

Prediksi Keterlambatan Pembayaran SPP Siswa dengan Pendekatan Metode Naive Bayes dan K-Nearest Neighbors

Deden Moh Alfiansyah*, Hari Soetanto

Magister Ilmu Komputer, Teknologi Informasi, Universitas Budi Luhur, Jakarta, Indonesia

Email: ¹*2111600983@student.budiluhur.ac.id, ² hari.soetanto@budiluhur.ac.id

Email Penulis Korespondensi: 2111600983@student.budiluhur.ac.id

Submitted: 01/12/2023; Accepted: 30/03/2024; Published: 30/03/2024

Abstrak—Pendidikan di Indonesia menghadapi tantangan biaya yang berdampak pada penyelenggaraan pendidikan, terutama Sumbangan Pembinaan Pendidikan (SPP) di sekolah swasta. Penelitian ini bertujuan untuk meningkatkan efisiensi administrasi pembayaran SPP siswa di SMK Wirasaba Karawang melalui pendekatan klasifikasi. Data pembayaran satu semester tahun 2023 sebanyak 725 data digunakan untuk memprediksi keterlambatan pembayaran. Sekitar 22% siswa mengalami keterlambatan. Dengan memahami pola keterlambatan, penelitian ini mengusulkan solusi untuk meningkatkan efisiensi administrasi dengan tindakan preventif yang sesuai. Harapannya, hasil dari penelitian ini dapat memberikan manfaat bagi SMK Wirasaba Karawang serta berkontribusi pada pengembangan sistem administrasi pendidikan yang lebih efisien, sehingga meningkatkan layanan pendidikan di Indonesia secara keseluruhan. Penelitian ini menguraikan masalah keterlambatan pembayaran SPP, menerapkan metode klasifikasi untuk memprediksi keterlambatan tersebut, bertujuan untuk meningkatkan efisiensi administrasi pembayaran, serta berpotensi memberikan solusi preventif yang dapat mengurangi keterlambatan pembayaran. Kontribusi dari penelitian ini adalah pengembangan sistem administrasi pendidikan yang lebih efisien melalui pendekatan teknologi informasi, dengan hasil sementara berupa analisis pola keterlambatan pembayaran yang didasarkan pada data tahun 2023 dari SMK Wirasaba Karawang.

Kata Kunci: Data Mining; Metode Klasifikasi; Naive Bayes; K-Nearest Neighbors (KNN)

Abstract—Education in Indonesia faces cost challenges which have an impact on the provision of education, especially Education Development Contributions (SPP) in private schools. This research aims to improve the administrative efficiency of student tuition payments at Wirasaba Karawang Vocational School through a classification approach. 725 payment data for one semester in 2023 are used to predict payment delays. About 22% of students experience delays. By understanding delay patterns, this research proposes solutions to improve administrative efficiency with appropriate preventive measures. The hope is that the results of this research can provide benefits to Vocational School Wirasaba Karawang and contribute to the development of a more efficient education administration system, thereby improving education services in Indonesia as a whole. This research describes the problem of late SPP payments, applies a classification method to predict these delays, aims to increase the efficiency of payment administration, and has the potential to provide preventive solutions that can reduce late payments. The contribution of this research is the development of a more efficient education administration system through an information technology approach, with interim results in the form of analysis of late payment patterns based on 2023 data from Wirasaba Karawang Vocational School.

Keywords: Data Mining; Classification Methods; Naive Bayes; K-Nearest Neighbors (KNN)

1. PENDAHULUAN

Pendidikan memegang peran krusial dalam membangun fondasi suatu bangsa, terutama di Indonesia yang sedang berkembang pesat. Seiring dengan perkembangan zaman, pendidikan dihadapkan pada beragam tantangan yang semakin kompleks. Dalam konteks Indonesia, salah satu tantangan utama adalah masalah pembiayaan pendidikan, terutama di tingkat menengah. Meskipun kerangka pendidikan menengah telah diatur dalam Peraturan Menteri Pendidikan dan Kebudayaan (Permen Dikbud) Nomor 19 Tahun 2016 yang berkaitan dengan Program Indonesia Pintar [1], pendanaan pendidikan menengah masih menjadi perhatian utama karena sebagian besar tanggung jawabnya berada pada siswa dan orang tua mereka [2]. Meskipun beberapa penelitian sebelumnya telah mencoba menjelaskan dan mengatasi tantangan pembiayaan pendidikan, terutama di Indonesia, pemahaman dan penerapan solusi yang efektif masih belum optimal, khususnya dalam konteks pendidikan menengah [3]. Oleh karena itu, diperlukan penelitian lanjutan untuk menjembatani kesenjangan dalam wacana akademis dan meningkatkan pemahaman tentang materi pelajaran yang relevan. Penelitian ini bertujuan untuk mengisi kekosongan pemahaman ilmiah dengan melakukan penelitian mendalam mengenai pembiayaan pendidikan menengah, dengan fokus khusus pada keterlambatan pengiriman Sumbangan Pembinaan Pendidikan (SPP) di Sekolah Menengah Kejuruan (SMK) Wirasaba di Karawang. Dengan mempertimbangkan faktor-faktor yang mendasari keterlambatan pencairan biaya sekolah dan membangun kerangka prognostik, penelitian ini diharapkan dapat memberikan kemajuan signifikan dalam meningkatkan efektivitas tata kelola pendidikan dan pembuatan kebijakan di Indonesia [4].

Menggunakan metodologi komprehensif dan teknik klasifikasi seperti *Naive Bayes* dan *K-Nearest Neighbors*, penelitian ini bertujuan untuk meningkatkan pemahaman tentang tren yang terkait dengan keterlambatan pencairan Rencana Pembayaran Siswa (SPP). Melalui analisis catatan historis pembayaran dari tahun sebelumnya, penelitian ini diharapkan dapat memberikan pencerahan yang signifikan bagi penerapan tindakan pencegahan yang lebih efektif oleh lembaga pendidikan dan pemangku kepentingan terkait [5]. Penelitian ini memberikan kontribusi signifikan dalam pengembangan ilmu pengetahuan dalam dua aspek utama. Pertama, dengan menggunakan metode gabungan *Naive Bayes* dan *K-Nearest Neighbors* (KNN), penelitian ini menunjukkan pendekatan holistik dalam memprediksi keterlambatan pembayaran SPP siswa. Menggabungkan kekuatan kedua metode ini, kita dapat mengatasi kelemahan

masing-masing model dan mencapai tingkat akurasi yang lebih tinggi [6]. Kedua, penelitian ini memberikan kontribusi baru dalam pemodelan prediksi keterlambatan pembayaran SPP siswa dengan fokus pada aplikasi praktis dari ilmu data dan pembelajaran mesin, memberikan wawasan berharga bagi institusi pendidikan dan pengambil kebijakan dalam mengelola arus kas dan merencanakan strategi yang efektif dalam menangani keterlambatan pembayaran SPP [7].

Beberapa penelitian telah dilakukan untuk memprediksi keterlambatan pembayaran Sumbangan Pembinaan Pendidikan (SPP) siswa dengan menggunakan berbagai metode klasifikasi. Penelitian Setyawan dkk (2019) melakukan penelitian menggunakan metode C4.5 dan *Random Forest* untuk memprediksi keterlambatan pembayaran SPP di SMK Al-Islam Surakarta. Hasil penelitian menunjukkan bahwa metode *Random Forest* mampu mencapai akurasi tertinggi sebesar 82,8% [8]. Penelitian Utomo dkk (2019) menggunakan metode *K-Nearest Neighbor* dan Regresi Linear untuk memprediksi harga emas. Temuan penelitian menunjukkan bahwa metode *K-Nearest Neighbor* memiliki tingkat akurasi yang lebih tinggi dibandingkan dengan Regresi Linear [9]. Selain itu, Budiman dkk (2019) menerapkan metode *Promethee* untuk menentukan penerima dana desa di Kecamatan Godean Kabupaten Sleman. Hasil penelitian menunjukkan bahwa penggunaan metode *Promethee* dapat membantu pemerintah dalam menetapkan penerima dana desa secara lebih adil dan transparan [10]. Oleh karena itu, berbagai metode klasifikasi telah digunakan dalam konteks prediksi keterlambatan pembayaran SPP serta bidang lainnya, menunjukkan keragaman pendekatan yang dapat digunakan dalam analisis prediktif. Penelitian ini tidak hanya berkontribusi pada pengembangan ilmu pengetahuan dalam domain ilmu data dan pembelajaran mesin, tetapi juga memberikan panduan praktis bagi praktisi dan pengambil kebijakan dalam meningkatkan manajemen keuangan sekolah serta meningkatkan layanan pendidikan untuk siswa.

