasifikasi juga akan menghasilkan sebuah rekomendasi kelas berdasarkan atribut atau kriteria tertentu. Seperti proses penentuan kelayakan penerima beasiswa yang membutuhkan atribut tertentu sehingga menghasilkan sebuah kesimpulan diterima atau ditolakanya calon pendaftar. Ayat di atas mengindikasikan dua kelas target yaitu *al-yusra* dan *al-'usra*. Adapun atribut yang digunakan diantaranya adalah dermawan atau bakhil, dan percaya atau tidak percaya adanya balasan pahala. Hal ini menunjukkan bahwa proses klasifikasi membutuhkan atribut-atribut yang relevan dalam proses penentuan kelas target atau rekomendasi.

BAB III

METODOLOGI PENELITIAN

3.1 Desain Penelitian

Penelitian ini menggunakan metode kuantitatif dengan pendekatan komparatif. Data yang digunakan dalam penelitian ini adalah data pendaftar beasiswa jalur KIP Kuliah dari Universitas Islam Negeri (UIN) Maulana Malik Ibrahim Malang tahun 2022. Data ini merupakan data sekunder yang diperoleh dari website resmi KIP Kuliah UIN Maulana Malik Ibrahim <https://kip.uin-malang.ac.id/> yang dikelola oleh Panitia Penerimaan Mahasiswa Baru <http://pmb.uin-malang.ac.id/>. Data ini akan diolah sehingga menghasilkan output sebagaimana tujuan penelitian yang disebutkan penulis.



Gambar 3.1 Diagram Alur Penelitian

Gambar 3.1 menampilkan Diagram langkah-langkah yang dilakukan dalam penelitian ini. Berdasarkan gambar tersebut, langkah-langkah yang dilakukan untuk melakukan penelitian adalah Input Data, Pre-processing data, Pengujian Model, Analisis dan Evaluasi:

3.1.1. Input Dataset

Data input yang digunakan adalah data penerima beasiswa KIP UIN Maulana Malik Ibrahim Malang tahun 2021. Data ini terdiri dari berbagai atribut, seperti id, gender, domisili, peng_ayah, peng_ibu, pekerjaan, tang_ortu, rerata_un, rerata_rapor, pres_aka, pres_non_aka, luas_rmh, pbb, rek_listrik, rekom, jalur. Data yang digunakan dalam penelitian ini bertipe data kategori.

3.1.2. Pre-processing data

Data yang telah diinput akan diolah terlebih dahulu untuk menghilangkan atau melakukan penyesuaian data yang tidak valid. Selain itu juga data yang tidak relevan tidak akan digunakan dalam penelitian ini. Pre-prosesing yang dilakukan dalam penelitian ini adalah normalisasi data, seleksi atribut, encoding data, dan splitting data. Setiap tahapan pre-processing yang dilakukan diharapkan dapat menghasilkan data yang berkualitas dan siap untuk digunakan oleh model klasifikasi.

3.1.3. Pengujian model

Data yang telah dilakukan pre-prosesing akan dimodelkan menggunakan algoritma klasifikasi dalam machine learning. Algoritma yang digunakan dalam penelitian ini adalah k-NN dan Naive Bayes. Kedua algoritma tersebut akan diuji menggunakan data penerima KIP untuk mengetahui performa kedua algoritma.

3.1.4. Analisis dan evaluasi

Hasil prediksi dari k-NN dan Naive Bayes akan dievaluasi dengan menggunakan data testing. Evaluasi yang dilakukan untuk mengukur performa kedua algoritma menggunakan metrik evaluasi yaitu akurasi, presisi, dan recall. Hasil evaluasi setiap akan dicari algoritma yang lebih unggul dalam menangani dataset penerima KIP.

3.2 Instrumen Penelitian

Penelitian ini dilaksanakan di Universitas Islam Negeri (UIN) Maulana Malik Ibrahim Malang Jalan Gajayana 50 Malang. Kebutuhan bahan dan alat yang digunakan dalam penelitian ini adalah:

- 3.2.1 Laptop dengan spesifikasi AMD Ryzen 3 3200U 2.60 GHz, RAM 12 GB, Integrated AMD Radeon Vega 3 Graphics, SSD 512 GB PCIe
- 3.2.2 Software atau Tools yaitu Windows 11 Home Single, Microsoft Excel Home and Student 2019, Jupyter Notebook.

3.3 Pengumpulan Data

Data yang digunakan dalam penelitian ini berasal dari sumber-sumber sebagai berikut:

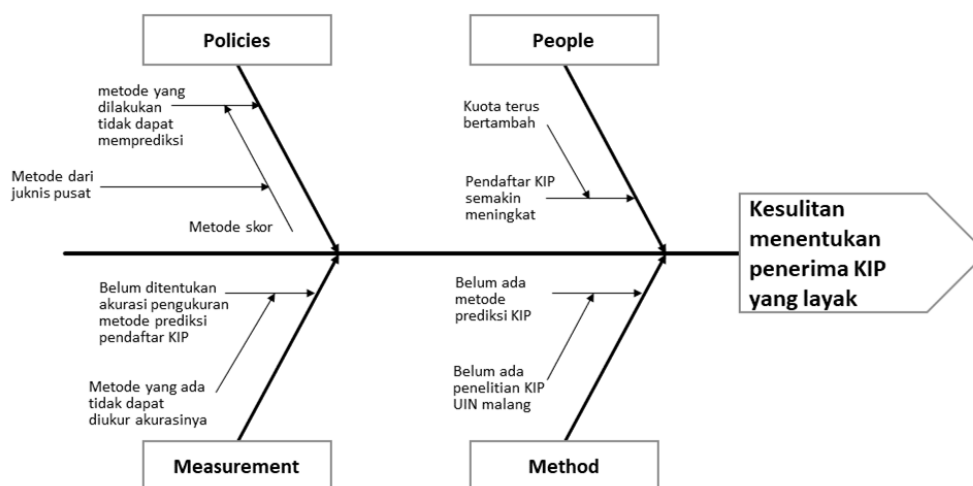
- 3.3.1 Dokumentasi Petunjuk Teknis KIP Kuliah UIN Maulana Malik Ibrahim tahun 2022
- 3.3.2 Wawancara dengan Panitia Penerimaan Mahasiswa Baru (PMB) UIN Maulana Malik Ibrahim bagian Pengolah data

3.3.3 Data pendaftar beasiswa KIP Kuliah di UIN Maulana Malik Ibrahim jalur Mandiri tulis, Mandiri prestasi, SNMPTN, SBMPTN, SPAN-PTKIN, dan UM-PTKIN tahun 2022.

3.4 Analisis *Fishbone*

Analisis fishbone adalah analisis permasalahan yang berpedoman pada diagram fishbone. Fishbone diagram adalah salah satu metode yang digunakan untuk menemukan akar masalah. Fishbone diagram kadang juga disebut cause and effect diagram atau diagram sebab akibat, atau juga dapat disebut Ishikawa diagram. Penyebutan Ishikawa diagram karena sesuai dengan penemunya yaitu karou Ishikawa seorang ilmuwan Jepang tahun 1943.(University, 2017)

Gambar 3.2 merupakan analisis fishbone pada penelitian prediksi kelayakan calon penerima beasiswa KIP Kuliah menggunakan K-NN dan Naive Bayes. Kemudian dari permasalahan yang teridentifikasi dalam diagram fishbone dijabarkan strategi penyelesaiannya pada tabel identifikasi permasalahan.



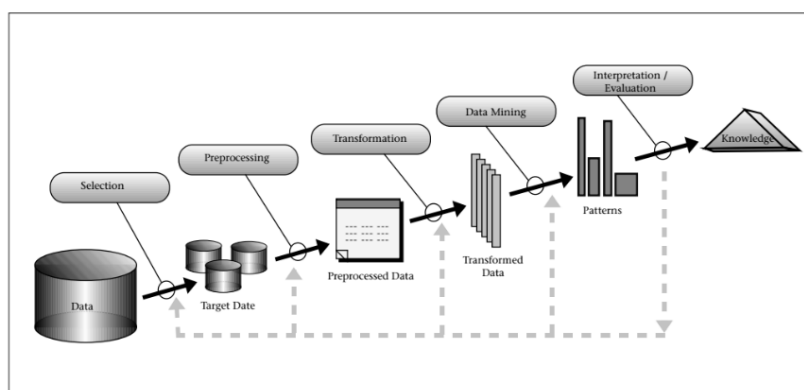
Gambar 3.2 diagram *fishbone*

Tabel 3.1 Identifikasi akar permasalahan berdasarkan diagram fishbone

No	Aspek-aspek yang diamati	Permasalahan	Strategi
1	Method	Belum ditemukan penelitian yang melakukan prediksi kelayakan penerima KIP di UIN Malang	Melakukan penelitian untuk memprediksi kelayakan dengan Teknik klasifikasi dalam Data Mining
2	People	Pendaftar beasiswa KIP di UIN Malang terus bertambah setiap tahun karena kuota beasiswa juga meningkat	Melakukan analisis terhadap data para penerima KIP sebelumnya sebagai evaluasi dalam menentukan penerima yang layak dan tepat sasaran dimasa yang akan datang
3	Measurement	Pengukuran akurasi metode penentuan kelayakan penerima KIP tidak dapat dilakukan karena metode penghitungan kelayakan yang dilakukan saat ini di UIN Malang menggunakan skor	Menggunakan metode klasifikasi seperti K-NN dan Naive Bayes yang dapat diukur tingkat akurasi dengan metode confusion matrix
4	Policies	Metode yang telah dilakukan oleh kampus untuk menentukan kelayakan penerima KIP tidak dapat melakukan prediksi karena metode yang digunakan merupakan kebijakan dari pusat yang tertuang dalam juknis	Menmbangun sebuah model prediksi untuk menentukan kelayakan penerima KIP sebagai bahan pertimbangan pihak kampus dalam mengambil keputusan yang efektif.

3.5 Analisis Data

Analisis data dalam penelitian ini mengikuti tahapan-tahapan yang digunakan dalam proses Knowledge Discovery in Database (KDD) sebagaimana yang ditunjukkan dalam gambar 3.3. Adapun tahapan-tahapan yang dilakukan ialah Selection, Preprocessing, Transformation, Data Mining, dan Evaluation.



Gambar 3.3 Langkah-langkah dalam proses KDD (Fayyad et al., 1996)

3.5.1 Data Selection

Data yang akan diolah dalam penelitian ini berasal dari data pendaftar beasiswa KIP Kuliah tahun 2022 dari jalur Mandiri tulis sebanyak 536 pendaftar, Mandiri prestasi sebanyak 57 pendaftar, SNMPTN sebanyak 64 pendaftar, SBMPTN sebanyak 77 pendaftar, SPAN-PTKIN sebanyak 64 pendaftar, dan UM-PTKIN sebanyak 246 pendaftar.

Tabel 3.2 data jumlah pendaftar KIP Kuliah

No	Data Jalur	Record
1	Mandiri-tulis	536
2	Mandiri-prestasi	57
3	SNMPTN	64
4	SBMPTN	77
5	SPAN-PTKIN	64
6	UM-PTKIN	246
Total		1044

Berdasarkan tabel di atas, pendaftar beasiswa KIP Kuliah dari berbagai jalur sebanyak 1044 data peserta. Jumlah tersebut merupakan gabungan dari pendaftar dari jalur enam jalur seperti yang terlihat di tabel. Adapun atribut yang dimiliki oleh setiap pendaftar berjumlah 17 atribut. Rincian atribut dan nilai yang dimiliki oleh setiap pendaftar dapat dilihat pada tabel di bawah ini:

Tabel 3.3 Atribut dan nilai data pendaftar KIP Kuliah

No	Atribut	Nilai
1	No peserta	Nomor pendaftaran
2	Nama	Nama lengkap pendaftar
3	Gender	1. P 2. L
4	Domisili	Kabupaten dan Provinsi pendaftar
5	Penghasilan ayah	1. alm 2. < 3 juta 3. $3 \leq x \leq 4,5$ juta 4. $4,5 < x \leq 6$ juta
6	Penghasilan ibu	1. almh 2. < 3 juta 3. $3 \leq x \leq 4,5$ juta 4. $4,5 < x \leq 6$ juta
7	Pekerjaan	1. alm 2. Buruh tani/petani 3. PNS/TNI/POLRI 4. Tukang 5. Wiraswasta

8	Tanggungan orang tua	1. < 5 orang 2. 5 - 6 orang	3. 7 - 8 orang 4. > 8 orang
9	Rerata UN	1. < 6 2. $6 \leq x \leq 7$ 3. > 8	4. $7 < x \leq 8$ 5. Belum UAN
10	Rerata raport	1. < 7 2. $7 \leq x \leq 8$	3. $8 < x \leq 9$ 4. > 9
11	Prestasi akademik	1. Tidak ada 2. Tingkat lokal 3. Tingkat regional	4. Tingkat nasional 5. Tingkat internasional
12	Prestasi non akademik	1. Tidak ada 2. Tingkat lokal 3. Tingkat regional	4. Tingkat nasional 5. Tingkat internasional
13	Luas rumah	1. < 40 m ² 2. $41 \leq x \leq 80$ m ²	3. $80 < x \leq 120$ m ² 4. > 120 m ²
14	PBB	1. < 100 ribu 2. $100 \leq x \leq 200$ ribu	3. $200 \leq x \leq 300$ ribu 4. > 300 ribu
15	Rek listrik	1. < 100 ribu 2. $100 \leq x \leq 200$ ribu	3. $200 \leq x \leq 300$ ribu 4. > 300 ribu
16	Skor	0 - 100	
17	Rekomendasi	1. Diterima 2. Ditolak	

Proses seleksi data untuk menentukan atribut mana saja yang akan digunakan dalam klasifikasi dilakukan dengan melakukan wawancara kepada pihak terkait yang mengelola beasiswa KIP Kuliah. Dalam hal ini peneliti melakukan wawancara kepada bagian kemahasiswaan dan alumni, dan kepada PTIPD. Kemudian dari 17 atribut yang tersedia akan dihapus beberapa atribut sehingga menyisakan 13 atribut yang akan digunakan. Nama-nama atribut yang akan digunakan ialah:

- | | | |
|---------------------|----------------------|----------------------|
| 1. ID | 6. Rerata UN | 10. Luas rumah |
| 2. Penghasilan ayah | 7. Rerata raport | 11. PBB |
| 3. Penghasilan ibu | 8. Prestasi akademik | 12. Rekening listrik |

- | | | |
|-------------------------|-----------------|-----------------|
| 4. Pekerjaan | 9. Prestasi non | 13. Rekomendasi |
| 5. Tanggungan orang tua | akademik | |

3.5.2 Data preprocessing

Aktivitas yang dilakukan dalam tahapan preprocessing dalam penelitian ini diantaranya adalah memperbaiki missing value, seperti pada atribut rekomendasi dimana banyak kolom yang tidak bernilai atau null. Selain itu memperbaiki kesalahan yang ditemukan pada data seperti kesalahan penulisan.

3.5.3 Data transformation

Proses transformasi data yang dilakukan adalah melakukan encoding nilai-nilai pada atribut data agar dapat diolah menggunakan algoritma klasifikasi. Proses encoding yang dilakukan berdasarkan range yang telah ditentukan oleh pihak pengelola beasiswa KIP Kuliah UIN Maulana Malik Ibrahim. Atribut yang nilai-nilainya akan dilakukan transformasi ialah penghasilan ayah dan ibu, pekerjaan orang tua, tanggungan orang tua, rerata UN, rerata rapor, prestasi akademik dan non akademik, PBB, listrik, dan rekomendasi. Tabel di bawah ini merupakan rincian proses transformasi yang dilakukan pada data pendaftar beasiswa KIP Kuliah.

Tabel 3.4 Proses transformasi data beasiswa KIP Kuliah

No	Atribut	Nilai	Kategori
1.	Penghasilan ayah dan ibu	alm	4
		< 3 juta	3
		$3 \leq x \leq 4,5$ juta	2
		$4,5 < x \leq 6$ juta	1
2.	Pekerjaan	Alm	5
		Tukang	4
		Buruh tani/petani	3

		Wiraswasta	2
		PNS/TNI/POLRI	1
3.	Tanggungan orang tua	< 5 orang	1
		5 - 6 orang	2
		7 - 8 orang	3
		> 8 orang	4
4.	Rerata UN	< 6	5
		$6 \leq x \leq 7$	4
		> 8	3
		$7 < x \leq 8$	2
		Belum UAN	1
5.	Rerata rapor	< 7	1
		$7 \leq x \leq 8$	2
		$8 < x \leq 9$	3
		> 9	4
6.	Prestasi akademik	Tidak ada	1
		Tingkat lokal	2
		Tingkat regional	3
		Tingkat nasional	4
		Tingkat internasional	5
7.	Prestasi non akademik	Tidak ada	1
		Tingkat lokal	2
		Tingkat regional	3
		Tingkat nasional	4
		Tingkat internasional	5
8.	Luas rumah	< 40 m ²	4
		$41 \leq x \leq 80$ m ²	3
		$80 < x \leq 120$ m ²	2
		> 120 m ²	1
9.	PBB	< 100 ribu	4
		$100 \leq x \leq 200$ ribu	3
		$200 \leq x \leq 300$ ribu	2
		> 300 ribu	1
10.	Rek listrik	< 100 ribu	4
		$100 \leq x \leq 200$ ribu	3
		$200 \leq x \leq 300$ ribu	2
		> 300 ribu	1
11.	Rekomendasi	Diterima	1
		Ditolak	0

3.5.4 Data Mining

Tahapan selanjutnya dalam proses KDD adalah Data mining. Proses Data mining dilakukan dengan Teknik klasifikasi menggunakan algoritma K-NN dan

Naive Bayes. Berikut ini implementasi algoritma Naive Bayes yang digunakan pada data pendaftar beasiswa KIP Kuliah UIN Maulana Malik Ibrahim:

- Tabel 3.5 merupakan Data Pendaftar KIP kuliah yang terdiri dari beberapa variabel atau atribut. Baris pertama menunjukkan nama variabel. Variabel-variabel pendaftar KIP diantaranya yaitu Peng_ayah, Pekerjaan, Tang_ortu, Nilai_rapor, Pres_aka, Rek_listrik, Peng_ibu, Nilai_un, Pres_non_aka, Luas_rmh, Pbb, Rekom.
- Tipe data pada pendaftar KIP adalah tipe data kategorikal. Untuk dapat digunakan dalam pemodelan Naive Bayes menggunakan Python, perlu dilakukan proses *encoding data* sehingga menghasilkan data numerik. Proses transformasi data dari kategorikal menjadi numerik di Python dilakukan dengan metode Ordinal Encoder. Metode Ordinal Encoder mengubah string menjadi bilangan bulat ordinal. Hasil transformasi data dari menjadi tipe data numerik dapat dilihat pada tabel 3.6.

Tabel 3.5 Data Pendaftar KIP

Peng_ayah	Pekerjaan	Tang_ortu	Rapor	Pres_aka	Rek_listrik	Pengibu	Un	Pres_non_aka	Luas_rmh	Pbb	Rekom
< 3 Juta	Buruh Tani / Petani	> 8 orang	8,01 - 9	Tingkat Lokal	< 100 ribu	< 3 Juta	> 8	Tidak Ada	< 40 m2	< 100 ribu	Ditolak
< 3 Juta	Buruh Tani / Petani	< 5 orang	> 9	Tingkat Lokal	< 100 ribu	< 3 Juta	> 8	Tingkat Regional	< 40 m2	< 100 ribu	Diterima
< 3 Juta	Buruh Tani / Petani	< 5 orang	8,01 - 9	Tingkat Regional	< 100 ribu	< 3 Juta	7,01 - 8	Tingkat Nasional	< 40 m2	< 100 ribu	Diterima
< 3 Juta	Buruh Tani / Petani	7 - 8 orang	8,01 - 9	Tingkat Nasional	< 100 ribu	< 3 Juta	Belum UAN	Tidak Ada	< 40 m2	< 100 ribu	Ditolak
< 3 Juta	Buruh Tani / Petani	5 - 6 orang	8,01 - 9	Tingkat Lokal	< 100 ribu	< 3 Juta	> 8	Tidak Ada	< 40 m2	< 100 ribu	Diterima
< 3 Juta	Buruh Tani / Petani	< 5 orang	8,01 - 9	Tingkat Lokal	< 100 ribu	< 3 Juta	> 8	Tingkat Nasional	41 - 80 m2	< 100 ribu	Diterima
< 3 Juta	Buruh Tani / Petani	< 5 orang	> 9	Tingkat Lokal	100 - 200 ribu	< 3 Juta	> 8	Tingkat Nasional	< 40 m2	100 - 200 ribu	Diterima
< 3 Juta	Buruh Tani / Petani	< 5 orang	8,01 - 9	Tingkat Regional	< 100 ribu	< 3 Juta	> 8	Tidak Ada	< 40 m2	< 100 ribu	Diterima
< 3 Juta	Buruh Tani / Petani	< 5 orang	8,01 - 9	Tingkat Lokal	< 100 ribu	< 3 Juta	> 8	Tingkat Nasional	< 40 m2	< 100 ribu	Ditolak

Tabel 3.6 Hasil Encode Data Pendaftar KIP

Peng_ayah	Pekerjaan	Tang_ortu	Rerata_rapor	Pres_aka	Rek_listrik	Peng_ibu	Rerata_un	Pres_non_aka	Luas_rmh	Pbb	Rekom
3	1	3	1	2	2	3	3	0	2	2	1.0
3	1	2	3	2	2	3	3	4	2	2	0.0
3	1	2	1	4	2	3	1	3	2	2	0.0
3	1	1	1	3	2	3	4	0	2	2	1.0
3	1	0	1	2	2	3	3	0	2	2	0.0
3	1	2	1	2	2	3	3	3	0	2	0.0
3	1	2	3	2	0	3	3	3	2	0	0.0
3	1	2	1	4	2	3	3	0	2	2	0.0
3	1	2	1	2	2	3	3	3	2	2	1.0

- Ditanya :

Tentukan rekomendasi kelayakan Pendaftar x jika diketahui data

pendukungnya seperti berikut:

Peng_ayah	Pekerjaan	Tang_ortu	Rerata_rapor	Pres_aka	Rek_listrik	Peng_ibu	Rerat_a_un	Pres_no_n_aka	Luas_rmh	Pbb	Rekom
< 3 Juta	Buruh Tani / Petani	> 8 orang	8,01 - 9	Tingkat Lokal	< 100 ribu	< 3 Juta	> 8	Tidak Ada	< 40 m2	< 100 ribu	?

Langkah pertama adalah membuat tabel frekuensi dan menentukan prior probability, dan evidence setiap atribut kelas. Disini akan dibuat tabel frekuensi dari atribut Pekerjaan, tang_ortu, dan Nilai_UN

Tabel frekuensi Pekerjaan	Class		Total	Evidence P(x)
	Tidak layak	Layak		
Buruh Tani/Petani	2	4	6	6/6
Total	2	4	6	
Prior P(c)	2/6	4/6		

Tabel frekuensi tang_ortu	Class		Total	Evidence P(x)
	Tidak layak	Layak		
<5 orang	0	3	3	3/6
5 - 6 orang	0	1	1	1/6
7 - 8 orang	1	0	1	1/6
>8 orang	1	0	1	1/6
Total	2	4	6	
Prior P(c)	2/6	4/6		

Tabel frekuensi Nilai_UN	Class		Total	Evidence P(x)
	Tidak layak	Layak		
Belum UAN	1	0	1	1/6
7,01 - 8	0	1	1	1/6
>8	1	3	4	4/6
Total	2	4	6	
Prior (c)	2/6	4/6		

Selanjutnya menghitung probabilitas masing-masing kelas berdasarkan rekomendasi layak atau tidak layak menggunakan teorema Bayes

a. Menghitung Probabilitas layak

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

$$\begin{aligned} P(A|B) &= P(\text{layak}|\text{Pendaftar } x) \\ P(B|A) &= \text{Probabilitas}(\text{Pendaftar } x|\text{layak}) \\ P(A) &= \text{Probabilitas}(\text{layak}) \\ P(B) &= \text{Probabilitas}(\text{Pendaftar } x) \end{aligned}$$

$$P(\text{layak}|\text{Pendaftar } x) = \frac{P(\text{Pendaftar } x|\text{layak})P(\text{layak})}{P(\text{Pendaftar } x)}$$

$$\begin{aligned} &P(\text{Layak}|\text{Pendaftar } x) = \\ &\frac{P(\text{pekerjaan} = \text{petani, tanggungan ortu} \Rightarrow 8 \text{ orang, Nilai UN} \Rightarrow 8|\text{Layak})P(\text{Layak})}{P(\text{pekerjaan} = \text{petani, tanggungan ortu} \Rightarrow 8 \text{ orang, rerata UN} \Rightarrow 8)} \end{aligned}$$

Selanjutnya nilai probabilitas diambilkan dari tabel frekuensi setiap atribut class

$$\begin{aligned} &P(\text{Layak}|\text{Pendaftar } x) \\ &= \frac{\left(\frac{\text{pekerjaan petani layak} * \text{tang ortu lebih dari sama dengan 8 layak} * \text{Nilai UN lebih dari sama dengan 8 layak}}{\text{jumlah pekerjaan layak} * \text{jumlah tang ortu layak} * \text{jumlah nilai UN layak}} \right) * \text{Probabilitas layak}}{\frac{\text{pekerjaan petani} * \text{tang ortu lebih dari sama dengan 8} * \text{Nilai UN lebih dari sam dengan 8}}{\text{jumlah pekerjaan} * \text{jumlah tang ortu} * \text{jumlah nilai UN}}} \\ &= \frac{\left(\frac{4}{6} * \frac{0}{6} * \frac{3}{6} \right) * \frac{4}{6}}{1 * 0,17 * 0,67} = \frac{0}{0,1139} = 0 \end{aligned}$$

b. Menghitung probabilitas tidak layak

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

$$\begin{aligned} P(A|B) &= P(\text{tidak layak}|\text{Pendaftar } x) \\ P(B|A) &= \text{Probabilitas}(\text{Pendaftar } x|\text{tidak layak}) \\ P(A) &= \text{Probabilitas}(\text{tidak layak}) \\ P(B) &= \text{Probabilitas}(\text{Pendaftar } x) \end{aligned}$$

$$P(\text{tidak layak}|X) = \frac{P(X|\text{tidak layak})P(\text{tidak layak})}{P(x)}$$

$$\begin{aligned} &P(\text{tidak layak}|X) \\ &= \frac{P(\text{pekerjaan ayah} = \text{petani, tanggungan ortu} \Rightarrow 8 \text{ orang, rerata UN} \Rightarrow 8|\text{tidak layak})P(\text{tidak layak})}{P(\text{pekerjaan ayah} = \text{petani, tanggungan ortu} \Rightarrow 8 \text{ orang, rerata UN} \Rightarrow 8)} \end{aligned}$$

Selanjutnya nilai probabilitas diambilkan dari tabel frekuensi setiap atribut class

$$P(\text{tidak layak} | \text{Pendaftar } x) = \frac{\left(\frac{\text{pekerjaan petani tidak layak}}{\text{jumlah pekerjaan tidak layak}} \cdot \frac{\text{tang ortu lebih dari sama dengan 8 tidak layak}}{\text{jumlah tang ortu tidak layak}} \cdot \frac{\text{Nilai UN lebih dari sama dengan 8 tidak layak}}{\text{jumlah nilai UN tidak layak}} \right) \cdot \text{Probabilitas tidak layak}}{\frac{\text{pekerjaan petani}}{\text{jumlah pekerjaan}} \cdot \frac{\text{tang ortu lebih dari sama dengan 8}}{\text{jumlah tang ortu}} \cdot \frac{\text{Nilai UN lebih dari sama dengan 8}}{\text{jumlah nilai UN}}}$$

$$\frac{\left(\frac{2}{6} * \frac{1}{6} * \frac{1}{6} \right) * \frac{2}{6}}{\frac{1}{6} * \frac{1}{6} * \frac{4}{6}} = \frac{(1 * 0,5 * 0,5) * 0,33}{1 * 0,17 * 0,67} = \frac{0,0825}{0,1139} = \mathbf{0,7243195785776}$$