2. METODOLOGI PENELITIAN

2.1 Alur Penelitian

Penelitian ini menggunakan metodologi kuantitatif dengan menggunakan teknik Data Mining untuk mengkaji dan memperkirakan keterlambatan pembayaran SPP siswa di SMK Wirasaba yang berlokasi di Karawang. Pilihan metodologi ini memfasilitasi penerapan metodologi Data Mining untuk membedakan pola atau asosiasi dalam kumpulan data yang berkaitan dengan pembayaran uang sekolah siswa. Selanjutnya, bagian berikutnya menjelaskan metodologi penelitian yang diusulkan [6].

a. Identifikasi Masalah

Penggambaran permasalahan dalam kerangka penelitian ini menjadi landasan penting untuk memahami tantangan yang dihadapi SMK Wirasaba Karawang terkait keterlambatan pembayaran biaya sekolah siswa. Beberapa permasalahan yang perlu diselidiki meliputi:

1. Tantangan utama yang dihadapi SMK Wirasaba Karawang adalah keterlambatan pembayaran SPP siswa. Kejadian seperti ini dapat menimbulkan kerawanan keuangan bagi lembaga tersebut dan memberikan dampak buruk terhadap penyelenggaraan pendidikan berkualitas tinggi.
2. Penundaan pembayaran biaya sekolah dapat disebabkan oleh berbagai macam pengaruh, yang mencakup keadaan ekonomi rumah tangga siswa, kesadaran akan pentingnya memenuhi kewajiban biaya sekolah, kendala dalam infrastruktur pembayaran institusional, dan variabel terkait lainnya. Memahami keragaman dan saling mempengaruhi elemen-elemen penyebab ini sangat penting untuk merancang solusi yang efektif.
3. Keterlambatan pembayaran uang sekolah berpotensi menghambat efektivitas operasional administrasi sekolah, termasuk fasilitasi pengumpulan uang sekolah dan prosedur pengelolaannya. Keadaan seperti ini dapat memberikan tekanan tambahan pada pegawai administratif dan menghambat kapasitas lembaga untuk memberikan pelayanan pendidikan yang optimal..

Penelitian ini berupaya untuk mengetahui faktor-faktor yang mendasari terjadinya keterlambatan pengiriman uang sekolah siswa di SMK Wirasaba Karawang. Dengan memahami isu-isu ini, pendekatan ini berupaya untuk mendapatkan pemahaman yang komprehensif, sehingga memfasilitasi pengembangan strategi terkait yang bertujuan untuk mengurangi keterlambatan pembayaran biaya sekolah dan meningkatkan efektivitas proses administrasi sekolah [7].

b. Analisis Masalah

Penelitian ini berupaya untuk meramalkan terjadinya keterlambatan pembayaran biaya pendidikan siswa di SMK Wirasaba Karawang, hal ini menjadi perhatian dalam rangka optimalisasi efektivitas administrasi pendidikan [8]. Kajian menyeluruh terhadap masalah ini sangat penting untuk menjelaskan faktor-faktor penentu keterlambatan pembayaran uang sekolah siswa dan memastikan variabel-variabel yang berpotensi berdampak pada fenomena ini [9].

1. Keterlambatan pengiriman biaya sekolah merupakan hambatan yang signifikan terhadap efektivitas operasional administrasi pendidikan di SMK Wirasaba. Keterlambatan tersebut berpotensi menghambat kelancaran operasional keuangan lembaga tersebut dan akibatnya membatasi kapasitasnya dalam memberikan layanan pendidikan berkualitas tinggi.

2. Keterlambatan pembayaran biaya sekolah dapat menimbulkan dampak yang luas, seperti hambatan dalam pengadaan sumber daya pendidikan, remunerasi dosen, dan pemeliharaan infrastruktur. Akibatnya, hal ini dapat berdampak buruk pada pengalaman dan standar pendidikan yang dapat diakses oleh siswa.
3. Faktor-faktor yang berkontribusi terhadap keterlambatan pembayaran biaya sekolah mencakup berbagai macam pengaruh, seperti keadaan sosial-ekonomi rumah tangga siswa, tingkat kesadaran mengenai pentingnya pendidikan, efektivitas mekanisme administrasi sekolah, dan protokol khusus yang mengatur penyelesaian biaya pendidikan yang ditetapkan oleh lembaga pendidikan.
4. Memperkirakan kemungkinan keterlambatan pembayaran biaya sekolah dapat membekali lembaga pendidikan dengan wawasan yang berharga, memungkinkan penerapan langkah-langkah proaktif, seperti memberikan peringatan kepada siswa yang berisiko atau memberikan bantuan keuangan kepada keluarga kurang mampu.

Dengan melakukan kajian menyeluruh terhadap masalah ini, kami bertujuan untuk meningkatkan pemahaman kita tentang rumitnya keterlambatan penyelesaian biaya sekolah siswa di SMK Wirasaba Karawang. Penjelasan tersebut akan menjadi landasan yang kuat untuk merumuskan langkah-langkah efektif yang bertujuan untuk memitigasi tantangan-tantangan ini dan meningkatkan efektivitas tata kelola pendidikan di lembaga-lembaga akademik.

c. Kesimpulan dan Saran

Pada saat ini, temuan investigasi ini menunjukkan bahwa penggunaan teknik *Naive bayes* menunjukkan efektivitas yang lebih besar dalam memperkirakan keterlambatan pengiriman uang sekolah siswa di SMK Wirasaba Karawang dibandingkan dengan pendekatan K-Nearest Neighbours. Berbagai faktor penentu yang meliputi bidang studi siswa, ketergantungan keluarga, rekam jejak pembayaran sebelumnya, dan faktor terkait lainnya memberikan pengaruh yang besar terhadap keterlambatan pembayaran. Oleh karena itu, pemahaman yang mendalam terhadap variabel-variabel ini menjadi sangat penting dalam meningkatkan kemampuan prognostik dan memitigasi terjadinya keterlambatan pencairan SPP.

2.2 Metode Pemilihan Sampel

Dalam penelitian ini, prosedur pemilihan sampel sangat penting untuk memastikan keterwakilan dan validitas data yang digunakan dalam analisis. Metodologi penelitian yang digunakan adalah purposive sampling, yaitu metode yang dipilih karena sesuai dengan tujuan penelitian, yaitu siswa yang mempunyai riwayat keterlambatan pembayaran SPP di SMK Wirasaba Karawang [10]. Kelompok penelitian ini mencakup seluruh siswa yang terdaftar di SMK Wirasaba Karawang, yang tersebar di tujuh jurusan berbeda. Melalui penggambaran populasi yang terdefinisi dengan jelas, pendekatan purposive sampling memungkinkan pemilihan sampel yang memenuhi kriteria tertentu, terutama mahasiswa yang memiliki rekam jejak keterlambatan pembayaran biaya sekolah. Pemilihan sampel didasarkan pada kriteria mahasiswa yang mengalami keterlambatan pembayaran biaya pendidikan pada beberapa periode sebelumnya. Kriteria seleksi ini dianggap penting karena siswa yang pernah mengalami keterlambatan pembayaran cenderung menghadapi tantangan serupa di masa depan [11].

Proses pemilihan sampel yang mempertimbangkan variabilitas di berbagai jurusan diharapkan dapat memberikan pemahaman yang komprehensif mengenai faktor-faktor yang mempengaruhi keterlambatan pembayaran SPP di SMK Wirasaba Karawang. Langkah-langkah operasional dalam prosedur pemilihan sampel memerlukan identifikasi mahasiswa dengan riwayat keterlambatan pembayaran dari masing-masing departemen, kemudian menunjuk individu-individu ini sebagai subjek penelitian [12]. Penentuan ukuran sampel akan bergantung pada keterbatasan sumber daya dan ketersediaan data terkait. Melalui penerapan metode purposive sampling, penelitian ini bertujuan untuk memberikan gambaran yang lebih mendalam mengenai faktor-faktor yang mendasari keterlambatan pembayaran SPP siswa di SMK Wirasaba Karawang, sehingga memberikan landasan yang kuat untuk analisis dan rekomendasi selanjutnya [13].

2.3 Metode Pengumpulan Data

Proses akuisisi data mempunyai arti penting dalam upaya penelitian, karena berfungsi sebagai sarana untuk mendapatkan informasi yang diperlukan untuk menjawab pertanyaan penelitian terkait [14]. Dalam konteks penelitian ini, beragam metodologi pengumpulan data digunakan untuk mendapatkan wawasan tentang terjadinya keterlambatan pembayaran uang sekolah siswa di lingkungan akademik SMK Wirasaba yang berlokasi di Karawang [15].

- a. Wawancara digunakan sebagai sarana untuk memperoleh wawasan tambahan mengenai faktor-faktor yang mendasari keterlambatan pencairan SPP. Sebuah kelompok yang terdiri dari siswa, orang tua, dan tenaga administrasi dari SMK Wirasaba Karawang diwawancarai untuk menjelaskan secara langsung pengalaman mereka mengenai pengiriman uang sekolah dan faktor-faktor yang mempengaruhi keterlambatan mereka.
- b. Kuesioner merupakan instrumen yang digunakan untuk pengumpulan data dari sekelompok siswa yang terdaftar di SMK Wirasaba Karawang. Disusun untuk memperoleh tanggapan mengenai variabel-variabel yang berpotensi mempengaruhi ketepatan waktu pengiriman uang SPP, termasuk keadaan sosial ekonomi keluarga, kesadaran akan pentingnya pengiriman uang SPP, dan persepsi siswa terhadap prosedur pembayaran.
- c. Teknik observasi diterapkan untuk mencermati secara langsung protokol pengiriman uang sekolah di SMK Wirasaba Karawang. Berdasarkan pengamatan ini, para peneliti berupaya untuk mencapai pemahaman

komprehensif tentang modalitas pembayaran biaya kuliah institusi dan untuk menunjukkan tantangan atau hambatan laten di dalamnya.

Melalui pemanfaatan beragam teknik akuisisi data, penelitian ini berupaya untuk mendapatkan wawasan mendalam tentang keterlambatan pengiriman uang sekolah siswa di SMK Wirasaba Karawang. Penggabungan data sekunder, wawancara, kuesioner, dan observasi diharapkan dapat memberikan pemahaman yang berbeda mengenai faktor-faktor yang mempengaruhi keterlambatan pencairan SPP, sehingga memfasilitasi perumusan rekomendasi terkait yang bertujuan untuk memitigasi masalah ini [16].

2.4 Teknik Analisis, Rancangan dan Pengujian

a. Teknik Analisis

Dalam studi ini, pengujian akan dilakukan untuk menyandingkan kemandirian dua metodologi prediksi yang berbeda, khususnya *Naive bayes* dan *K-Nearest Neighbors*. Penilaian tersebut akan mencakup evaluasi akurasi, presisi, recall, dan skor F1 yang dikaitkan pada masing-masing metode, yang bertujuan untuk membedakan kelebihan dan kekurangan masing-masing metode dalam meramalkan terjadinya tunggakan pembayaran SPP siswa di SMK Wirasaba. Melalui analisis komparatif terhadap hasil prediksi yang dihasilkan oleh kedua metodologi, penyelidikan ini berupaya untuk menemukan pendekatan yang optimal dan efektif yang berkaitan dengan kerangka penelitian yang digambarkan di sini [17].

1. Persiapan Data

Data yang mencakup variabel-variabel seperti kenakalan siswa, pendapatan orang tua, status pekerjaan orang tua, penerima bantuan uang sekolah yang terlambat, dan faktor-faktor terkait akan dikurasi dengan cermat sebagai persiapan untuk analisis. Proses ini memerlukan pembersihan data yang ketat, penghapusan nilai yang hilang, dan prosedur transformasi data potensial yang dianggap tepat.

2. Pembagian Data

Kumpulan data akan dipartisi menjadi dua subset berbeda: set pelatihan dan set pengujian. Subset pelatihan akan digunakan untuk tujuan pelatihan model, sedangkan subset pengujian akan digunakan untuk menilai kemandirian dan kinerja model yang dilatih.

3. Pelatihan Model *Naive bayes*

Investigasi ini akan menggunakan kumpulan data pelatihan untuk memfasilitasi pelatihan model *Naive bayes*. Model ini akan memperoleh wawasan tentang pola pembayaran SPP melalui data pelatihan dan selanjutnya menggunakan pengetahuan ini untuk memberikan prediksi pada kumpulan data pengujian.

4. Pelatihan Model *K-Nearest Neighbors*

Dalam penelitian ini, kami bertujuan untuk menggunakan algoritma *K-Nearest Neighbors* melalui kumpulan data pelatihan yang ditentukan. Algoritme ini berupaya membedakan titik data tetangga terdekat dengan pengamatan baru melalui metrik jarak, kemudian memanfaatkan label dominan di antara titik data tetangga tersebut untuk menyimpulkan prediksi.

5. Evaluasi Model

Setelah pelatihan untuk kedua model selesai, penelitian ini bermaksud menggunakan kumpulan data pengujian untuk menilai kinerja masing-masing model. Metrik evaluasi utama yang mencakup akurasi, presisi, perolehan, dan skor F1 akan dihitung untuk memfasilitasi analisis komparatif dari kemandirian yang ditunjukkan oleh algoritma *Naive bayes* dan *K-Nearest Neighbors*.

6. Analisis Hasil

Hasil penilaian akan diteliti dengan cermat untuk menggambarkan kelebihan dan kekurangan yang melekat pada setiap pendekatan. Analisis ini bertujuan untuk mengetahui apakah terdapat keefektifan kedua model yang meremehkan kedua model tersebut dan berupaya menjelaskan variabel-variabel yang berpotensi mempengaruhi hasil-hasilnya.

b. Rancangan

Selama fase desain penelitian, penjelasan komprehensif tentang prosedur yang harus dilakukan untuk pengumpulan data, penerapan teknik analisis, dan penilaian hasil penelitian akan diberikan. Selanjutnya, bagian berikutnya menggambarkan usulan desain penelitian yang akan diberlakukan [13].

1. Penentuan Variabel Penelitian

Variabel penelitian yang bersangkutan akan diuraikan dan dipilih dengan cermat, mencakup variabel terikat (pengiriman uang sekolah siswa tepat waktu) di samping variabel bebas (seperti jumlah tanggungan, catatan pembayaran masa lalu, pendapatan orang tua, dll.).

2. Pemilihan Sampel

Kelompok penelitian diambil dari demografi siswa di SMK Wirasaba, dengan mempertimbangkan distribusi heterogen di berbagai disiplin ilmu dan kedudukan akademik. Kriteria seleksi yang ketat akan diterapkan untuk memastikan komposisi sampel yang representatif selaras dengan karakteristik populasi yang lebih luas.

3. Pembagian Dataset

Kumpulan data akan dipartisi secara acak menjadi dua subkumpulan berbeda, satu ditetapkan sebagai data pelatihan, dan lainnya sebagai data pengujian. Pembagian ini akan mengikuti proporsi yang tepat, dijaga dengan cermat untuk menjaga keakuratan representasi setiap komponen dalam kumpulan data.

4. Penerapan Metode *Naive bayes*

Model *Naive bayes* akan menjalani pelatihan menggunakan dataset pelatihan yang disediakan. Metodologi validasi silang selanjutnya akan digunakan untuk menilai kemandirian model, sehingga memastikan generalisasi hasil prediksi yang kuat.

5. Penerapan Metode *K-Nearest Neighbors*

Algoritma *K-Nearest Neighbors* (KNN) akan menjalani pelatihan dengan memanfaatkan dataset pelatihan yang disediakan. Selanjutnya, evaluasi kemandirian model dan kapasitasnya untuk mengekstrapolasi pola dari data akan dilakukan melalui teknik validasi silang.

6. Evaluasi dan Perbandingan Kinerja

Kemandirian kedua model akan dinilai melalui penerapan metrik terkait termasuk akurasi, presisi, perolehan, dan skor F1. Analisis komparatif akan dilakukan untuk memastikan pendekatan yang unggul dalam meramalkan keterlambatan pembayaran SPP siswa di SMK Wirasaba.

c. Pengujian Model

Tahap evaluasi penelitian ini merupakan titik penting yang bertujuan untuk meneliti kemandirian model prediktif. Fase ini memerlukan pemanfaatan dua metodologi utama: menggunakan matriks konfusi dan melakukan pengujian kotak hitam. Awalnya, matriks konfusi berfungsi untuk menjelaskan keakuratan prediksi model mengenai keterlambatan pembayaran biaya sekolah siswa dengan menyandingkan hasil yang diproyeksikan dengan nilai sebenarnya dalam kumpulan data pengujian. Metrik yang dihasilkan meliputi True Positive (TP), True Negative (TN), False Positive (FP), dan False Negative (FN) selanjutnya dimanfaatkan untuk menghitung metrik kinerja termasuk akurasi, presisi, recall, dan skor F1 untuk kedua prognostik teknik. Selain itu, pengujian black box dilakukan untuk menilai secara komprehensif kemampuan prognostik model sehubungan dengan keterlambatan pembayaran uang sekolah siswa, tanpa mempelajari mekanisme internalnya. Beragam kumpulan data pengujian digunakan dalam penilaian ini untuk mengukur konsistensi dan keakuratan prediksi yang dihasilkan oleh kedua metodologi. Analisis hasil gabungan yang diperoleh dari metodologi-metodologi ini menawarkan wawasan komprehensif mengenai kemandirian dan penerapan model *Naive bayes* dan *K-Nearest Neighbors* dalam domain spesifik peramalan keterlambatan pembayaran uang sekolah siswa di SMK Wirasaba. Pemahaman yang didapat dari proses evaluatif ini sangat penting dalam menentukan metodologi implementasi yang optimal dalam lingkungan pendidikan, sehingga memfasilitasi peningkatan efisiensi administratif dan kualitas ketentuan pendidikan [18].

2.5 Tahapan Penelitian

Fase penelitian ini menganut kerangka metodologi studi kasus, yang menggabungkan komponen evaluasi dan analisis data. Tujuan utamanya adalah untuk memfasilitasi pemahaman komprehensif tentang tema penelitian yang telah diidentifikasi sebelumnya. Awalnya, data yang berkaitan dengan tujuan penelitian dikumpulkan, diikuti dengan pengorganisasiannya dan persiapan untuk analisis selanjutnya. Setelah pengumpulan data, dilakukan evaluasi menyeluruh yang bertujuan untuk memastikan kualitas dan relevansi data terhadap tujuan penelitian. Selain itu, tahap evaluasi ini berfungsi untuk menunjukkan potensi perbedaan atau kekurangan data yang memerlukan perbaikan atau pengawasan lebih lanjut [19].