Karena nilai probabilitas tidak layak lebih besar dibandingkan probabilitas layak,

maka probabilitas rekomendasi untuk pendaftar x adalah **tidak layak**

Kemudian untuk implementasi K-NN pada data pendaftar beasiswa KIP Kuliah

UIN Maulana Malik Ibrahim sebagai berikut:

Diketahui

A1 = Peng_ayah

A5 = Rerata_UN

A9 = Luas_rumah

A2 = Peng_ibu

A6 = Rerata_rapot

A10 = PBB

A3 = Pekerjaan_ayah

A7 = Prestasi_aka

A11 = Rek_listrik

A4 = tangg_ortu

A8 = Prestasi_non_aka

Tabel 3.7 encoding data beasiswa untuk perhitungan dalam K-NN

NO	ID	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	Rekomendasi
1	42265600148	3	3	3	4	4	3	2	1	4	4	4	Ditolak
2	42221700208	3	3	3	1	4	4	2	3	4	4	4	Diterima
3	42253101508	3	3	3	1	3	3	3	4	4	4	4	Diterima
4	42253101498	3	3	3	3	1	3	4	1	4	4	4	Ditolak
5	42253101344	3	3	3	2	4	3	2	1	4	4	4	Diterima
6	42231900280	3	3	3	1	4	3	2	4	3	4	4	Diterima

Untuk melakukan perhitungan algoritma K-NN terlebih perlu dihitung jarak euclidian antara data uji dengan setiap data training. Misalnya data yang diprediksi adalah data pertama, maka berikut ini tahapan perhitungan K-NN:

- Tentukan nilai K, misalnya 3

b. Hitung jarak euclidian

Tabel 3.8 Menghitung jarak euclidian

Data	Euclidian distance
Data 1	$[(3-3)^2 + (3-3)^2 + (3-3)^2 + (4-4)^2 + (4-4)^2 + (3-3)^2 + (2-2)^2 + (1-1)^2 + (4-4)^2 + (4-4)^2 + (4-4)^2]^{1/2} = 0$
Data 2	$[(3-3)^2 + (3-3)^2 + (3-3)^2 + (4-1)^2 + (4-4)^2 + (3-4)^2 + (2-2)^2 + (1-3)^2 + (4-4)^2 + (4-4)^2 + (4-4)^2]^{1/2} = 3,741657$
Data 3	$[(3-3)^2 + (3-3)^2 + (3-3)^2 + (4-1)^2 + (4-3)^2 + (3-3)^2 + (2-3)^2 + (1-4)^2 + (4-4)^2 + (4-4)^2 + (4-4)^2]^{1/2} = 4,472136$
Data 4	$[(3-3)^2 + (3-3)^2 + (3-3)^2 + (4-3)^2 + (4-1)^2 + (3-3)^2 + (2-4)^2 + (1-1)^2 + (4-4)^2 + (4-4)^2 + (4-4)^2]^{1/2} = 3,741657$
Data 5	$[(3-3)^2 + (3-3)^2 + (3-3)^2 + (4-2)^2 + (4-4)^2 + (3-3)^2 + (2-2)^2 + (1-1)^2 + (4-4)^2 + (4-4)^2 + (4-4)^2]^{1/2} = 2$
Data 6	$[(3-3)^2 + (3-3)^2 + (3-3)^2 + (4-1)^2 + (4-4)^2 + (3-3)^2 + (2-2)^2 + (1-4)^2 + (4-3)^2 + (4-4)^2 + (4-4)^2]^{1/2} = 4,358899$

c. Urutkan hasil perhitungan jarak dari yang terkecil

Tabel 3.9 Mengurutkan hasil perhitungan jarak

Data	Distance	Rekomendasi	K=3
Data 1	0	Ditolak	Ditolak
Data 5	2	Diterima	Diterima
Data 4	3,741657	Ditolak	Ditolak
Data 2	3,741657	Diterima	
Data 6	4,358899	Diterima	
Data 3	4,472136	Diterima	

d. Jika k=3 maka hasil voting 3 tetangga terdekat terbanyak adalah **ditolak**

Ditolak [data 1, data 4] > diterima [data 5]

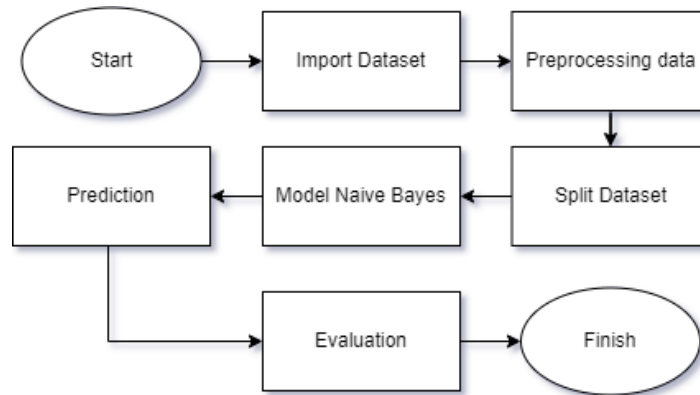
3.5.5 Evaluasi

Proses evaluasi yang dilakukan ialah mengukur akurasi setiap metode menggunakan confusion matrix. Kemudian dari beberapa percobaan dengan skenario yang berbeda-beda akan dilakukan perbandingan antara akurasi metode K-NN dengan akurasi metode Naive Bayes.

BAB IV

METODE NAIVE BAYES

4.1 Desain Metode Naive Bayes



Gambar 4.1 Desain Metode Naive Bayes

Pada bagian ini akan menganalisis metode naive Bayes dalam melakukan klasifikasi dan prediksi. Gambar 4.1 menunjukkan langkah-langkah proses metode naive Bayes. Adapun tahapan-tahapannya dijelaskan sebagai berikut:

4.1.1. *Import Dataset*

Import dataset dilakukan dengan menggunakan library pandas di Python. File dataset akan dipanggil menggunakan `read_csv` di Python kemudian disimpan di dalam sebuah variabel. Dataset yang digunakan harus berkualitas agar model dapat digunakan dengan optimal. Dataset yang digunakan adalah dataset pendaftar KIP Kuliah tahun 2022. Gambar 4.1 adalah potongan kode untuk memanggil file `df_encoded` yang bertipe csv menggunakan library pandas.

```
1 import pandas as pd
2 #import dataset
3 df_encoded = pd.read_csv('df_encoded.csv', sep=',')
4 df_encoded.head()
```

Gambar 4.2 Proses import dataset

4.1.2. Preprocessing data

Setelah memanggil dataset, fitur-fitur (atribut-atribut) yang digunakan untuk memprediksi kelayakan beasiswa KIP dipisahkan dari kolom target rekom. Fitur-fitur ini disimpan dalam variabel X, sementara kolom target rekom disimpan dalam variabel y. Gambar 4.2 menunjukkan potongan perintah untuk memisahkan feature dan target pada dataset KIP menggunakan perintah drop kolom.

```

1 # Memisahkan feature dan target
2 X = df_encoded.drop(['rekom'], axis=1)
3 y = df_encoded['rekom']

```

Gambar 4.3 Proses memisahkan feature dan target

Dataset kemudian dibagi menjadi data latih dan data uji menggunakan fungsi `train_test_split` dari library `sklearn.model_selection`. Dalam kode pada gambar 4.3, data uji atau test size yang digunakan disimpan dalam variabel `test_sizes`, dengan tetap menjaga proporsi kelas yang seimbang menggunakan `random_state=42` untuk memastikan hasil yang konsisten. Program akan melakukan perulangan sebanyak jumlah data uji yang digunakan yaitu 8 kali pengujian.

```

9 # Range test size dari 0.10 hingga 0.30
10 test_sizes = [0.10, 0.15, 0.20, 0.25, 0.30, 0.35, 0.40, 0.45]
11
12 for test_size in test_sizes:
13     # Split dataset menjadi training set dan testing set
14     X_train, X_test, y_train, y_test = train_test_split
15     (X, y, test_size=test_size, random_state=42)

```

Gambar 4.4 Pemisahan data training dan data uji

4.1.3. Model Naive Bayes

Model Naive Bayes dengan menggunakan Gaussian Naive Bayes diinisialisasi menggunakan library `sklearn.naive_Bayes`. Model ini akan

digunakan untuk melatih data latih dan melakukan prediksi pada data uji. Gambar 4.4 menunjukkan perintah untuk melakukan inisialisasi model naive Bayes menggunakan library skicit learn.

```
21 # Inisialisasi model Naive Bayes
22 nb = GaussianNB()
23
24 # Melatih model pada data training
25 nb.fit(X_train, y_train)
```

Gambar 4.5 Implementasi model Naive Bayes

Model Naive Bayes kemudian dilatih menggunakan dataset yang telah dilakukan split sebelumnya. Data yang dilatih adalah data training yang terbagi menjadi X_train sebagai feature dan y_train sebagai target. Proses melatih model Naive Bayes dilakukan menggunakan metode *fit* pada objek model **nb**. Gambar 4.4 adalah perintah untuk melakukan fitting atau melatih model Naive Bayes.

4.1.4. Prediksi

Setelah model dilatih, klasifikasi kelayakan beasiswa KIP dilakukan pada data uji menggunakan metode predict pada objek model nb. Hasil prediksi disimpan dalam variabel y_pred_nb. Gambar 4.5 menunjukkan proses prediksi yang diimplementasikan pada pemrograman python.

```
27 # Melakukan prediksi pada data testing
28 y_pred_nb = nb.predict(X_test)
```

Gambar 4.6 Proses prediksi dengan Naive Bayes

4.1.5. Evaluasi Performa

Untuk mengevaluasi performa model Naive Bayes, beberapa metrik evaluasi seperti akurasi, presisi, dan recall dihitung menggunakan library sklearn.metrics. Dalam kode di atas, metrik-metrik ini dihitung dengan membandingkan nilai aktual dari kolom target y_test dengan hasil prediksi

y_pred_nb. Hasil evaluasi performa model ini disimpan dalam list results untuk setiap ukuran data uji yang berbeda. Gambar 4.6 merupakan proses untuk menampilkan hasil evaluasi performa naive Bayes berdasarkan nilai akurasi, presisi, dan recall.

```

30 # Menghitung evaluasi model
31 acc_nb = accuracy_score(y_test, y_pred_nb)
32 prec_nb = precision_score(y_test, y_pred_nb)
33 rec_nb = recall_score(y_test, y_pred_nb)
34
35 # Menyimpan hasil evaluasi ke dalam list
36 results.append((test_size, jumlah_data, jumlah_data2, acc_nb, prec_nb, rec_nb)

```

Gambar 4.7 Melakukan evaluasi Model Naive Bayes

4.2 Pengujian Naive Bayes

4.2.1. Skenario pengujian

Pengujian menggunakan algoritma klasifikasi Naive Bayes dilakukan dengan variasi pembagian data menjadi data training dan data testing yang berbeda. Jumlah keseluruhan dataset adalah 1044 kemudian akan dipisah sesuai dengan prosentase yang ditentukan untuk mengetahui pengaruh variasi pembagian data terhadap performa model naive Bayes. Tabel 4.1 menunjukkan rancangan skenario pengujian Naive Bayes. Skenario ini bertujuan untuk melihat bagaimana performa model Naive Bayes dalam memprediksi kelayakan calon penerima beasiswa KIP.

4.2.2.1 Pada pengujian pertama, data akan dibagi menjadi data training sebanyak 90% dari total data dan data testing sebanyak 10% dari total data. Model Naive Bayes akan dilatih menggunakan 939 data training dan diuji kinerjanya menggunakan 105 data testing.

- 4.2.2.2 Pengujian kedua, data akan dibagi menjadi data training sebanyak 85% dari total data dan data testing sebanyak 15% dari total data. Model Naive Bayes akan dilatih menggunakan 887 data training dan diuji kinerjanya menggunakan 157 data testing.
- 4.2.2.3 Pada pengujian ketiga, data akan dibagi menjadi data training sebanyak 80% dari total data dan data testing sebanyak 20% dari total data. Model Naive Bayes akan dilatih menggunakan 835 data training dan diuji kinerjanya menggunakan 209 data testing.
- 4.2.2.4 Pada pengujian keempat, data akan dibagi menjadi data training sebanyak 75% dari total data dan data testing sebanyak 25% dari total data. Model Naive Bayes akan dilatih menggunakan 783 data training dan diuji kinerjanya menggunakan 261 data testing.
- 4.2.2.5 Pada pengujian kelima, data akan dibagi menjadi data training sebanyak 70% dari total data dan data testing sebanyak 30% dari total data. Model Naive Bayes akan dilatih menggunakan 730 data training dan diuji kinerjanya menggunakan 314 data testing.
- 4.2.2.6 Pada pengujian keenam, data akan dibagi menjadi data training sebanyak 65% dari total data dan data testing sebanyak 35% dari total data. Model Naive Bayes akan dilatih menggunakan 678 data training dan diuji kinerjanya menggunakan 366 data testing.
- 4.2.2.7 Pada pengujian ketujuh, data akan dibagi menjadi data training sebanyak 60% dari total data dan data testing sebanyak 40% dari total data. Model

Naive Bayes akan dilatih menggunakan 626 data training dan diuji kinerjanya menggunakan 418 data testing.

4.2.2.8 Pada pengujian kedelapan, data akan dibagi menjadi data training sebanyak 55% dari total data dan data testing sebanyak 45% dari total data. Model Naive Bayes akan dilatih menggunakan 574 data training dan diuji kinerjanya menggunakan 470 data testing.

Setiap pengujian akan diukur kinerjanya dengan metrik-metrik evaluasi yang sesuai, seperti akurasi, presisi, atau recall untuk mengevaluasi seberapa baik model dapat melakukan klasifikasi pada data testing yang belum pernah dilihat sebelumnya. Hasil pengujian ini dapat membantu dalam menentukan pembagian data training dan data testing yang optimal untuk penggunaan model Naive Bayes pada dataset KIP.

Tabel 4.1 Skenario pengujian naive Bayes

Pengujian	Train-test split	Training	Test
1	90-10	939	105
2	85-15	887	157
3	80-20	835	209
4	75-25	783	261
5	70-30	730	314
6	65-35	678	366
7	60-40	626	418
8	55-45	574	470

4.2.2. Hasil pengujian

4.2.2.9 Confusion matrix pengujian ke-1

Tabel 4.2 menunjukkan hasil confusion matrix Naive Bayes untuk test size 0.10 dan training size 0.90. Data testing yang digunakan sebanyak 105. Hasilnya menunjukkan terdapat 6 data yang sebenarnya masuk dalam kelas "diterima" dan

model berhasil memprediksi dengan benar 6 data sebagai "diterima" (True Positive). Kemudian terdapat 60 data yang sebenarnya masuk dalam kelas "ditolak" dan model berhasil memprediksi dengan benar sebagai "ditolak" (True Negative). Ada 32 data yang seharusnya memiliki kelas ditolak, tetapi diprediksi sebagai diterima (False Positive). Ada 7 data yang seharusnya berkelas diterima, tetapi diprediksi sebagai ditolak (False negative)

Tabel 4.2 confusion matrix naive Bayes pengujian 1

Naive Bayes test size 0.10		Prediksi	
		Diterima	Ditolak
Aktual	Diterima	6	32
	Ditolak	7	60

4.2.2.10 Confusion matrix pengujian ke-2

Tabel 4.3 menunjukkan hasil confusion matrix naive Bayes untuk test size 0.15 dan training size 0.85. Data testing yang digunakan sebanyak 157. Hasilnya menunjukkan terdapat 8 data yang sebenarnya masuk dalam kelas "diterima" dan model berhasil memprediksi sebagai "diterima" (True Positive). Kemudian terdapat 93 data yang sebenarnya masuk dalam kelas "ditolak" dan model berhasil memprediksi dengan sebagai "ditolak" (True Negative). Ada 46 data yang seharusnya berkelas ditolak, tetapi diprediksi sebagai diterima (False Positive). Ada 10 data yang seharusnya berkelas diterima, tetapi diprediksi sebagai ditolak (False negative)

Tabel 4.3 confusion matrix naive Bayes pengujian 2

Naive Bayes test size 0.15		Prediksi	
		Diterima	Ditolak
Aktual	Diterima	8	46
	Ditolak	10	93

4.2.2.11 Confusion matrix pengujian ke-3

Tabel 4.4 menunjukkan hasil confusion matrix naive Bayes untuk test size 0.20 dan training size 0.80. Data testing yang digunakan sebanyak 209. Hasilnya menunjukkan terdapat 13 data yang sebenarnya masuk dalam kelas "diterima" dan model berhasil memprediksi sebagai "diterima" (True Positive). Kemudian terdapat 124 data yang sebenarnya masuk dalam kelas "ditolak" dan model berhasil memprediksi dengan sebagai "ditolak" (True Negative). Ada 60 data yang seharusnya berkelas ditolak, tetapi diprediksi sebagai diterima (False Positive). Ada 12 data yang seharusnya berkelas diterima, tetapi diprediksi sebagai ditolak (False negative)

Tabel 4.4 confusion matrix naive Bayes pengujian 3

Naive Bayes test size 0.20		Prediksi	
		Diterima	Ditolak
Aktual	Diterima	13	60
	Ditolak	12	124

4.2.2.12 Confusion matrix pengujian ke-4

Tabel 4.5 menunjukkan hasil confusion matrix naive Bayes untuk test size 0.25 dan training size 0.75. Data testing yang digunakan sebanyak 261. Hasilnya menunjukkan terdapat 17 data yang sebenarnya masuk dalam kelas "diterima" dan model berhasil memprediksi sebagai "diterima" (True Positive). Kemudian terdapat 151 data yang sebenarnya masuk dalam kelas "ditolak" dan model berhasil memprediksi dengan sebagai "ditolak" (True Negative). Ada 76 data yang seharusnya berkelas ditolak, tetapi diprediksi sebagai diterima (False Positive). Ada 17 data yang seharusnya berkelas diterima, tetapi diprediksi sebagai ditolak (False negative)

Tabel 4.5 confusion matrix naive Bayes pengujian 4

Naive Bayes test size 0.25		Prediksi	
		Diterima	Ditolak
Aktual	Diterima	17	76
	Ditolak	17	151

4.2.2.13 Confusion matrix pengujian ke-5

Tabel 4.6 menunjukkan hasil confusion matrix naive Bayes untuk test size 0.30 dan training size 0.70. Data testing yang digunakan sebanyak 314. Hasilnya menunjukkan terdapat 18 data yang sebenarnya masuk dalam kelas "diterima" dan model berhasil memprediksi sebagai "diterima" (True Positive). Kemudian terdapat 171 data yang sebenarnya masuk dalam kelas "ditolak" dan model berhasil memprediksi dengan sebagai "ditolak" (True Negative). Ada 97 data yang seharusnya berkelas ditolak, tetapi diprediksi sebagai diterima (False Positive). Ada 28 data yang seharusnya berkelas diterima, tetapi diprediksi sebagai ditolak (False negative)

Tabel 4.6 confusion matrix naive Bayes pengujian 5

Naive Bayes test size 0.30		Prediksi	
		Diterima	Ditolak
Aktual	Diterima	18	97
	Ditolak	28	171

4.2.2.14 Confusion matrix pengujian ke-6

Tabel 4.7 menunjukkan hasil confusion matrix naive Bayes untuk test size 0.35 dan training size 0.65. Data testing yang digunakan sebanyak 366. Hasilnya menunjukkan terdapat 21 data yang sebenarnya masuk dalam kelas "diterima" dan model berhasil memprediksi sebagai "diterima" (True Positive). Kemudian terdapat 205 data yang sebenarnya masuk dalam kelas "ditolak" dan model

berhasil memprediksi dengan sebagai "ditolak" (True Negative). Ada 110 data yang seharusnya berkelas ditolak, tetapi diprediksi sebagai diterima (False Positive). Ada 30 data yang seharusnya berkelas diterima, tetapi diprediksi sebagai ditolak (False negative)

Tabel 4.7 confusion matrix naive Bayes pengujian 6

Naive Bayes test size 0.35		Prediksi	
		Diterima	Ditolak
Aktual	Diterima	21	110
	Ditolak	30	205

4.2.2.15 Confusion matrix pengujian ke-7

Tabel 4.8 menunjukkan hasil confusion matrix naive Bayes untuk test size 0.40 dan training size 0.60. Data testing yang digunakan sebanyak 418. Hasilnya menunjukkan terdapat 25 data yang sebenarnya masuk dalam kelas "diterima" dan model berhasil memprediksi sebagai "diterima" (True Positive). Kemudian terdapat 229 data yang sebenarnya masuk dalam kelas "ditolak" dan model berhasil memprediksi dengan sebagai "ditolak" (True Negative). Ada 126 data yang seharusnya berkelas ditolak, tetapi diprediksi sebagai diterima (False Positive). Ada 38 data yang seharusnya berkelas diterima, tetapi diprediksi sebagai ditolak (False negative)

Tabel 4.8 confusion matrix naive Bayes pengujian 7

Naive Bayes test size 0.40		Prediksi	
		Diterima	Ditolak
Aktual	Diterima	25	126
	Ditolak	38	229

4.2.2.16 Confusion matrix pengujian ke-8

Tabel 4.9 menunjukkan hasil confusion matrix naive Bayes untuk test size 0.45 dan training size 0.55. Data testing yang digunakan sebanyak 470. Hasilnya menunjukkan terdapat 30 data yang sebenarnya masuk dalam kelas "diterima" dan model berhasil memprediksi sebagai "diterima" (True Positive). Kemudian terdapat 260 data yang sebenarnya masuk dalam kelas "ditolak" dan model berhasil memprediksi dengan sebagai "ditolak" (True Negative). Ada 129 data yang seharusnya berkelas ditolak, tetapi diprediksi sebagai diterima (False Positive). Ada 51 data yang seharusnya berkelas diterima, tetapi diprediksi sebagai ditolak (False negative)

Tabel 4.9 confusion matrix naive Bayes pengujian 8

Naive Bayes test size 0.45		Prediksi	
		Diterima	Ditolak
Aktual	Diterima	30	129
	Ditolak	51	260

Berdasarkan confusion matrix pada uji coba pertama sampai kedelapan, maka diperoleh nilai akurasi, presisi, dan recall yang berbeda. Hal ini terlihat dalam tabel 4.10 yang menampilkan hasil uji coba naive Bayes. Perbedaan ini dipengaruhi salah satunya oleh variasi test size yang digunakan. Berikut analisis performa Naive Bayes berdasarkan skor evaluasi yang dihasilkan:

Tabel 4.10 hasil uji coba metode naive Bayes

Uji coba	Test size	Akurasi	Presisi	Recall
1	0.10	63	65	65
2	0.15	64	67	90
3	0.20	66	67	91
4	0.25	64	67	90
5	0.30	60	64	86
6	0.35	62	65	87

Lanjutan tabel 4.10				
7	0.40	61	65	86
8	0.45	62	67	84
Rata-rata		62,75	65,88	84,88

i. Akurasi

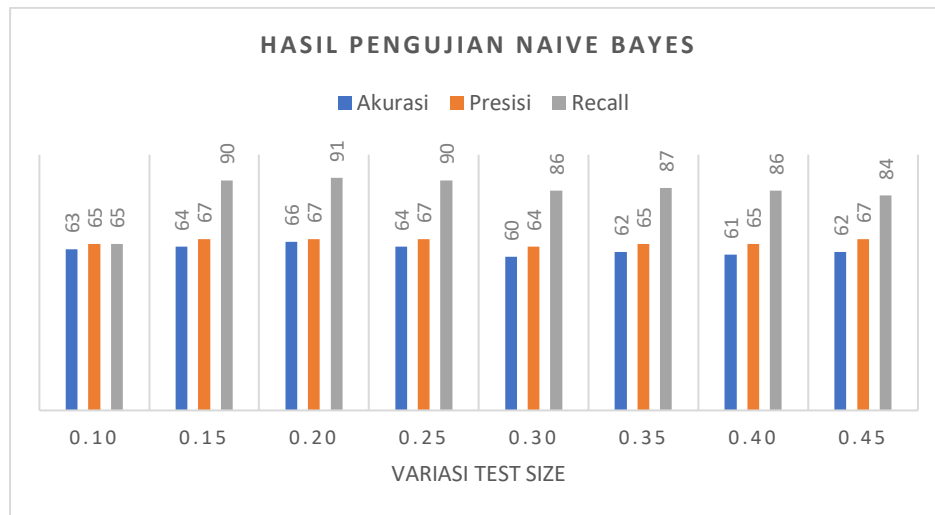
Rentang akurasi yang diperoleh untuk test size antara 0.10 hingga 0.45 adalah antara 60% hingga 66%. Akurasi tertinggi berada pada uji coba ketiga dengan test size 0.20 yang menghasilkan akurasi 66%. Pada uji coba selanjutnya hasil akurasi stabil bahkan cenderung turun. Rata-rata akurasi adalah 62,75%. Ini adalah tingkat akurasi yang dapat diterima, namun masih ada ruang untuk perbaikan.

ii. Presisi

Presisi menggambarkan seberapa baik model dalam mengidentifikasi calon penerima beasiswa KIP dengan benar dari seluruh data yang diprediksi positif. Rentang presisi yang diperoleh untuk test size antara 0.10 hingga 0.45 adalah antara 64% hingga 67%. Rata-rata presisi yang diperoleh adalah 65,88%. Ini adalah tingkat presisi yang baik walaupun masih dapat ditingkatkan.

iii. Recall

Rentang recall yang diperoleh untuk test size antara 0.10 hingga 0.45 adalah antara 84% hingga 91%. Recall menggambarkan seberapa baik model dalam mengidentifikasi calon penerima beasiswa KIP dengan benar dari seluruh calon penerima beasiswa KIP yang sebenarnya. Semakin tinggi nilai recall, semakin sedikit calon yang memenuhi syarat yang salah terklasifikasi sebagai tidak memenuhi syarat.



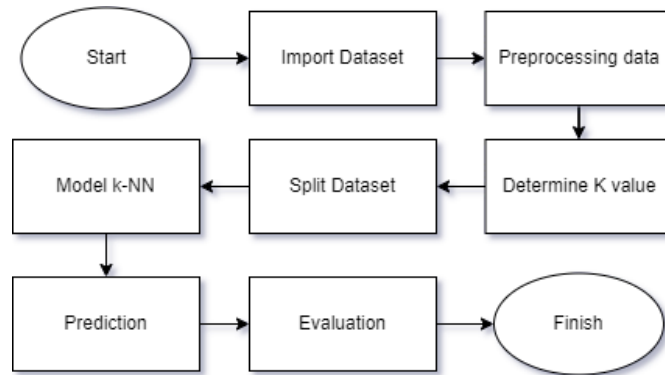
Gambar 4.8 Hasil pengujian naive Bayes

Berdasarkan gambar 4.7 hasil pengujian naive Bayes, dapat dilihat bahwa model Naive Bayes memberikan performa yang cukup baik dalam memprediksi kelayakan calon penerima beasiswa KIP. Namun, terdapat variasi dalam performa model tergantung pada ukuran data uji yang digunakan. Hal ini menunjukkan bahwa ukuran data uji dapat mempengaruhi performa model, dan ukuran data uji yang lebih besar dapat memberikan hasil yang lebih konsisten.