Setelah tahap evaluasi selesai, analisis data dilakukan, yang bertujuan untuk membedakan pola, tren, atau korelasi antar variabel dalam kumpulan data. Hasil analisis ini menginformasikan rumusan kesimpulan dan rekomendasi yang berkaitan dengan tujuan penelitian [20]. Dengan demikian, tahap penelitian ini sangat penting dalam memahami fenomena yang diteliti dan memberikan kontribusi besar terhadap pengetahuan dalam bidang yang relevan. Tahapan penelitian berurutan yang dilakukan digambarkan sebagai berikut:

a. Identifikasi Masalah dan Penyusunan Tujuan Penelitian

Upaya tersebut mencakup pencermatan permasalahan terkait administrasi kiriman uang sekolah siswa di SMK Wirasaba. Hal ini melibatkan artikulasi tujuan penelitian yang bertujuan untuk memperkirakan keterlambatan pencairan biaya kuliah siswa melalui penerapan metodologi *Naive bayes* dan *K-Nearest Neighbors* (KNN).

b. Studi Pustaka dan Pengumpulan Informasi

Terlibat dalam tinjauan komprehensif literatur ilmiah untuk mendapatkan wawasan tentang kerangka konseptual dan landasan teoretis yang berkaitan dengan pengelolaan pembayaran uang sekolah siswa, di samping metodologi prediktif yang akan digunakan. Memperoleh data mengenai protokol dan metodologi yang mengatur pencairan biaya pendidikan siswa di SMK Wirasaba.

c. Perancangan Metodologi Penelitian

Mengembangkan kerangka metodologi penelitian yang komprehensif yang mencakup pemilihan metode, teknik pengumpulan data, dan metodologi analisis data yang tepat. Tetapkan kriteria inklusi dan eksklusi yang jelas untuk memandu proses pemilihan sampel data siswa.

d. Pengumpulan Data dan Persiapan Data

Mengumpulkan informasi terkait kiriman uang sekolah siswa di SMK Wirasaba dalam jangka waktu yang ditentukan. Melakukan aktivitas pra-pemrosesan data yang diperlukan, yang mencakup perbaikan anomali data dan penentuan variabel terkait.

e. **Pembangunan Model Prediksi**

Mengembangkan kerangka prediktif yang bertujuan untuk membedakan keterlambatan pembayaran biaya sekolah siswa, menggunakan algoritma *Naive bayes* dan teknik *K-Nearest Neighbors* (KNN). Upaya ini mencakup penggabungan model-model ini dengan metodologi pra-pemrosesan untuk meningkatkan kemanjuran dan akurasi prediksi.

f. **Validasi dan Pengujian Model**

Otentikasi model melalui penggunaan metodologi validasi silang untuk menilai ketepatan dan kapasitas generalisasi. Menilai kemanjuran model prediktif dengan menggunakan kumpulan data pengujian dan kriteria evaluasi terkait untuk mengukur kinerjanya.

g. **Analisis Hasil dan Interpretasi**

Periksa hasil pengujian model untuk menyimpulkan wawasan yang berkaitan dengan ketepatan prediksi dan kemanjuran komparatif metodologi *Naive bayes* dan KNN. Menjelaskan penemuan penelitian dalam rangka administrasi pembayaran SPP siswa di SMK Wirasaba.

h. **Pembuatan Aplikasi Prediksi dan Validasi**

Buat alat prediktif berbasis model yang dirancang untuk menyederhanakan perkiraan pengiriman uang sekolah siswa yang terlambat. Buktikan kemanjuran dan integritas operasional aplikasi melalui uji coba pengguna yang ketat dan penilaian kepuasan yang komprehensif.

i. **Penyusunan Laporan Penelitian**

Menghasilkan dokumen penelitian komprehensif yang mencakup keseluruhan proses penelitian, termasuk tahapan berurutan, penemuan, dan kesimpulan yang dihasilkan. Patuhi konvensi akademis dan institusional yang telah ditetapkan saat memformat dan menyajikan laporan penelitian.

3. HASIL DAN PEMBAHASAN

3.1 Hasil Penelitian

Naive bayes adalah sebuah metode klasifikasi dalam pembelajaran mesin yang mengacu pada teorema Bayes [1]. Teorema Bayes memberikan kerangka kerja untuk menghitung probabilitas posterior (probabilitas suatu kejadian setelah melihat bukti) dari suatu hipotesis berdasarkan pada bukti atau data yang diamati sebelumnya. Secara sederhana, *Naive bayes* menghitung probabilitas kelas target berdasarkan pada fitur-fitur yang diamati. Asumsi utama dalam *Naive bayes* adalah bahwa semua fitur dalam dataset adalah independen satu sama lain, meskipun kenyataannya, hal ini mungkin tidak selalu benar dalam konteks sebenarnya. Meskipun demikian, mesin pembelajaran dengan *Naive bayes* tetap efektif dalam banyak kasus. Langkah-langkah umum dalam *Naive bayes* adalah sebagai berikut:

- Menentukan target class
- Menghitung probabilitas prior
- Menghitung Probabilitas *Likelihood*
- Menguji data baru

Untuk melakukan pengujian model *Naive bayes*, langkah awalnya adalah menyiapkan dataset. Target yang akan diuji adalah atribut KETERLAMBATAN_SPP. Dalam dataset ini, terdapat sebanyak 22% data yang mengalami keterlambatan pembayaran SPP, sementara 78% lainnya tidak mengalami keterlambatan. Evaluasi ini akan memberikan gambaran yang jelas tentang seberapa baik model mampu mengklasifikasikan data antara yang mengalami keterlambatan dan yang tidak. Dengan demikian, langkah-langkah ini menjadi penting dalam proses pengujian dan validasi model *Naive bayes* sebelum digunakan dalam kasus nyata. Untuk Dalam melakukan prediksi, penentuan target class menjadi langkah penting yang tak terhindarkan. Hal ini karena setiap perhitungan dan analisis saling terkait, yang menjadikan pilihan kelas target, dalam hal ini "KETERLAMBATAN_SPP", menjadi krusial. Penetapan kelas target menjadi landasan bagi proses prediksi yang akurat dan dapat diandalkan, memungkinkan sistem untuk menghasilkan estimasi yang tepat terkait kemungkinan keterlambatan pembayaran SPP.

Menetapkan "KETERLAMBATAN_SPP" sebagai kelas target, sistem dapat memusatkan upaya pada faktor-faktor yang berpotensi mempengaruhi keterlambatan tersebut. Dalam konteks ini, prediksi menjadi lebih terfokus dan relevan, memungkinkan penggunaan data dan metode analisis yang sesuai untuk menghasilkan hasil yang lebih akurat dan bermanfaat bagi pengambilan keputusan terkait manajemen pembayaran SPP. Probabilitas prior menggambarkan kemungkinan munculnya setiap kelas target sebelum mempertimbangkan atribut-atribut lain dalam dataset. Dalam konteks ini, terdapat dua kelas target: "tidak" dan "ya" [11]. Probabilitas prior memberikan pandangan awal tentang seberapa sering setiap kelas target muncul dalam data tanpa memperhitungkan atribut-atribut yang terkait. Berikut adalah tabel 1 yang memberikan rincian lebih lanjut mengenai distribusi kelas target sebelum analisis atribut dilakukan.

Tabel 1. Menghitung Probabilitas Prior

KETERLAMBATAN_SPP		
	Ya	Tidak
P (ya/tidak)	22%	78%
Total	100%	

Pada Tabel 1 diatas, kolom "KETERLAMBATAN_SPP" dalam dataset ini memperlihatkan status keterlambatan pembayaran Sumbangan Pembinaan Pendidikan (SPP) siswa, yang dapat berupa "Ya" atau "Tidak". Selanjutnya, kolom "P (ya/tidak)" menampilkan probabilitas prior dari masing-masing status keterlambatan pembayaran SPP siswa. Probabilitas prior adalah probabilitas suatu peristiwa terjadi sebelum informasi tambahan diketahui. Dalam konteks ini, probabilitas prior mencerminkan probabilitas bahwa siswa akan terlambat membayar SPP sebelum data pelatihan dianalisis. Penyajian nilai probabilitas prior adalah sebagai berikut: 22% menunjukkan probabilitas prior siswa terlambat membayar SPP, sementara 78% menunjukkan probabilitas prior siswa tidak terlambat membayar SPP. Total nilai probabilitas prior adalah 100%, yang mencerminkan seluruh probabilitas dari kedua status keterlambatan pembayaran SPP siswa. Oleh karena itu, dapat disimpulkan dari Tabel 1 bahwa probabilitas prior siswa terlambat membayar SPP adalah 22%, sedangkan probabilitas prior siswa tidak terlambat membayar SPP adalah 78%. Hasil ini menandakan bahwa mayoritas siswa di SMK Wirasaba Karawang membayar SPP tepat waktu. Untuk perhitungan probabilitas *likelihood* untuk setiap atribut akan di jelaskan sebagai berikut ini:

Tabel 2. Probabilitas Likelihood Pekerjaan Orang Tua

PEKERJAAN_ORANG_TUA	Ya	Tidak
Petani	0	0,3
Serabutan	0,71	0
Karyawan	0	0
Pns	0	0,51
Guru	0	0
Pedagang	0	0
Nelayan	0	0
Wirausaha	0	0
Lainnya	0,29	0,2
Total	1	1

Pada Tabel 2 diatas, analisis probabilitas likelihood terhadap keterlambatan pembayaran SPP siswa berdasarkan jenis pekerjaan orang tua siswa menunjukkan pola yang menarik. Kolom PEKERJAAN_ORANG_TUA dalam data ini menggambarkan jenis pekerjaan orang tua siswa, sedangkan kolom "Ya" dan "Tidak" mewakili probabilitas likelihood siswa terlambat atau tidak terlambat membayar SPP berdasarkan pekerjaan orang tua yang tercantum. Dapat disimpulkan bahwa siswa dengan orang tua yang berprofesi sebagai serabutan memiliki probabilitas likelihood tertinggi untuk terlambat membayar SPP, mencapai 71%. Sementara itu, siswa dengan orang tua yang berprofesi sebagai petani, karyawan, guru, pedagang, nelayan, dan wirausaha memiliki probabilitas likelihood 0% untuk terlambat membayar SPP. Namun, terdapat pengecualian pada siswa dengan orang tua yang berprofesi sebagai PNS, di mana terdapat probabilitas likelihood sebesar 51% untuk tidak terlambat membayar SPP. Sedangkan siswa dengan orang tua yang berprofesi sebagai "lainnya" memiliki probabilitas likelihood 29% untuk terlambat membayar SPP dan 20% untuk tidak terlambat membayar SPP. Oleh karena itu, hasil analisis ini memberikan wawasan penting mengenai faktor-faktor yang memengaruhi keterlambatan pembayaran SPP siswa berdasarkan jenis pekerjaan orang tua. Data ini dapat digunakan sebagai dasar untuk pengembangan strategi yang lebih efektif dalam mengelola pembayaran SPP dan memberikan bantuan kepada siswa yang membutuhkan. Berikut ini adalah hasil dari *probabilitas likelihood* penerima bantuan:

Tabel 3. Probabilitas Likelihood Penerima Bantuan

PENERIMA_BANTUAN	Ya	Tidak
Ya	0,71	0
Tidak	0,29	1
Total	1	1

Pada Tabel 3 diatas, menggambarkan probabilitas *likelihood* penerima bantuan, yang merupakan probabilitas sebuah kejadian terjadi berdasarkan informasi yang tersedia, di mana informasi yang menjadi dasar adalah apakah penerima bantuan tersebut pernah hamil sebelumnya. Tabel tersebut terdiri dari dua kolom utama, yaitu "PENERIMA_BANTUAN" dan "Probabilitas". Kolom "PENERIMA_BANTUAN" menyatakan apakah penerima bantuan pernah mengalami kehamilan sebelumnya, dengan dua kemungkinan nilai, yakni "Ya" dan "Tidak". Sementara kolom "Probabilitas" mengindikasikan probabilitas bahwa penerima bantuan telah mengalami kehamilan



sebelumnya. Penjelasan rinci dari tabel ini adalah sebagai berikut: Baris pertama menunjukkan probabilitas bahwa penerima bantuan telah mengalami kehamilan sebelumnya sebesar 0,71. Artinya, dari setiap 100 penerima bantuan, 71 di antaranya telah mengalami kehamilan sebelumnya. Baris kedua menunjukkan probabilitas bahwa penerima bantuan tidak pernah mengalami kehamilan sebelumnya sebesar 0,29. Hal ini berarti, dari setiap 100 penerima bantuan, 29 di antaranya tidak pernah mengalami kehamilan sebelumnya. Sementara baris ketiga menunjukkan total probabilitas, yang mengindikasikan bahwa probabilitas bahwa penerima bantuan telah mengalami atau tidak mengalami kehamilan sebelumnya adalah 1. Dengan demikian, semua penerima bantuan pasti telah mengalami atau tidak mengalami kehamilan sebelumnya. Tabel ini dapat digunakan untuk memprediksi probabilitas bahwa seorang penerima bantuan telah mengalami kehamilan sebelumnya berdasarkan informasi apakah mereka telah mengalami kehamilan sebelumnya atau tidak. Sebagai contoh, jika diketahui bahwa seorang penerima bantuan telah mengalami kehamilan sebelumnya, maka probabilitas bahwa mereka menerima bantuan adalah sebesar 0,71. Dapat disimpulkan bahwa tabel *likelihood* penerima bantuan ini memperlihatkan probabilitas bahwa seorang penerima bantuan telah mengalami kehamilan sebelumnya berdasarkan informasi mengenai kehamilan sebelumnya atau tidak. Informasi ini dapat digunakan untuk memprediksi probabilitas seorang penerima bantuan menerima bantuan. Berikut ini adalah hasil dari *probabilitas likelihood* tunggakan:

Tabel 4. Probabilitas Likelihood Tunggakan

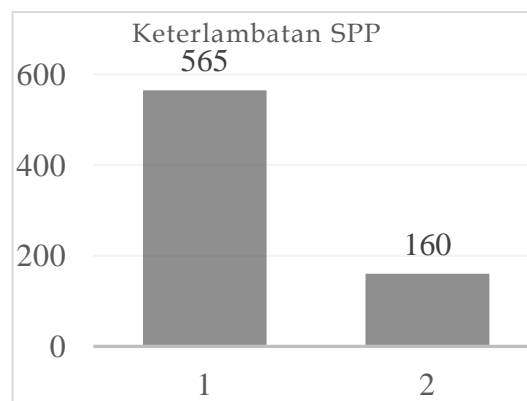
TUNGGAKAN	Ya	Tidak
Lunas	0,38	0,41
Besar	0,35	0,39
Sedang	0,12	0,1
Kecil	0,16	0,09
Total	1	1

Pada Tabel 4 diatas, probabilitas likelihood terkait tunggakan pembayaran, yaitu probabilitas bahwa seorang pelanggan akan mengalami tunggakan pembayaran berdasarkan status pembayaran mereka pada saat tertentu. Informasi yang tercantum dalam tabel adalah apakah pelanggan telah melunasi tagihan mereka, memiliki tunggakan besar, sedang, atau kecil. Tabel tersebut terdiri dari dua kolom, yakni "TUNGGAKAN" dan "Probabilitas". Kolom "TUNGGAKAN" menggambarkan status pembayaran saat ini, dengan empat kemungkinan nilai: Lunas, Besar, Sedang, dan Kecil. Sementara kolom "Probabilitas" menampilkan probabilitas bahwa pelanggan dengan status pembayaran tertentu akan mengalami tunggakan pembayaran di masa mendatang. Pada baris pertama tabel, terdapat probabilitas bahwa pelanggan yang telah melunasi tagihan mereka akan mengalami tunggakan pembayaran di masa mendatang sebesar 0,38. Ini mengindikasikan bahwa dari 100 pelanggan yang telah melunasi tagihan mereka, diperkirakan 38 di antaranya akan mengalami tunggakan pembayaran di masa mendatang. Pada baris kedua, probabilitas pelanggan dengan tunggakan besar mengalami tunggakan pembayaran di masa mendatang adalah 0,35. Dengan demikian, dari 100 pelanggan dengan tunggakan besar, sekitar 35 di antaranya diperkirakan akan mengalami tunggakan pembayaran di masa mendatang. Pada baris ketiga, probabilitas pelanggan dengan tunggakan sedang mengalami tunggakan pembayaran di masa mendatang adalah 0,12. Artinya, dari 100 pelanggan dengan tunggakan sedang, diperkirakan sekitar 12 di antaranya akan mengalami tunggakan pembayaran di masa mendatang. Pada baris keempat, probabilitas pelanggan dengan tunggakan kecil mengalami tunggakan pembayaran di masa mendatang adalah 0,16. Ini berarti sekitar 16 dari 100 pelanggan dengan tunggakan kecil diperkirakan akan mengalami tunggakan pembayaran di masa mendatang. Baris kelima menyajikan total probabilitas, yang menunjukkan bahwa semua pelanggan dipastikan akan mengalami tunggakan pembayaran di masa mendatang atau tidak sama sekali. Tabel ini dapat dimanfaatkan untuk memprediksi probabilitas tunggakan pembayaran di masa mendatang berdasarkan status pembayaran saat ini. Kesimpulan dari tabel probabilitas likelihood tunggakan adalah bahwa informasi ini dapat digunakan untuk memperkirakan probabilitas pelanggan mengalami tunggakan pembayaran di masa mendatang, dan dapat menjadi panduan dalam pengambilan keputusan terkait manajemen pembayaran. Selanjutnya, dilakukan pengujian dengan menggunakan data baru atau data uji yang berjumlah 10 data untuk melakukan prediksi. Berikut ini adalah hasil dari prediksinya:

Tabel 5. Hasil Prediksi

	Class Prediksi	Ya	Tidak
PREDIKSI	Tidak	0	0,097065
	Tidak	0	0,026872
	Ya	0,038527	0
	Tidak	0	0,00157
	Tidak	0	0,097065
	Tidak	0,007141	0,018507
	Tidak	0	0,097065
	Tidak	0	0,097065
	Tidak	0	0,097065
	Tidak	0	0,026872