BAB V

METODE K-NEAREST NEIGHBOR (k-NN)

5.1 Desain Metode k-NN



Gambar 5.1 Desain Metode k-Nearest Neighbor

Pada bagian ini metode k-NN akan diuraikan dengan menampilkan gambar desain metode dan tahapan-tahapan yang dilakukan untuk pemodelan. Gambar 5.1 adalah desain metode k-NN yang digunakan. Adapun tahapan-tahapan dalam melakukan analisis k-NN sebagai berikut:

5.1.1. Import dataset

Pada tahap ini, dataset diimpor ke dalam jupyter notebook menggunakan library pandas. Dataset ini adalah dataset penerima beasiswa KIP dan dataset disimpan dalam dataframe dengan nama `df_kip`. Gambar 5.2 menampilkan proses import dataset untuk diprediksi dengan model k-NN. Dataset yang digunakan adalah data berformat csv.

```
1 import pandas as pd
2 #import dataset
3 df_kip = pd.read_csv('dataset kip 2022.csv', sep=';')
4 df_kip.head()
```

Gambar 5.2 Import dataset untuk proses k-NN

5.1.2. Preprocessing data

Pada tahap ini akan dilakukan preprocessing data untuk memastikan kualitas data yang baik sebelum melanjutkan dengan pemodelan. Gambar 5.3 merupakan normalisasi data yaitu mengganti data yang tidak sesuai. Seperti kesalahan pada atribut `rerata_rapor` dimana ada data berformat tanggal. Mengganti data yang tidak sesuai pada kolom `'rerata_rapor'` dan `'rerata_un'` dilakukan menggunakan metode `'replace()'`.

```

1 # ganti data pada kolom rerata_rapor dan rerata_un yang tidak sesuai
2
3 ganti = {
4     '08-Jul': '7-8',
5     '07-Aug': '7-8'
6 }
7
8 df_kip['rerata_rapor'] = df_kip['rerata_rapor'].replace(ganti)
9 df_kip['rerata_un'] = df_kip['rerata_un'].replace('07-Jun', '6 - 7')
10 df_kip.head()

```

Gambar 5.3 proses mengganti data yang tidak sesuai

Proses selanjutnya ialah melakukan seleksi kolom atau atribut. Atribut yang tidak relevan akan dibuang yaitu `no`, `id`, `domisili`, `jalur`. Gambar 5.4 menunjukkan proses pembuangan kolom dengan menggunakan metode `drop()` sebelum dilakukan encoding data.

```

1 # menghapus kolom yang tidak dilakukan encode
2 X = df_kip.drop(['no', 'id', 'domisili', 'jalur'], axis=1)
3 X.head()

```

Gambar 5.4 seleksi kolom sebelum proses encode

Encoding data dilakukan menggunakan library `skicit-learn` dan metode yang digunakan adalah `OrdinalEncoder`. Gambar 5.5 menunjukkan proses encoding data pada `DataFrame X` kemudian disimpan dalam variabel `encoded_data`

kemudian hasil encoding akan ditampilkan dalam format DataFrame dan disimpan pada variabel `encoded_df`.

```

1 from sklearn.preprocessing import OrdinalEncoder
2
3 # Buat objek OrdinalEncoder dengan parameter categories yang sesuai
4 ordinal_encoder = OrdinalEncoder()
5
6 # Lakukan ordinal encoding pada DataFrame X
7 encoded_data = ordinal_encoder.fit_transform(X)
8
9 # Ubah hasil encoding menjadi DataFrame
10 encoded_df = pd.DataFrame(encoded_data, columns=X.columns)
11
12 # Tampilkan hasil encoding
13 encoded_df.head()

```

Gambar 5.5 Proses encoding data

5.1.3. Split data

Setelah preprocessing data, dataset dibagi menjadi training set dan testing set menggunakan `train_test_split()` dari modul `sklearn.model_selection`. Gambar 5.6 menunjukkan proses pemisahan data menjadi training set dan test set. Nilai X dan y diperoleh dari proses pemisahan menggunakan `drop()`.

```

16 # Split dataset menjadi training set dan testing set
17 X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=test_size, random_state=42)

```

Gambar 5.6 proses split data untuk k-NN

Gambar 5.7 adalah proses memisahkan feature yang disimpan dalam variabel X dan target yang disimpan dalam variabel y. Variabel X berisi atribut selain target dan variabel y berisi 1 atribut yang merupakan target. Hasil dari pemisahan dataset akan dijadikan training set dan testing set.

```

1 # Memisahkan feature dan target
2 X = encoded_df.drop(['rekom'], axis=1)
3 y = encoded_df['rekom']

```

Gambar 5.7 Proses pemisahan feature dan target

5.1.4. Tentukan nilai k

Pada tahap ini, dilakukan iterasi untuk setiap nilai k dalam rentang 3 hingga 15 dengan mengambil bilangan ganjil. Nilai k ini menentukan jumlah tetangga yang akan digunakan untuk memprediksi label suatu data baru.

5.1.5. Model k-NN

Setelah nilai k ditentukan, model k-NN (KNeighborsClassifier) diinisialisasi dengan menggunakan nilai k yang sesuai dan metrik jarak 'euclidean'. Model ini akan digunakan untuk melatih dan memprediksi data.

```
19 # Inisialisasi model kNN
20 knn = KNeighborsClassifier(n_neighbors=k, metric='euclidean')
21
22 # Melatih model menggunakan training set
23 knn.fit(X_train, y_train)
```

Gambar 5.8 Proses pemisahan feature dan target

5.1.6. Prediksi

Setelah model dievaluasi, langkah terakhir adalah menggunakan model yang telah dilatih untuk melakukan prediksi pada data baru. Pada kode tersebut, dilakukan prediksi pada testing set menggunakan metode `predict()` pada model k-NN.

```
25 # Melakukan prediksi pada testing set
26 y_pred_knn = knn.predict(X_test)
```

Gambar 5.9 Proses Prediksi pada k-NN

5.1.7. Evaluasi model

Setelah model k-NN dilatih dengan menggunakan training set, dilakukan evaluasi performa model menggunakan testing set. Gambar 5.10 menampilkan kode untuk menghitung nilai confusion matrix yang diperoleh oleh model k-NN. Hasilnya dijadikan untuk menghitung nilai akurasi, presisi, dan recall.

```

36 # Menampilkan confusion matrix
37 cm_knn = confusion_matrix(y_test, y_pred_knn)

```

Gambar 5.10 Menampilkan confusion matrix k-NN

Pada gambar 5.11, metrik evaluasi yang digunakan adalah akurasi, presisi, dan recall. Akurasi mengukur seberapa tepat model dalam memprediksi keseluruhan data, sedangkan presisi dan recall mengukur kualitas prediksi dalam kelas tertentu.

```

28 # Menghitung evaluasi model
29 acc_knn = accuracy_score(y_test, y_pred_knn)
30 prec_knn = precision_score(y_test, y_pred_knn)
31 rec_knn = recall_score(y_test, y_pred_knn)

```

Gambar 5.11 Menampilkan matrik evaluasi k-NN

5.2 Skenario pengujian

Skenario pengujian K-Nearest Neighbors (K-NN) dalam penelitian ini dilakukan dengan memvariasikan dua parameter utama, yaitu nilai test size (jumlah data uji) dan nilai k (jumlah tetangga terdekat yang dipertimbangkan dalam klasifikasi). Tujuan dari pengujian ini adalah untuk menentukan nilai k terbaik yang memberikan performa klasifikasi yang optimal.

5.2.1 Variasi Test Size

Test size mengacu pada ukuran dataset uji yang digunakan untuk menguji model K-NN. Pada skenario pengujian yang digunakan, nilai test size yang digunakan adalah yaitu 10%, 15%, 20%, 25%, 30%, 35%, 40%, dan 45%. Setiap test size merepresentasikan jumlah data yang digunakan sebagai data uji dari total dataset yang tersedia. Total pengujian yang dilakukan berdasarkan variasi test size berjumlah 8 kali pengujian. Setiap pengujian hasilnya akan dievaluasi menggunakan matrik evaluasi.

Tabel 5.1 Skenario pengujian k-NN

Pengujian variasi test size	Train-test split	Training	Test
1	90-10	939	105
2	85-15	887	157
3	80-20	835	209
4	75-25	783	261
5	70-30	730	314
6	65-35	678	366
7	60-40	626	418
8	55-45	574	470

5.2.2 Variasi K

Parameter k mengacu pada jumlah tetangga terdekat yang akan diambil dalam proses klasifikasi. Nilai parameter k yang akan digunakan adalah nilai bilangan bulat ganjil mulai dari 3, 5, 7, 9, 11, 13, dan 15. Pengujian dilakukan dengan melakukan iterasi pengujian variasi test size kemudian dilanjutkan dengan melakukan variasi nilai k. Pengujian ini diharapkan dapat mencari nilai k terbaik berdasarkan nilai matrik akurasi, presisi, dan recall.

Tabel 5.2 skenario pengujian k-NN dengan variasi k

No	Pengujian variasi k	k
1	Iterasi pengujian variasi test size	3
2		5
3		7
4		9
5		11
6		13
7		15

Pada tabel 5.2 skenario pengujian k-NN dengan variasi k, pengujian dilakukan sebanyak tujuh kali dengan rincian sebagai berikut:

5.2.2.1 Pengujian pertama dilakukan dengan melakukan variasi test size mulai dari 10% sampai 45% dengan jumlah dataset 1044. Pengujian dilakukan

dengan mengambil nilai $k = 3$ dan kemudian akan diambil nilai matrik akurasi, presisi, dan recall.

5.2.2.2 Pengujian kedua dilakukan dengan melakukan variasi test size mulai dari 10% sampai 45% dengan jumlah dataset 1044. Pengujian dilakukan dengan mengambil nilai $k = 5$ kemudian akan dihitung nilai matrik akurasi, presisi, dan recall.

5.2.2.3 Pengujian ketiga dilakukan dengan melakukan variasi test size mulai dari 10% sampai 45% dengan jumlah dataset 1044. Pengujian dilakukan dengan mengambil nilai $k = 7$ kemudian akan dihitung nilai matrik akurasi, presisi, dan recall.

5.2.2.4 Pengujian keempat dilakukan dengan melakukan variasi test size mulai dari 10% sampai 45% dengan jumlah dataset 1044. Pengujian dilakukan dengan mengambil nilai $k = 9$ kemudian akan dihitung nilai matrik akurasi, presisi, dan recall.

5.2.2.5 Pengujian kelima dilakukan dengan melakukan variasi test size mulai dari 10% sampai 45% dengan jumlah dataset 1044. Pengujian dilakukan dengan mengambil nilai $k = 11$ kemudian akan dihitung nilai matrik akurasi, presisi, dan recall.

5.2.2.6 Pengujian keenam dilakukan dengan melakukan variasi test size mulai dari 10% sampai 45% dengan jumlah dataset 1044. Pengujian dilakukan dengan mengambil nilai $k = 13$ kemudian akan dihitung nilai matrik akurasi, presisi, dan recall.

5.2.2.7 Pengujian ketujuh dilakukan dengan melakukan variasi test size mulai dari 10% sampai 45% dengan jumlah dataset 1044. Pengujian dilakukan dengan mengambil nilai $k = 15$ kemudian akan dihitung nilai matrik akurasi, presisi, dan recall.

Setiap pengujian akan diukur kinerjanya dengan metrik-metrik evaluasi seperti akurasi, presisi, atau recall untuk mengevaluasi seberapa baik model dapat melakukan klasifikasi pada data testing yang belum pernah dilihat sebelumnya. Hasil pengujian ini dapat membantu dalam menentukan nilai k yang optimal untuk penggunaan model k -NN pada dataset KIP.

5.3 Hasil pengujian k-NN

5.3.1. Matrik Evaluasi

Pada tabel hasil pengujian, Ukuran data uji atau test size adalah proporsi data yang digunakan untuk menguji model. Ketika test size meningkat, akurasi cenderung mengalami penurunan. Hal ini karena model memiliki lebih sedikit data untuk dilatih ketika ukuran pengujian besar.

Kemudian jika memperhatikan variasi k , dengan meningkatnya nilai k , akurasi model umumnya meningkat. Ini karena nilai k yang lebih besar akan membuat model lebih tahan terhadap noise. Namun, akurasi mulai stabil setelah $k = 9$. Hal ini menunjukkan bahwa tidak ada pengaruh yang besar ketika nilai k diatas 9.

Akurasi model atau persentase data yang diklasifikasikan dengan benar cenderung meningkat dengan meningkatnya nilai k hingga titik tertentu. Akurasi tertinggi diperoleh ketika nilai $k=11$ dengan nilai akurasi 67%. Berdasarkan rata-

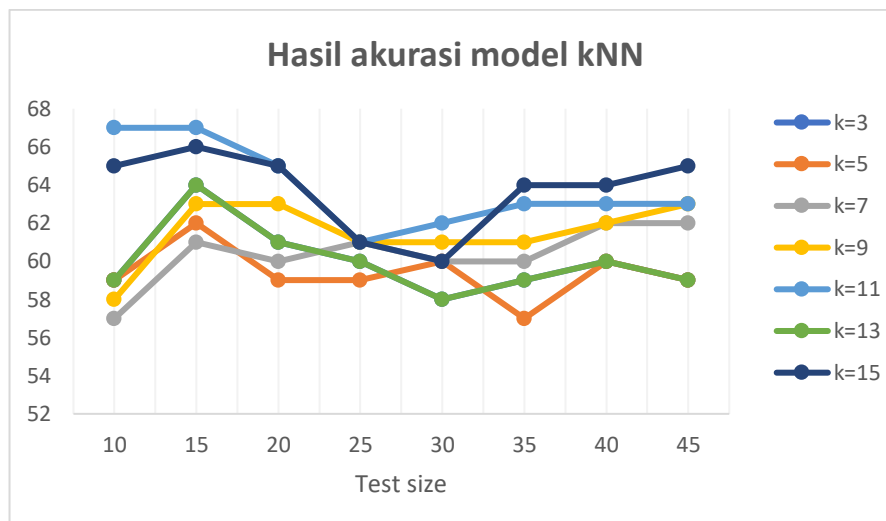
rata akurasi, nilai akurasi semakin meningkat mulai $k=7$ kemudian cenderung stabil setelah $k=11$.

Tabel 5.3 Hasil pengujian akurasi k-NN dengan variasi k

Test size	k=3	k=5	k=7	k=9	k=11	k=13	k=15
10	59	59	57	58	67	59	65
15	64	62	61	63	67	64	66
20	61	59	60	63	65	61	65
25	60	59	61	61	61	60	61
30	58	60	60	61	62	58	60
35	59	57	60	61	63	59	64
40	60	60	62	62	63	60	64
45	59	59	62	63	63	59	65
Rata-rata	60,00	59,38	60,38	61,50	63,88	60,00	63,75

Tabel 5.3 merupakan hasil pengujian k-NN dengan variasi nilai k. pengujian dilakukan dengan test size 10 sampai 45. Setiap test size akan diuji untuk mencari nilai akurasi yang diperoleh antara nilai $k=3$ sampai $k=15$. Hasil pengujian setiap nilai k akan dihitung rata-rata. Nilai $k=11$ memperoleh nilai rata-rata paling tinggi yaitu 63,88 dan nilai rata-rata akurasi terendah diperoleh nilai $k=5$ dengan nilai 59,38. Nilai akurasi tertinggi yang diperoleh berada pada posisi $k=11$ dengan test size 10 dan 15 yaitu 67%.

Gambar 5.12 menunjukkan hasil percobaan untuk mencari nilai k terbaik untuk algoritma K-Nearest Neighbors (K-NN) berdasarkan nilai akurasi yang diperoleh. Eksperimen memvariasikan ukuran uji dan nilai k, hasil akurasi akan dicatat untuk setiap kombinasi. Berdasarkan gambar tersebut menunjukkan bahwa nilai k antara 7 dan 11 adalah nilai yang optimal untuk dataset KIP karena memiliki nilai rata-rata yang cenderung naik.



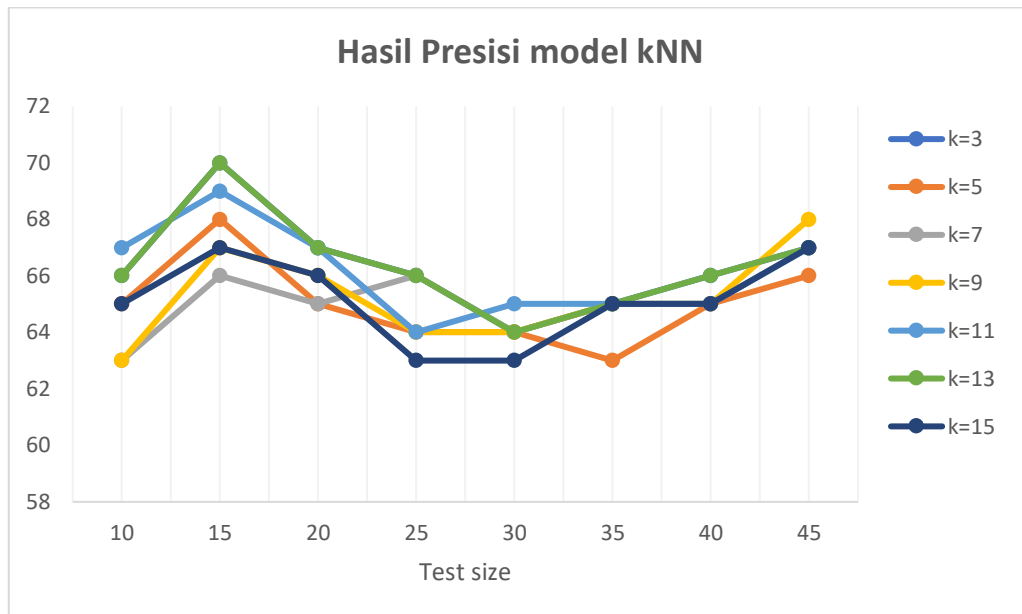
Gambar 5.12 Grafik hasil pengujian k-NN dengan variasi k

Tabel 5.4 menunjukkan hasil pengujian untuk menentukan nilai optimal k untuk model K-nearest neighbor (K-NN). Pengujian dilakukan dengan memvariasikan test size dan nilai k, kemudian mengukur presisi model. Nilai presisi rata-rata meningkat sampai titik tertentu, mulai k=3 sampai k=13, kemudian mulai menurun. Ini menunjukkan bahwa ada nilai k optimal untuk model ini, dan nilainya berada di antara 7 dan 9.

Berdasarkan gambar 5.13, nilai k terbaik untuk model ini adalah 7 atau 9. Hal ini karena memiliki rata-rata presisi paling tinggi. Presisi cenderung menurun saat k diatas 9.

Tabel 5.4 Hasil pengujian nilai presisi k-NN

Test size	k=3	k=5	k=7	k=9	k=11	k=13	k=15
10	66	65	63	63	67	66	65
15	70	68	66	67	69	70	67
20	67	65	65	66	67	67	66
25	66	64	66	64	64	66	63
30	64	64	64	64	65	64	63
35	65	63	65	65	65	65	65
40	66	65	65	65	65	66	65
45	67	66	67	68	67	67	67
Rata-rata	66,38	65,00	65,13	65,25	66,13	66,38	65,13



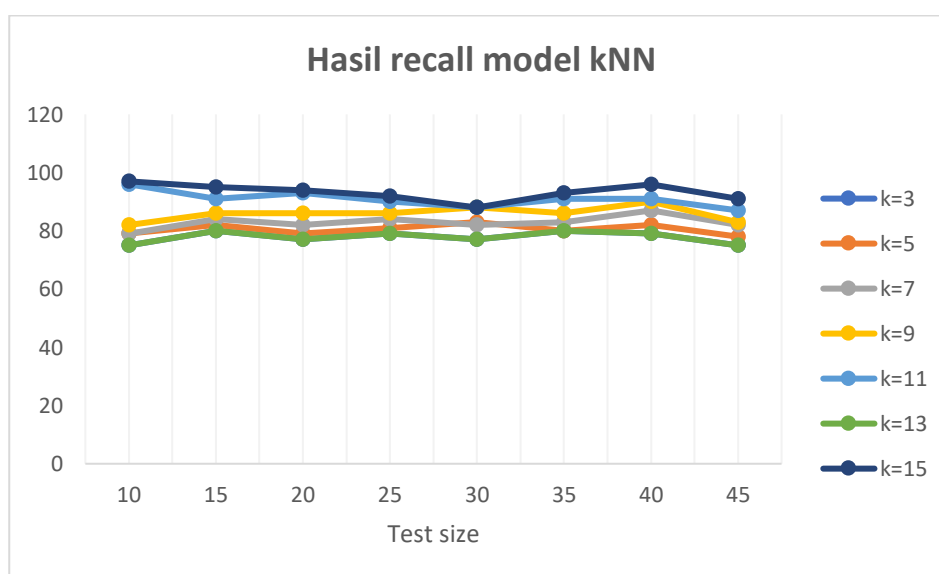
Gambar 5.13 Hasil pengujian nilai presisi k-NN dengan variasi k

Tabel 5.5 menunjukkan hasil pengujian untuk menentukan nilai optimal k pada model K-nearest neighbor (K-NN). Pengujian dilakukan dengan memvariasikan test size dan nilai k untuk mengukur recall model. Nilai recall rata-rata dari setiap k cenderung meningkat sampai titik tertentu, kemudian menurun. Rata-rata Nilai recall terus meningkat mulai k=5 sampai k=11 kemudian mengalami penurunan pada k=13 sebelum naik kembali pada posisi k=15. Ini menunjukkan bahwa ada nilai k optimal untuk model ini, yang salah satunya dipengaruhi oleh semakin besarnya nilai k. Rata-rata nilai recall tertinggi berada pada nilai k=15.

Berdasarkan gambar 5.14, nilai k terbaik untuk model ini adalah k=11 atau k=15. Hal ini karena nilai k tersebut memiliki rata-rata recall paling tinggi. Recall menurun saat k=13.

Tabel 5.5 Hasil Pengujian nilai recall k-NN

Test size	k=3	k=5	k=7	k=9	k=11	k=13	k=15
10	75	79	79	82	96	75	97
15	80	82	84	86	91	80	95
20	77	79	82	86	93	77	94
25	79	81	84	86	90	79	92
30	77	83	82	88	88	77	88
35	80	80	83	86	91	80	93
40	79	82	87	90	91	79	96
45	75	78	82	83	87	75	91
Rata-rata	77,75	80,5	82,88	85,88	90,88	77,75	93,25



Gambar 5.14 Hasil pengujian nilai recall k-NN dengan variasi k

Berdasarkan tabel 5.6, perubahan rata-rata metrik akurasi, presisi, dan recall seiring dengan perubahan nilai k. Pengamatan yang dapat diambil dari tabel 5.6 yaitu bahwa rata-rata Akurasi meningkat ketika nilai k meningkat dari 3 hingga 11, namun, kemudian menurun pada nilai k 13 dan naik kembali pada nilai k 15. Nilai k 11 memiliki rata-rata akurasi tertinggi. Rata-rata Presisi mengalami fluktuasi dengan nilai k yang berbeda. Namun, secara keseluruhan, presisi cenderung tetap relatif stabil dengan perubahan nilai k, dengan nilai tertinggi terjadi pada k = 11. Rata-rata Recall meningkat secara signifikan ketika nilai k

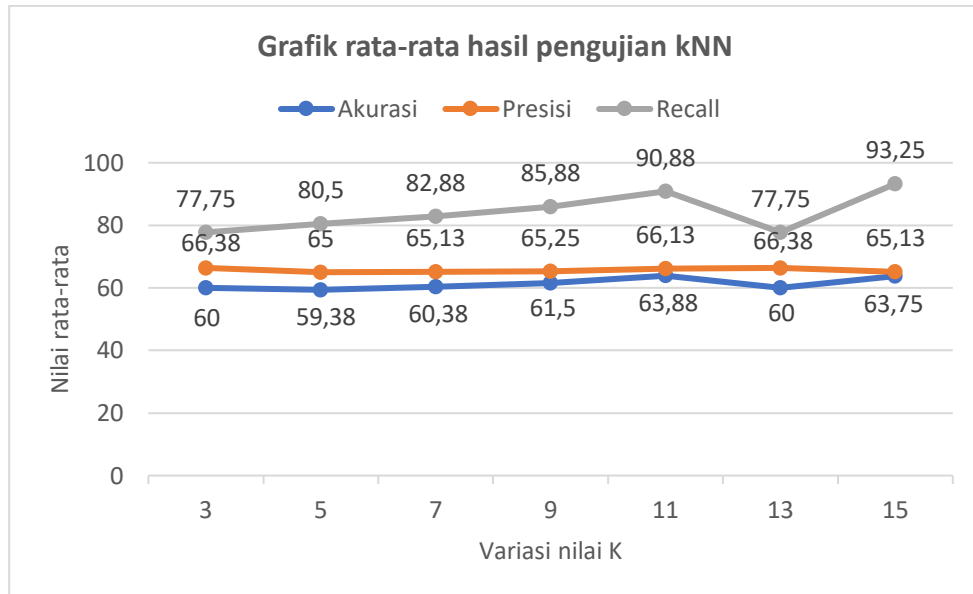
meningkat dari 3 hingga 11, dan kemudian tetap tinggi pada $k = 13$ dan $k = 15$. Nilai $k = 15$ memiliki rata-rata recall tertinggi.

Berdasarkan hasil pengujian, nilai k terbaik untuk model K-NN ini bisa dipilih berdasarkan metrik yang ingin dioptimalkan. Jika fokus pada akurasi, maka $k=11$ mungkin menjadi pilihan yang baik. Namun, jika prioritasnya adalah recall, maka $k=15$ bisa menjadi pilihan terbaik. Untuk presisi, $k=11$ juga tampak menjadi pilihan yang layak karena memiliki nilai presisi yang baik dan stabil.

Jika memperhitungkan keseimbangan antara ketiga metrik (akurasi, presisi, dan recall), maka nilai k terbaik adalah ketika nilai $k=11$ karena memiliki akurasi tertinggi, tetapi juga memiliki nilai presisi dan recall yang tidak terlalu jelek. Oleh karena itu model k-NN yang dipilih dan akan dibandingkan dengan model naive Bayes yang digunakan dalam penelitian ini adalah k-NN dengan $k=11$.

Tabel 5.6 Rata-rata hasil pengujian k-NN dengan variasi k

Nilai k	Rata-rata hasil pengujian		
	Akurasi	Presisi	Recall
3	60	66,38	77,75
5	59,38	65	80,5
7	60,38	65,13	82,88
9	61,50	65,25	85,88
11	63,88	66,13	90,88
13	60	66,38	77,75
15	63,75	65,13	93,25



Gambar 5.15 Rata-rata hasil pengujian k-NN

5.3.2. Confusion Matrix

5.3.2.1. Confusion matrix pengujian ke-1

Tabel 5.7 menunjukkan hasil confusion matrix k-NN untuk test size 0.10 dan training size 0.90. Data testing yang digunakan sebanyak 105. Hasilnya menunjukkan terdapat 6 data yang sebenarnya masuk dalam kelas "diterima" dan model berhasil memprediksi dengan benar 6 data sebagai "diterima" (True Positive). Kemudian terdapat 64 data yang sebenarnya masuk dalam kelas "ditolak" dan model berhasil memprediksi dengan benar 64 data sebagai "ditolak" (True Negative). Ada 32 data yang seharusnya berkelas ditolak, tetapi diprediksi sebagai diterima (False Positive). Ada 3 data yang seharusnya berkelas diterima, tetapi diprediksi sebagai ditolak (False negative)

Tabel 5.7 Confusion matrix k-NN pengujian ke-1

k-NN test size 0.10		Prediksi	
		Diterima	Ditolak
Aktual	Diterima	6	32
	Ditolak	3	64

5.3.2.2. Confusion matrix pengujian ke-2

Tabel 5.8 menunjukkan hasil confusion matrix k-NN untuk test size 0.15 dan training size 0.85. Data testing yang digunakan sebanyak 157. Hasilnya menunjukkan terdapat 11 data yang sebenarnya masuk dalam kelas "diterima" dan model berhasil memprediksi sebagai "diterima" (True Positive). Kemudian terdapat 94 data yang sebenarnya masuk dalam kelas "ditolak" dan model berhasil memprediksi dengan sebagai "ditolak" (True Negative). Ada 43 data yang seharusnya berkelas ditolak, tetapi diprediksi sebagai diterima (False Positive). Ada 9 data yang seharusnya berkelas diterima, tetapi diprediksi sebagai ditolak (False negative)

Tabel 5.8 Confusion matrix k-NN pengujian ke-2

k-NN test size 0.15		Prediksi	
		Diterima	Ditolak
Aktual	Diterima	11	43
	Ditolak	9	94