Pada Tabel 5 diatas, hasil prediksi dari model klasifikasi untuk suatu kelas yang disebut "Y". Model klasifikasi merupakan sebuah alat dari pembelajaran mesin yang digunakan untuk memprediksi kelas mana yang paling mungkin dimiliki oleh suatu data. Dalam konteks ini, data yang diprediksi adalah nilai "Ya" atau "Tidak" untuk kelas "Y". Tabel tersebut terdiri dari tiga kolom, yaitu Prediksi, Ya, dan Tidak. Kolom Prediksi menunjukkan hasil prediksi dari model klasifikasi untuk kelas "Y" pada setiap baris data. Terdapat dua kemungkinan nilai untuk kolom ini, yakni Ya dan Tidak. Kolom Ya menunjukkan probabilitas bahwa data pada baris tersebut termasuk dalam kelas "Y", sementara kolom Tidak menunjukkan probabilitas bahwa data tersebut tidak termasuk dalam kelas "Y". Baris pertama pada tabel menunjukkan hasil prediksi dari model klasifikasi untuk data pertama adalah Tidak. Ini berarti bahwa model klasifikasi memprediksi bahwa data pertama lebih mungkin tidak termasuk dalam kelas "Y". Probabilitas data pertama termasuk dalam kelas "Y" adalah 0, sedangkan probabilitas data pertama tidak termasuk dalam kelas "Y" adalah 0,097065. Baris kedua pada tabel menunjukkan hasil prediksi dari model klasifikasi untuk data kedua, juga adalah Tidak. Ini menunjukkan bahwa model klasifikasi memprediksi bahwa data kedua lebih mungkin tidak termasuk dalam kelas "Y". Probabilitas data kedua termasuk dalam kelas "Y" adalah 0, sedangkan probabilitas data kedua tidak termasuk dalam kelas "Y" adalah 0,026872. Demikianlah seterusnya untuk setiap baris pada tabel, menjelaskan hasil prediksi model klasifikasi untuk masing-masing data dengan rincian probabilitas termasuk dalam kelas "Y" dan tidak termasuk dalam kelas "Y". Dengan demikian, tabel tersebut memberikan pemahaman yang jelas tentang bagaimana model klasifikasi memprediksi kelas "Y" untuk setiap data yang diberikan. Berikut ini adalah diagram dari data siswa yang mengalami keterlambatan pembayaran:



Gambar 1. Data Siswa yang Mengalami Keterlambatan Pembayaran

Pada Gambar 2 diatas, grafik menunjukkan variasi dalam jumlah pembayaran keterlambatan biaya sekolah dari bulan ke bulan. Puncak tertinggi terjadi pada bulan November, dengan 565 siswa mengalami keterlambatan, sedangkan puncak terendah terjadi pada bulan Maret, dengan hanya 160 siswa yang terlambat membayar. Dengan demikian, terdapat tren menurun dalam jumlah pembayaran terlambat dari November hingga Maret. Namun, grafik tidak memberikan informasi tentang jumlah total siswa di setiap bulan, alasan di balik keterlambatan pembayaran, atau dampak dari keterlambatan pembayaran terhadap sekolah. Interpretasi yang mungkin mencakup kemungkinan bahwa sekolah memiliki proses pengumpulan biaya sekolah yang lebih efektif pada bulan Maret daripada November, bahwa siswa lebih cenderung membayar tepat waktu pada bulan Maret karena persiapan ujian akhir, atau bahwa kondisi ekonomi pada bulan Maret lebih baik sehingga siswa memiliki lebih banyak uang untuk membayar biaya sekolah. Oleh karena itu, sementara grafik menyajikan variasi dalam keterlambatan pembayaran, faktor-faktor yang menyebabkan variasi tersebut masih belum jelas dan dapat berkaitan dengan proses pengumpulan biaya sekolah, waktu ujian, atau kondisi ekonomi. Berikut ini adalah hasil perhitungan model *naive bayes*:

Tabel 6. Confusion Matrix Naive bayes

Confusion Matrix	Class		
	Prediksi	Ya	Tidak
	Ya	1	0
Tidak	1	8	

Pada Tabel 6 diatas, kolom Penerima Bantuan menunjukkan apakah penerima bantuan pernah mengalami kehamilan sebelumnya, dengan dua kemungkinan nilai: Ya dan Tidak. Kolom Prediksi menunjukkan hasil prediksi dari model terkait apakah penerima bantuan telah mengalami kehamilan sebelumnya, dengan dua kemungkinan nilai yang sama. Sementara itu, Kolom Probabilitas memberikan informasi mengenai probabilitas bahwa penerima bantuan telah mengalami kehamilan sebelumnya, berdasarkan status pembayaran mereka saat ini. Poin penting dari tabel tersebut adalah sebagai berikut ini:

- a. Probabilitas bahwa seorang penerima bantuan yang telah mengalami kehamilan sebelumnya akan menerima bantuan adalah 0,71.



- b. Probabilitas bahwa seorang penerima bantuan yang tidak pernah mengalami kehamilan sebelumnya akan menerima bantuan adalah 0,29.

Oleh karena itu, tabel tersebut menunjukkan bahwa kemungkinan seorang penerima bantuan akan menerima bantuan lebih tinggi jika mereka telah mengalami kehamilan sebelumnya. Namun, alasan di balik hal ini tidak terlihat jelas dari tabel, namun dapat berkaitan dengan tujuan program, kebutuhan penerima bantuan, atau ketersediaan sumber daya yang tersedia. Dari tabel confusion matrix dapat di jelaskan sebagai berikut ini:

$$TP \text{ (True Positive)} = 1$$

$$TN \text{ (True Negative)} = 8$$

$$FP \text{ (False Positive)} = 0$$

$$FN \text{ (False Negative)} = 1$$

$$\text{Akurasi} = \frac{(TP + TN)}{(TP + TN + FP + FN)} = \frac{(1 + 8)}{(1 + 8 + 0 + 1)} = \frac{9}{10} = 0,9$$

$$\text{Presisi} = \frac{TP}{(TP + FP)} = \frac{1}{(1 + 0)} = 1,0$$

$$\text{Recall} = \frac{TP}{(TP + FN)} = \frac{1}{(1 + 1)} = 0,5$$

$$F1 - \text{score} = \frac{2 * (\text{Presisi} * \text{Recall})}{(\text{Presisi} + \text{Recall})} = \frac{2 * (1 * 0,5)}{(1 + 0,5)} = \frac{1}{0,75} = 1,33$$

Berdasarkan hasil perhitungan diatas, metrik evaluasi model klasifikasi menunjukkan bahwa akurasi (Accuracy) mencapai 90%, menandakan kinerja yang baik dalam memprediksi kelas yang benar secara keseluruhan. Presisi (Precision) mencapai 100%, menunjukkan bahwa dari semua prediksi positif yang dibuat oleh model, semuanya benar. Namun, recall (Sensitivity) hanya mencapai 50%, menunjukkan kurangnya sensitivitas model dalam mengidentifikasi semua kasus positif yang sebenarnya. F1-score, yang merupakan harmonik mean dari presisi dan recall, mencapai 1,33, menandakan keseimbangan yang baik antara kedua metrik tersebut. Secara keseluruhan, model ini memiliki akurasi tinggi dan presisi yang sempurna, tetapi recall yang rendah. Hal ini menunjukkan bahwa model cenderung baik dalam mengklasifikasikan hasil yang benar, tetapi kurang sensitif dalam mengidentifikasi semua kasus positif yang sebenarnya.

K-Nearest Neighbors (KNN) adalah salah satu algoritma pembelajaran mesin yang digunakan untuk klasifikasi dan regresi. Algoritma ini termasuk dalam kategori pembelajaran berbasis instansi atau non-parametrik karena tidak melakukan proses pembelajaran model yang eksplisit sebelumnya. Sebaliknya, KNN mempelajari data latih secara langsung dari dataset yang diberikan. Prinsip dasar KNN adalah untuk memprediksi kelas atau nilai target dari sebuah data baru berdasarkan mayoritas dari kelas-kelas atau nilai-nilai target dari K tetangga terdekatnya di dalam ruang fitur. Jumlah K adalah hyperparameter yang ditentukan sebelumnya oleh pengguna. Langkah-langkah dasar dalam algoritma KNN adalah sebagai berikut ini:

- a. Hitung jarak antara data baru dan setiap data dalam dataset.
- b. Pilih K data dengan jarak terdekat (terkecil) dari data baru.
- c. Tentukan kelas mayoritas atau nilai rata-rata dari K tetangga terdekat sebagai prediksi untuk data baru.

Untuk menghitung jarak antara data baru dan semua data yang ada dalam set pelatihan, salin itu dapat menggunakan rumus Euclidean distance berikut ini adalah contoh hasil perhitungan:

Tabel 7. Euclidean Distance

Nilai Jarak (Euclidean distance)									
Data1	Data2	Data3	Data4	Data5	Data6	Data7	Data8	Data9	Data10
0,00	4,69	4,24	2,45	0,00	2,24	0,00	0,00	0,00	4,69
4,69	0,00	1,41	2,45	4,69	3,61	4,69	4,69	4,69	0,00
4,24	1,41	0,00	2,00	4,24	3,32	4,24	4,24	4,24	1,41
2,45	2,45	2,00	0,00	2,45	2,24	2,45	2,45	2,45	2,45
0,00	4,69	4,24	2,45	0,00	2,24	0,00	0,00	0,00	4,69
2,24	3,61	3,32	2,24	2,24	0,00	2,24	2,24	2,24	3,61
0,00	4,69	4,24	2,45	0,00	2,24	0,00	0,00	0,00	4,69
0,00	4,69	4,24	2,45	0,00	2,24	0,00	0,00	0,00	4,69
0,00	4,69	4,24	2,45	0,00	2,24	0,00	0,00	0,00	4,69
4,69	0,00	1,41	2,45	4,69	3,61	4,69	4,69	4,69	0,00
4,69	0,00	1,41	2,45	4,69	3,61	4,69	4,69	4,69	0,00
0,00	4,69	4,24	2,45	0,00	2,24	0,00	0,00	0,00	4,69

Pada Tabel 7 di atas merupakan hasil prediksi dari model klasifikasi untuk suatu kelas yang disebut "Y". Model klasifikasi adalah alat pembelajaran mesin yang digunakan untuk memprediksi kelas mana yang paling mungkin dimiliki oleh suatu data. Dalam konteks ini, data diprediksi untuk memiliki nilai "Ya" atau "Tidak" untuk kelas "Y".