5.3.2.3. Confusion matrix pengujian ke-3

Tabel 5.9 menunjukkan hasil confusion matrix k-NN untuk test size 0.20 dan training size 0.80. Data testing yang digunakan sebanyak 209. Hasilnya menunjukkan terdapat 10 data yang sebenarnya masuk dalam kelas "diterima" dan model berhasil memprediksi sebagai "diterima" (True Positive). Kemudian terdapat 126 data yang sebenarnya masuk dalam kelas "ditolak" dan model berhasil memprediksi dengan sebagai "ditolak" (True Negative). Ada 63 data yang seharusnya berkelas ditolak, tetapi diprediksi sebagai diterima (False Positive). Ada 10 data yang seharusnya berkelas diterima, tetapi diprediksi sebagai ditolak (False negative)

Tabel 5.9 Confusion matrix k-NN pengujian ke-3

k-NN test size 0.20		Prediksi	
		Diterima	Ditolak
Aktual	Diterima	10	63
	Ditolak	10	126

5.3.2.4. Confusion matrix pengujian ke-4

Tabel 5.10 menunjukkan hasil confusion matrix k-NN untuk test size 0.25 dan training size 0.75. Data testing yang digunakan sebanyak 261. Hasilnya menunjukkan terdapat 8 data yang sebenarnya masuk dalam kelas "diterima" dan model berhasil memprediksi sebagai "diterima" (True Positive). Kemudian terdapat 152 data yang sebenarnya masuk dalam kelas "ditolak" dan model berhasil memprediksi dengan sebagai "ditolak" (True Negative). Ada 85 data yang seharusnya berkelas ditolak, tetapi diprediksi sebagai diterima (False Positive). Ada 16 data yang seharusnya berkelas diterima, tetapi diprediksi sebagai ditolak (False negative)

Tabel 5.10 Confusion matrix k-NN pengujian ke-4

k-NN test size 0.25		Prediksi	
		Diterima	Ditolak
Aktual	Diterima	8	85
	Ditolak	16	152

5.3.2.5. Confusion matrix pengujian ke-5

Tabel 5.11 menunjukkan hasil confusion matrix k-NN untuk test size 0.30 dan training size 0.70. Data testing yang digunakan sebanyak 314. Hasilnya menunjukkan terdapat 19 data yang sebenarnya masuk dalam kelas "diterima" dan model berhasil memprediksi sebagai "diterima" (True Positive). Kemudian terdapat 175 data yang sebenarnya masuk dalam kelas "ditolak" dan model

berhasil memprediksi dengan sebagai "ditolak" (True Negative). Ada 96 data yang seharusnya berkelas ditolak, tetapi diprediksi sebagai diterima (False Positive). Ada 24 data yang seharusnya berkelas diterima, tetapi diprediksi sebagai ditolak (False negative)

Tabel 5.11 Confusion matrix k-NN pengujian ke-5

k-NN test size 0.30		Prediksi	
		Diterima	Ditolak
Aktual	Diterima	19	96
	Ditolak	24	175

5.3.2.6. Confusion matrix pengujian ke-6

Tabel 5.12 menunjukkan hasil confusion matrix k-NN untuk test size 0.35 dan training size 0.65. Data testing yang digunakan sebanyak 366. Jika keseluruhan data berjumlah 1044 maka data trainingnya adalah jumlah data dikurangi data testing yaitu 678 atau 65% dari keseluruhan dataset. Hasilnya menunjukkan terdapat 17 data yang sebenarnya masuk dalam kelas "diterima" dan model berhasil memprediksi sebagai "diterima" (True Positive). Kemudian terdapat 213 data yang sebenarnya masuk dalam kelas "ditolak" dan model berhasil memprediksi dengan sebagai "ditolak" (True Negative). Ada 114 data yang seharusnya berkelas ditolak, tetapi diprediksi sebagai diterima (False Positive). Ada 22 data yang seharusnya berkelas diterima, tetapi diprediksi sebagai ditolak (False negative).

Tabel 5.12 Confusion matrix k-NN pengujian ke-6

k-NN test size 0.35		Prediksi	
		Diterima	Ditolak
Aktual	Diterima	17	114
	Ditolak	22	213

5.3.2.7. Confusion matrix pengujian ke-7

Tabel 5.13 menunjukkan hasil confusion matrix k-NN untuk test size 0.40 dan training size 0.60. Data testing yang digunakan sebanyak 418. Hasilnya menunjukkan terdapat 18 data yang sebenarnya masuk dalam kelas "diterima" dan model berhasil memprediksi sebagai "diterima" (True Positive). Kemudian terdapat 244 data yang sebenarnya masuk dalam kelas "ditolak" dan model berhasil memprediksi dengan sebagai "ditolak" (True Negative). Ada 133 data yang seharusnya berkelas ditolak, tetapi diprediksi sebagai diterima (False Positive). Ada 23 data yang seharusnya berkelas diterima, tetapi diprediksi sebagai ditolak (False negative)

Tabel 5.13 Confusion matrix k-NN pengujian ke-7

k-NN test size 0.40		Prediksi	
		Diterima	Ditolak
Aktual	Diterima	18	133
	Ditolak	23	244

5.3.2.8. Confusion matrix pengujian ke-8

Tabel 5.14 menunjukkan hasil confusion matrix k-NN untuk test size 0.45 dan training size 0.55. Data testing yang digunakan sebanyak 470. Hasilnya menunjukkan terdapat 26 data yang sebenarnya masuk dalam kelas "diterima" dan model berhasil memprediksi sebagai "diterima" (True Positive). Kemudian terdapat 270 data yang sebenarnya masuk dalam kelas "ditolak" dan model berhasil memprediksi dengan sebagai "ditolak" (True Negative). Ada 133 data yang seharusnya berkelas ditolak, tetapi diprediksi sebagai diterima (False

Positive). Ada 41 data yang seharusnya berkelas diterima, tetapi diprediksi sebagai ditolak (False negative)

Tabel 5.14 Confusion matrix k-NN pengujian ke-8

k-NN test size 0.45		Prediksi	
		Diterima	Ditolak
Aktual	Diterima	26	133
	Ditolak	41	270

Skenario pengujian yang dilakukan menggunakan model k-NN pada dataset penerima KIP Kuliah menghasilkan sebuah model terbaik yang akan dilakukan komparasi dengan model klasifikasi yang lain, dalam penelitian ini adalah Naive Bayes. Pemilihan Naive Bayes sebagai pembanding dilakukan karena Naive Bayes merupakan model klasifikasi yang simpel dan mudah untuk diterapkan.

Adapun model k-NN penggunaannya cenderung lebih rumit karena harus mencari nilai k terbaik, kemudian setelah diperoleh nilai k terbaik akan dilakukan pengujian dengan variasi test size. Untuk mencari jarak dalam k-NN sebenarnya dapat dilakukan dengan beberapa metode selain euclidian, yaitu Manhattan atau minkowski. Namun dalam penelitian ini metrik euclidian yang digunakan.

BAB VI

PEMBAHASAN

6.1 Perbandingan k-NN dan Naive Bayes

Pengujian setiap metode dilakukan dengan memvariasikan data uji mulai dari 90 – 10 sampai 55 – 45. Metode k-NN dan naive Bayes masing-masing diuji sebanyak delapan kali. Hasil pengujian masing-masing metode dapat dilihat pada tabel 6.1. Selain itu, nilai rata-rata presisi dan recall K-NN juga lebih tinggi daripada Naive Bayes. Hal ini menunjukkan bahwa K-NN lebih mampu memprediksi calon penerima KIP yang benar-benar layak menerima bantuan.

Tabel 6.1 perbandingan hasil pengujian k-NN dan Naive Bayes

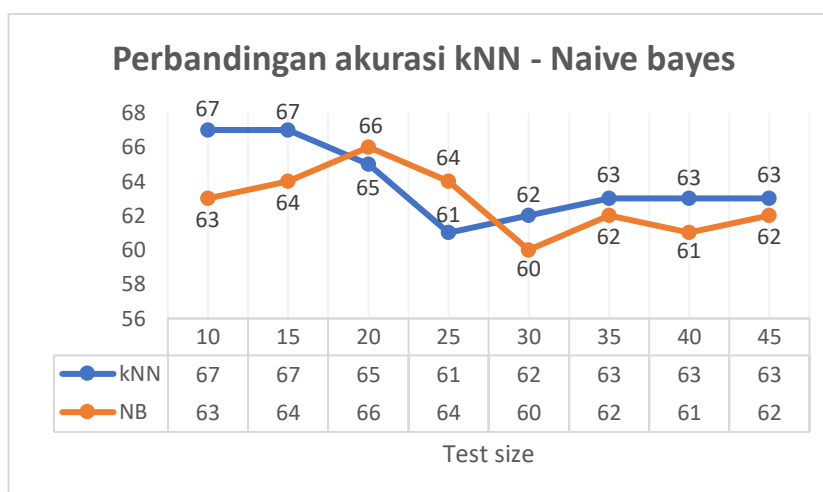
Test size	k-Nearest Neighbor (k=11)			Naive Bayes		
	Akurasi	Presisi	Recall	Akurasi	Presisi	Recall
90-10	67	67	96	63	65	90
85-15	67	69	91	64	67	90
80-20	65	67	93	66	67	91
75-25	61	64	90	64	67	90
70-30	62	65	88	60	64	86
65-35	63	65	91	62	65	87
60-40	63	65	91	61	65	86
55-45	63	67	87	62	67	84
Rata-rata	63,88	66,13	90,88	62,75	65,88	88,00

6.2 Analisis Matrik Evaluasi

6.2.1 Perbandingan Akurasi

Akurasi K-NN cenderung meningkat seiring dengan peningkatan ukuran data uji. Hal ini menunjukkan bahwa K-NN semakin mampu memprediksi kelayakan calon penerima KIP dengan baik seiring dengan meningkatnya jumlah data yang digunakan. Adapun untuk akurasi Naive Bayes cenderung menurun

seiring dengan peningkatan ukuran data uji. Hal ini menunjukkan bahwa Naive Bayes semakin tidak mampu memprediksi kelayakan calon penerima KIP dengan baik seiring dengan meningkatnya jumlah data yang digunakan. Nilai rata-rata akurasi K-NN ternyata lebih tinggi, yaitu 63,88%, dibandingkan dengan nilai rata-rata akurasi Naive Bayes yang hanya 62,75%. Gambar 6.1 memperlihatkan trend perolehan akurasi metode k-NN dan Naive Bayes. K-NN memiliki nilai akurasi yang lebih tinggi ketika test size 10 dan 15 yaitu 67% dan Naive Bayes memiliki akurasi tertinggi pada test size 20 yaitu 66.

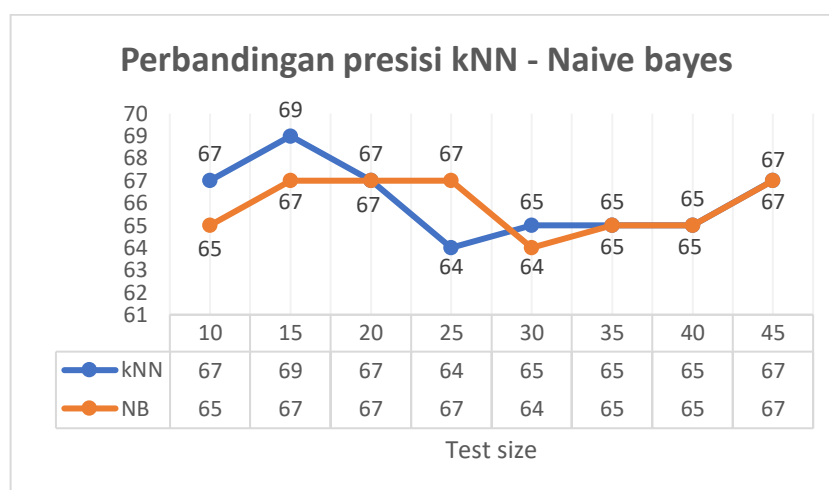


Gambar 6.1 Perbandingan Akurasi k-NN dan Naive Bayes

6.2.2 Perbandingan Presisi

Presisi K-NN cenderung meningkat seiring dengan peningkatan ukuran data uji. Kemudian mengalami penurunan setelah test size 15. Sementara Naive Bayes cenderung lebih stabil walaupun tidak memperoleh nilai presisi yang lebih tinggi dari k-NN. Pada pengujian dengan test size 15 sampai 25 Naive Bayes memiliki nilai presisi stabil yaitu 67 sementara k-NN memperoleh nilai tertinggi yaitu 69. Hal ini menunjukkan bahwa K-NN mampu memprediksi calon penerima KIP yang benar-benar layak menerima bantuan. Sedangkan Presisi Naive Bayes

cenderung lebih rendah walaupun lebih stabil seiring dengan peningkatan ukuran data uji. Nilai rata-rata presisi K-NN ternyata lebih tinggi, yaitu 66,13%, dibandingkan dengan nilai rata-rata presisi Naive Bayes yang hanya 65,88%. Gambar 6.2 menunjukkan perbandingan k-NN dan Naive Bayes dari nilai presisi. K-NN mampu mendapatkan nilai tertinggi yaitu 69 dibandingkan Naive Bayes yang hanya mendapatkan nilai 67.

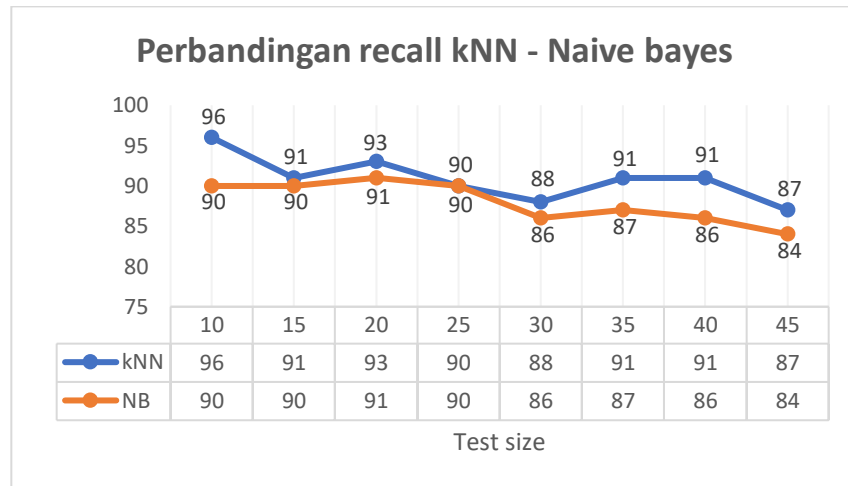


Gambar 6.2 Perbandingan Presisi k-NN dan Naive Bayes

6.1.3 Perbandingan *Recall*

K-NN cenderung memiliki nilai recall yang selalu lebih tinggi dibandingkan Naive Bayes seiring dengan peningkatan ukuran data uji. Nilai recall yang diperoleh kedua metode semakin menurun dengan meningkatnya data uji. Hal ini menunjukkan bahwa semakin tinggi data uji semakin rendah nilai recall. K-NN mampu mendapatkan skor recall tertinggi sampai 97 sementara Naive Bayes hanya mampu mendapatkan skor tertinggi sampai 91. Gambar 6.3 menunjukkan perbandingan k-NN dan Naive Bayes dari nilai recall. Terlihat jika K-NN hampir selalu mengungguli Naive Bayes mulai dari pengujian pertama sampai kedelapan.

Nilai rata-rata recall dari k-NN lebih tinggi yaitu 90,88% dibandingkan dengan naive Bayes yang hanya 88,00%.



Gambar 6.3 Perbandingan Recall k-NN dan Naive Bayes

Berdasarkan analisis trend yang ditemukan pada skor akurasi, presisi, dan recall tersebut, dapat disimpulkan bahwa metode K-NN lebih unggul daripada metode Naive Bayes untuk memprediksi kelayakan calon penerima KIP. Hal ini karena K-NN memiliki akurasi, presisi, dan recall yang lebih tinggi. Oleh karena itu, jika ingin memprediksi kelayakan calon penerima KIP, disarankan untuk menggunakan metode K-NN sebagai pilihan utama.

BAB VII

PENUTUP

5.1 Kesimpulan

Berdasarkan hasil pengujian yang dilakukan untuk menganalisis performa metode klasifikasi kelayakan calon penerima Kartu Indonesia Pintar (KIP) menggunakan k-Nearest Neighbor (k-NN) dan Naive Bayes, dapat diambil beberapa kesimpulan sebagai berikut:

1. Metode K-Nearest Neighbor (k-NN) menunjukkan performa yang lebih baik dibandingkan dengan metode Naive Bayes dalam melakukan klasifikasi kelayakan calon penerima beasiswa KIP Kuliah. k-NN memiliki tingkat akurasi, presisi, dan recall yang lebih tinggi daripada Naive Bayes, menandakan kemampuannya dalam mengidentifikasi calon penerima beasiswa yang benar-benar layak.
2. Pemilihan nilai k terbaik pada metode k-NN memberikan pengaruh yang positif dalam meningkatkan performa metode k-NN. Nilai k=11 merupakan nilai k terbaik dari nilai k bilangan ganjil antara 3 sampai 15. Maka, pemilihan nilai k terbaik diperlukan sebelum menggunakan metode k-NN untuk melakukan klasifikasi kelayakan calon penerima beasiswa KIP.
3. Peningkatan ukuran data uji memberikan dampak positif pada kinerja kedua metode. Semakin besar ukuran data uji yang digunakan, kinerja prediksi dari kedua metode cenderung meningkat. Oleh karena itu, disarankan untuk menggunakan sebanyak mungkin data uji yang tersedia agar hasil klasifikasi menjadi lebih andal dan akurat.

5.2 Saran

Berdasarkan kesimpulan yang diperoleh dari penelitian ini, berikut adalah beberapa saran untuk pengembangan dan penerapan klasifikasi kelayakan calon penerima beasiswa KIP Kuliah:

1. k-Nearest Neighbor (k-NN) menunjukkan performa yang lebih baik dalam penelitian ini, tetapi tetap perlu dilakukan eksplorasi lebih lanjut dalam penggunaan metode-metode lain atau pengembangan model prediksi yang lebih canggih. Misalnya, bisa dicoba metode lain seperti Support Vector Machine (SVM) atau Decision Tree untuk melihat apakah mereka dapat memberikan hasil yang lebih baik dalam prediksi kelayakan penerima KIP.
2. Penyetelan parameter yang tepat untuk kedua metode memang sangat penting, terutama pada k-NN, untuk mendapatkan hasil yang optimal. Percobaan dengan berbagai nilai k dalam k-NN atau melakukan penyetelan pada parameter lainnya dapat membantu meningkatkan kinerja prediksi.
3. Pengumpulan data yang lebih luas dan representatif tentang calon penerima KIP dapat meningkatkan ketepatan dan generalisasi model prediksi. Data yang lebih kaya akan membantu mengurangi bias dan meningkatkan keakuratan prediksi.
4. Jika hasil prediksi dari model telah terbukti andal, disarankan untuk mengintegrasikan model prediksi ini dengan sistem yang berhubungan langsung dengan program bantuan penerima KIP. Dengan demikian, proses seleksi penerima manfaat bisa lebih efisien dan akurat.

DAFTAR PUSTAKA

- Adeniyi, D. A., Wei, Z., & Yongquan, Y. (2018). Automated web usage data mining and recommendation system using K-nearest neighbor (KNN) classification method, Applied KJ Cha et al. *Journal of Mechanical Science and Technology*, 32(6), 2453–2462.
- Adi, S. (2018). Implementasi Algoritma Naive Bayes Classifier Untuk Klasifikasi Penerima Beasiswa PPA Di Universitas Amikom Yogyakarta. *Jurnal Mantik Penusa*, 22(1).
- Agama, D. (2011). *Al-Qur'an dan Tafsirnya (Edisi yang Disempurnakan)*. Widya Cahaya.
- Agama, K. (2020). *Beasiswa Bidikmisi Menjadi KIP Kuliah, Ini Dampaknya*. Kementerian Agama. <https://kemenag.go.id/read/beasiswa-bidikmisi-menjadi-kip-kuliah-ini-dampaknya-jj0pp>
- Agama, M. (2020). *KMA No. 361 Tentang Pedoman KIP Kuliah*. Kementerian Agama. <https://simpuh.kemenag.go.id/regulasi/KMA NO. 361 TAHUN 2020 TENTANG PEDOMAN KIP KULIAH .pdf>
- Ani, P. A., & Andri, A. (2022). Penerapan Algoritma Naive Bayes Untuk Klasifikasi Mahasiswa Penerima KIP Pada Universitas Bina Darma. *Bina Darma Conference on Computer Science (BDCSS)*, 4(1), 172–180.
- Arifin, T., & Syalwah, S. (2020). Prediksi keberhasilan immunotherapy pada penyakit kutil dengan menggunakan algoritma naïve bayes. *Jurnal Responsif: Riset Sains Dan Informatika*, 2(1), 38–43.
- Basori, R. (2020). *KIP Kuliah dan Perluasan Akses Pendidikan Tinggi*. Kementerian Agama. <https://kemenag.go.id/read/kip-kuliah-dan-perluasan-akses-pendidikan-tinggi-lmamo>
- Budiman, B., & Umami, I. (2022). Implementasi Algoritma Naïve Bayes Untuk Menentukan Calon Penerima Beasiswa Di SMK YPM 14 Sumobito Jombang. *Jurnal Teknologi Dan Sistem Informasi Bisnis*, 4(2), 446–454.
- Dahri, D., Agus, F., & Khairina, D. M. (2017). *Metode Naive Bayes Untuk Penentuan Penerima Beasiswa Bidikmisi Universitas Mulawarman*.
- Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI Magazine*, 17(3), 37.
- Ferdousy, Elma Zannatul, Md Mafijul Islam, and M. A. M. (2013). combination of naive bayes classifier and K-Nearest Neighbor (cNK) in the classification beased predictive models. *Computer and Information Science*, 6(3), 48.
- Id, I. D. (2021). *Machine Learning: Teori, Studi Kasus dan Implementasi Menggunakan Python (Vol. 1)*. Unri Press.
- Undang-Undang (UU) No 12 Tahun 2012 Tentang Pendidikan Tinggi, (2012).

<https://diktis.kemenag.go.id/prodi/dokumen/UU-Nomor-12-Tahun-2012-ttg-Pendidikan-Tinggi.pdf>

- Jiang, L., Zhang, H., & Cai, Z. (2006). *Dynamic K-Nearest-Neighbor Naive Bayes with Attribute Weighted BT - Fuzzy Systems and Knowledge Discovery* (L. Wang, L. Jiao, G. Shi, X. Li, & J. Liu (eds.); pp. 365–368). Springer Berlin Heidelberg.
- Kholil, M. (2018). Penerapan Metode K Nearest Neighbord Dalam Proses Seleksi Penerima Beasiswa. *Proceeding Seminar Nasional Sistem Informasi Dan Teknologi Informasi, 1*(1), 13–18.
- Lee, S. (2015). Hybrid Naïve Bayes K-nearest neighbor method implementation on speech emotion recognition. *2015 IEEE Advanced Information Technology, Electronic and Automation Control Conference (IAEAC)*, 349–353. <https://doi.org/10.1109/IAEAC.2015.7428573>
- Mulyani, A., Kurniadi, D., Nashrulloh, M. R., Julianto, I. T., & Regita, M. (2022). The Prediction Of PPA And KIP-Kuliah Scholarship Recipients Using Naive Bayes Algorithm. *Jurnal Teknik Informatika (Jutif)*, 3(4), 821–827.
- Nikmatun, I. A., & Waspada, I. (2019). Implementasi Data Mining untuk Klasifikasi Masa Studi Mahasiswa Menggunakan Algoritma K-Nearest Neighbor. *Simetris: Jurnal Teknik Mesin, Elektro Dan Ilmu Komputer*, 10(2), 421–432.
- Noviana, D., Susanti, Y., & Susanto, I. (2019). Analisis Rekomendasi Penerima Beasiswa Menggunakan Algoritma K-Nearest Neighbor (K-NN) dan Algoritma C4. 5. *Seminar & Conference Proceedings of UMT*.
- Rektor, S. K. (2022). *Petunjuk Teknis Seleksi Program Kartu Indonesia Pintar Kuliah Universitas Islam Negeri Maulana Malik Ibrahim Malang Bagi Mahasiswa Angkatan Tahun 2022*. Universitas Islam Negeri Maulana Malik Ibrahim; <https://kip.uin-malang.ac.id/>. <https://kip.uin-malang.ac.id/docs/Juknis-KIP-Tahun-2022.pdf>
- Risman, H., Nugroho, D., & Utami, Y. R. W. (2015). Penerapan Metode K-Nearest Neighbor Pada Aplikasi Penentu Penerima Beasiswa Mahasiswa Di STMIK Sinar Nusantara Surakarta. *Jurnal Teknologi Informasi Dan Komunikasi (TIKOMSiN)*, 3(2).
- Riyyan, M., & Firdaus, H. (2022). Perbandingan Algoritma Naive Bayes dan KNN Terhadap Data Penerimaan Beasiswa (Studi Kasus Lembaga Beasiswa Baznas Jabar). *Jurnal Informatika Dan Rekayasa Elektronik*, 5(1), 1–10.
- Romadhon, M. R., & Kurniawan, F. (2021). A comparison of naive Bayes methods, logistic regression and KNN for predicting healing of Covid-19 patients in Indonesia. *2021 3rd East Indonesia Conference on Computer and Information Technology (Eiconcit)*, 41–44.
- Suganda, G., Asfi, M., Subagio, R. T., & Kusuma, R. P. (2022). Penentuan

- Penerima Bantuan Beasiswa Kartu Indonesia Pintar (KIP) Kuliah Menggunakan Naive Bayes Classifier. *JSiI (Jurnal Sistem Informasi)*, 9(2), 193–199.
- Sulistiani, H. (2018). *Penerapan Algoritma Klasifikasi Sebagai Pendukung Keputusan Pemberian Beasiswa Mahasiswa*.
- Sumarlin, S. (2015). Implementasi Algoritma K-Nearest Neighbor Sebagai Pendukung Keputusan Klasifikasi Penerima Beasiswa PPA dan BBM. *JSINBIS (Jurnal Sistem Informasi Bisnis)*, 5(1), 52–62.
- Sumiah, A., & Mirantika, N. (2020). Perbandingan Metode K-Nearest Neighbor dan Naive Bayes untuk Rekomendasi Penentuan Mahasiswa Penerima Beasiswa pada Universitas Kuningan. *Buffer Informatika*, 6(1), 1–14.
- University, B. (2017). *Fishbone diagram*.
<https://sis.binus.ac.id/2017/05/15/fishbone-diagram/>
- Verdhan, V. (2020). *Supervised Learning with Python*. Apress Berkeley, CA.
<https://doi.org/https://doi.org/10.1007/978-1-4842-6156-9>
- Wahbah, A.-Z. (2013). Tafsir Al-Munir, terj. Oleh Abdul Hayyie Al-Kattani Dkk, *Jilid, 15*, 556.
- Wibawa, A. P., Guntur, M., Purnama, A., Akbar, M. F., & Dwiyanto, F. A. (2018). Metode-metode Klasifikasi. *Prosiding Seminar Ilmu Komputer Dan Teknologi Informasi*, 3(1).
- Zhang, Y. (2018). The development of Bayesian theory and its applications in business and bioinformatics. *IOP Conference Series: Earth and Environmental Science*, 128(1), 12120.
- Zuhaili, W. (n.d.). *Surat An-Nisa' ayat 85*. TafsirWeb. Retrieved February 1, 2023, from <https://tafsirweb.com/1617-surat-an-nisa-ayat-85.html>
- ديوان الإمام الشافعي. مكتبة ابن سينا. (n.d.). الشافعي, ا
دار إحياء الكتب (ed.). م. ف. ع. الباقي) سنن ابن ماجه. (n.d.). القزويني, ا. م. أ. ع. ا. م. ب. ي
العربية