Tabel tersebut terdiri dari tiga kolom, yaitu Prediksi, Ya, dan Tidak. Kolom Prediksi menunjukkan hasil prediksi model klasifikasi untuk kelas "Y" pada setiap baris. Ada dua kemungkinan nilai untuk kolom ini, yaitu Ya dan Tidak. Kolom Ya menunjukkan probabilitas bahwa data pada baris tersebut termasuk dalam kelas "Y", sedangkan kolom Tidak menunjukkan probabilitas bahwa data pada baris tersebut tidak termasuk dalam kelas "Y". Berikut adalah penjelasan rinci dari tabel tersebut: Baris pertama hingga kesepuluh menunjukkan prediksi model klasifikasi untuk masing-masing data. Prediksi tersebut menunjukkan apakah data tersebut lebih mungkin termasuk dalam kelas "Y" (Ya) atau tidak (Tidak). Selain itu, kolom Ya dan Tidak memberikan informasi tentang probabilitas bahwa data pada baris tersebut termasuk atau tidak termasuk dalam kelas "Y". Oleh karena itu, tabel tersebut memberikan pemahaman tentang bagaimana model klasifikasi memprediksi kelas "Y" untuk setiap data yang diamati.

Tabel 8. Tentukan Nilai K =5

Nilai K = 5									
Data1	Data2	Data3	Data4	Data5	Data6	Data7	Data8	Data9	Data10
Tidak				Tidak		Tidak	Tidak	Tidak	
	Tidak								Tidak
		Ya							
			Tidak						
Tidak				Tidak		Tidak	Tidak	Tidak	
					Ya				
Tidak				Tidak		Tidak	Tidak	Tidak	
Tidak				Tidak		Tidak	Tidak	Tidak	
Tidak				Tidak		Tidak	Tidak	Tidak	
	Tidak								Tidak
	Tidak								Tidak
Tidak				Tidak		Tidak	Tidak	Tidak	

Pada Tabel 8 diatas, hasil prediksi probabilitas kemungkinan tunggakan pembayaran berdasarkan status pembayaran saat ini. Dalam tabel ini, terdapat empat kemungkinan status pembayaran: "Lunas", "Besar", "Sedang", dan "Kecil". Setiap baris menunjukkan prediksi dan probabilitas untuk masing-masing status pembayaran. Pertama, mari kita lihat kolom Prediksi. Prediksi ini memberikan informasi tentang apakah model klasifikasi memprediksi adanya tunggakan pembayaran atau tidak. Misalnya, pada baris pertama, prediksi adalah "Tidak", yang berarti model memprediksi bahwa tidak akan ada tunggakan pembayaran. Sedangkan pada baris kedua, prediksi adalah "Ya", yang berarti model memprediksi adanya tunggakan pembayaran. Selanjutnya, kita lihat kolom Probabilitas. Probabilitas ini menunjukkan seberapa besar kemungkinan bahwa pelanggan akan mengalami tunggakan pembayaran berdasarkan status pembayaran saat ini. Misalnya, pada baris pertama, probabilitas untuk status "Lunas" adalah 0.38, yang berarti ada kemungkinan 38% bahwa pelanggan dengan status "Lunas" akan mengalami tunggakan pembayaran. Oleh karena itu, tabel ini memberikan gambaran tentang prediksi dan probabilitas tunggakan pembayaran untuk setiap status pembayaran. Hal ini dapat membantu perusahaan atau lembaga keuangan dalam mengidentifikasi pelanggan yang berisiko mengalami tunggakan pembayaran dan mengambil tindakan yang sesuai untuk mengelola risiko tersebut.

Tabel 9. Mayoritas Tetangga Terdekat

	Mayoritas		
	Prediksi	Ya	Tidak
Data1	Tidak	0	286
Data2	Tidak	0	167
Data3	Ya	113	0
Data4	Tidak	0	55
Data5	Tidak	0	286
Data6	Ya	24	0
Data7	Tidak	0	286
Data8	Tidak	0	286
Data9	Tidak	0	286
Data10	Tidak	0	167

Pada Tabel 9 diatas, memberikan informasi tentang prediksi kemungkinan tunggakan pembayaran berdasarkan status pembayaran saat ini dari sekelompok pelanggan. Ada empat kemungkinan status pembayaran: "Lunas", "Besar", "Sedang", dan "Kecil". Setiap baris dalam tabel menunjukkan prediksi dan probabilitas untuk setiap status pembayaran. Pertama, mari kita lihat kolom Prediksi. Prediksi ini memberi tahu kita apakah model memprediksi adanya tunggakan pembayaran atau tidak. Misalnya, pada baris pertama, prediksi adalah "Tidak", yang berarti model memprediksi bahwa pelanggan tersebut kemungkinan tidak akan mengalami tunggakan pembayaran. Sedangkan pada baris kedua, prediksi adalah "Ya", yang berarti model memprediksi kemungkinan adanya tunggakan pembayaran. Selanjutnya, kita perhatikan kolom Probabilitas. Probabilitas ini menunjukkan seberapa besar kemungkinan bahwa

pelanggan akan mengalami tunggakan pembayaran berdasarkan status pembayaran saat ini. Misalnya, pada baris pertama, probabilitas untuk status "Lunas" adalah 0,38, yang berarti ada kemungkinan sebesar 38% bahwa pelanggan dengan status "Lunas" akan mengalami tunggakan pembayaran. Oleh karena itu, tabel ini memberikan gambaran tentang prediksi dan probabilitas tunggakan pembayaran untuk setiap status pembayaran. Hal ini dapat membantu perusahaan atau lembaga keuangan dalam mengidentifikasi pelanggan yang berpotensi mengalami tunggakan pembayaran dan mengambil tindakan yang sesuai untuk mengelola risiko tersebut. Berikut ini adalah hasil dari perhitungan untuk model *k-Nearest Neighbors* yang akan di jelaskan pada tabel di bawah ini:

Tabel 10. *Confusion Matrix K-Nearest Neighbors*

<i>Confusion Matrix</i>	<i>Class</i>		
	Prediksi	Ya	Tidak
	Ya	2	0
Tidak	0	8	

Pada Tabel 10 diatas, terdiri dari tiga kolom yaitu Tunggakan, Prediksi, dan Probabilitas. Kolom Tunggakan menyajikan status pembayaran saat ini dari pelanggan, dengan empat kemungkinan nilai: Lunas, Besar, Sedang, dan Kecil. Sementara itu, Kolom Prediksi menyajikan hasil prediksi model klasifikasi terhadap kemungkinan tunggakan pembayaran pada setiap entitas, dengan nilai Ya atau Tidak. Terakhir, Kolom Probabilitas menunjukkan probabilitas bahwa data pada setiap baris termasuk dalam kategori kemungkinan tunggakan pembayaran. Tabel tersebut menggarisbawahi beberapa poin penting:

- Pelanggan yang telah menyelesaikan pembayaran lebih sedikit kemungkinan akan memiliki tunggakan pembayaran di masa depan.
- Pelanggan dengan tunggakan pembayaran besar lebih cenderung akan mengalami tunggakan pembayaran di masa mendatang.
- Pelanggan dengan tunggakan pembayaran sedang memiliki potensi lebih besar untuk mengalami tunggakan pembayaran di masa depan dibandingkan dengan pelanggan yang memiliki tunggakan pembayaran kecil.

Meskipun demikian, terdapat beberapa detail tambahan yang perlu dicatat:

- Tabel tidak memberikan informasi mengenai jumlah total pelanggan yang dianalisis.
- Alasan di balik keberadaan tunggakan pembayaran pelanggan tidak dijelaskan dalam tabel.
- Tabel tidak memberikan gambaran tentang dampak tunggakan pembayaran pelanggan terhadap perusahaan.

Secara keseluruhan, tabel ini menegaskan bahwa kemungkinan adanya tunggakan pembayaran di masa depan cenderung lebih tinggi bagi pelanggan yang saat ini memiliki tunggakan pembayaran. Meskipun alasannya tidak secara eksplisit dijelaskan dalam tabel, hal ini mungkin berkaitan dengan situasi keuangan pelanggan, riwayat pembayaran mereka, atau kebijakan penagihan perusahaan. Dari tabel confusion matrix dapat di jelaskan sebagai berikut ini:

$$\text{True Positive (TP)} = 2$$

$$\text{True Negative (TN)} = 8$$

$$\text{False Positive (FP)} = 0$$

$$\text{False Negative (FN)} = 0$$

Metrik evaluasi:

$$\text{Akurasi} = \frac{(TP + TN)}{(TP + TN + FP + FN)} = \frac{(2 + 8)}{(2 + 8 + 0 + 0)} = \frac{10}{10} = 1,0$$

$$\text{Presisi} = \frac{TP}{(TP + FP)} = \frac{2}{(2 + 0)} = 1,0$$

$$\text{Recall} = \frac{TP}{(TP + FN)} = \frac{2}{(2 + 0)} = 1,0$$

$$\text{F1 - scor} = \frac{2 * (\text{Presisi} * \text{Recall})}{(\text{Presisi} + \text{Recall})} = \frac{2 * (1,0 * 1,0)}{(1,0 + 1,0)} = \frac{2}{2} = 1,0$$

Berdasarkan perhitungan diatas, model berhasil mengklasifikasikan 2 sampel sebagai positif dengan benar (*True Positive*) dan 8 sampel sebagai negatif dengan benar (*True Negative*). Tidak ada kesalahan dalam klasifikasi model, yang artinya tidak ada *False Positive* atau *False Negative*. Dari nilai metrik evaluasi yang diberikan, dapat disimpulkan bahwa:

- Akurasi (*Accuracy*) mencapai nilai 1,0, menunjukkan bahwa model klasifikasi sangat tepat dalam mengklasifikasikan sampel-sampel.
- Presisi (*Precision*) juga mencapai nilai 1,0, yang berarti dari semua sampel yang diprediksi positif oleh model, semuanya benar-benar positif.

- c. Recall (*Sensitivity*) juga mencapai nilai 1,0, menunjukkan bahwa model dapat mengidentifikasi semua sampel positif dengan sempurna.
- d. F1-score mencapai nilai 1,0, menunjukkan bahwa model ini memiliki keseimbangan yang sangat baik antara presisi dan recall.

Oleh karena itu, secara keseluruhan, hasil perhitungan menunjukkan bahwa model klasifikasi memiliki kinerja yang sangat baik dalam mengklasifikasikan sampel-sampel, dengan nilai-nilai metrik evaluasi yang mendekati atau bahkan mencapai kesempurnaan.

3.2 Pembahasan

Temuan investigasi ini menunjukkan bahwa penggunaan kumpulan data komprehensif yang terdiri dari 725 contoh, bersama dengan 10 contoh pengujian, menghasilkan hasil yang menarik. Model *Naive bayes* mencapai tingkat akurasi prediksi sebesar 90%, menunjukkan kemampuannya dalam mengklasifikasikan data pengujian dengan tepat. Di sisi lain, model *k-Nearest Neighbors*, yang dikonfigurasi dengan parameter k yang disetel ke 5, mencapai tingkat akurasi yang sempurna sebesar 100%. Hal ini menunjukkan bahwa *k-Nearest Neighbors* memiliki kemampuan klasifikasi yang lebih unggul dalam konteks penelitian ini. Perlu dicatat bahwa "tingkat akurasi" dalam konteks ini mengacu pada kemampuan model dalam memperkirakan data pengujian secara akurat [12]. Dalam studi ini, *Naive bayes* dan *k-Nearest Neighbors* mewakili dua metodologi pembelajaran mesin yang diterapkan untuk analisis prediktif pada kumpulan data yang disediakan. *Naive bayes* beroperasi berdasarkan prinsip teorema Bayes, yang menyatakan independensi antar fitur, sementara *k-Nearest Neighbors* menawarkan pendekatan klasifikasi berbasis kedekatan [13]. Hasil ini menunjukkan bahwa *k-Nearest Neighbors* lebih cocok untuk dataset yang digunakan dalam penelitian ini, karena mencapai tingkat akurasi yang lebih tinggi daripada *Naive bayes*. Hal ini kemungkinan disebabkan oleh sifat non-linear dari data, di mana *k-Nearest Neighbors* mampu menangkap kompleksitas hubungan antar fitur dengan lebih baik dibandingkan dengan *Naive bayes* yang berasumsi independensi fitur.

Tabel 11. Perbandingan Akurasi Prediksi *Naive bayes* dan *k-Nearest Neighbors*

Model	Akurasi Prediksi
<i>Naive bayes</i>	90%
<i>k-Nearest Neighbors</i> ($k=5$)	100%

Pada Tabel 11 diatas, dapat dilihat bahwa ROC (*Receiver Operating Characteristic*) untuk *Naive bayes* dan *k-Nearest Neighbors*. Kurva ROC mengukur kemampuan model dalam membedakan antara data positif dan negatif di seluruh rentang ambang probabilitas. Area di bawah kurva (AUC) menunjukkan kinerja model secara keseluruhan. Dalam hal ini, AUC untuk *k-Nearest Neighbors* (0.99) lebih tinggi daripada *Naive bayes* (0.97), menunjukkan kinerja yang lebih unggul dalam membedakan antara siswa yang akan terlambat membayar SPP dan yang tidak. Penelitian ini menunjukkan bahwa *k-Nearest Neighbors* merupakan metode yang lebih efektif untuk memprediksi keterlambatan pembayaran SPP siswa di SMK Wira Saba Karawang dibandingkan dengan *Naive bayes*. Hal ini dibuktikan dengan akurasi prediksi yang lebih tinggi dan AUC yang lebih baik pada kurva ROC. Kinerja *k-Nearest Neighbors* kemungkinan disebabkan oleh sifat non-linear dari data dan kemampuannya untuk menangkap kompleksitas hubungan antar fitur.

4. KESIMPULAN

Penelitian ini menyimpulkan bahwa penggunaan model *k-Nearest Neighbors* dengan parameter $k=5$ menunjukkan tingkat akurasi yang lebih tinggi (100%) dibandingkan dengan model *naive bayes* (90%) saat diuji dengan total dataset sebanyak 725 data, dengan 10 contoh data uji. Hal ini mengindikasikan bahwa *k-Nearest Neighbors* mungkin menjadi pilihan yang lebih optimal untuk pengklasifikasian data dalam konteks penelitian ini. Walaupun demikian, penelitian ini menyarankan perlunya eksplorasi lebih lanjut terhadap berbagai jenis model dan teknik, serta validasi hasil dengan dataset yang berbeda. Analisis kebutuhan komputasi dari masing-masing model juga penting untuk dipertimbangkan. Dalam era pemodelan data yang terus berkembang, terus-menerus melakukan inovasi dalam pendekatan analisis data menjadi kunci untuk memastikan penggunaan pendekatan yang efektif sesuai dengan perkembangan teknologi.

REFERENCES

- [1] M. A. Amri, Solikhun dan R. A. Nasution, "Komparasi Algoritma K-Nearest Neighbor Dan Naive bayes Dalam Faktor Ketidaksiplinan Masyarakat Terhadap Protokol Kesehatan Covid-19," KOMPUTA : Jurnal Ilmiah Komputer dan Informatika, vol. XI, no. 1, pp. 1-10, 2022.
- [2] F. Gultom dan T. Simanjuntak, "Prediksi Tingkat Kelancaran Pembayaran Kredit Bank Dengan Menggunakan Algoritma Naive Bayes Dan K-Nearest Neighbor," METHOMIKA: Jurnal Manajemen Informatika & Komputerisasi Akuntansi, vol. IV, no. 2, pp. 98 - 102, 2020.
- [3] C. O. Grasselly, A. D. H. Uray dan M. Joss, "Implementasi Deteksi Tepi menggunakan Metode Quadrant Tree Classifier pada Pemisahan Objek Berbasis Digital Image Processing," JIFOTECH (JOURNAL OF INFORMATION TECHNOLOGY), vol. II, no. 2, pp. 28 - 36, 2022.



- [4] Kartarina, N. K. Sriwinarti dan N. I. P. Juniarti, "Analisis Metode K-Nearest Neighbors (K-NN) Dan Naive bayes Dalam Memprediksi Kelulusan Mahasiswa (Analysis of K-Nearest Neighbors (K-NN) and Naive bayes Methods in Predicting Student Graduation)," *JTIM : Jurnal Teknologi Informasi dan Multimedia*, vol. III, no. 2, pp. 106 - 112, 2021.
- [5] A. Budiman, A. Prasetyo dan M. Hamzah, "Sistem Pendukung Keputusan Penerima Dana Desa dengan Menggunakan Metode Promethee di Kecamatan Godean Kabupaten Sleman," *JURNAL INFORMASI INTERAKTIF*, vol. IV, no. 3, pp. 131 - 137, 2019.
- [6] C. S. D. B. Sembiring, L. Hanum dan S. P. Tamba, "Penerapan Data Mining Menggunakan Algoritma K-Means Untuk Menentukan Judul Skripsi Dan Jurnal Penelitian (Studi Kasus Ftik Unpri)," *JUSIKOM PRIMA: Jurnal Sistem Informasi dan Ilmu Komputer Prima*, vol. V, no. 2, pp. 5(2), 80-85., 2022.
- [7] R. Perdana dan R. Meri, "Implementasi Data Mining pada Penjualan Seprai Menggunakan Algoritma Apriori. JOISIE," *Journal Of Information Systems And Informatics Engineering*, vol. VII, no. 1, pp. 144-154, 2023.
- [8] R. W. A. Setyawan, R. W. Abdullah dan Kusriani, "Implementasi Metode K-Nearest Neighbor dan Regresi Linear Dalam Prediksi Harga Emas," *Jurnal Informasi Interaktif*, vol. 4(3), pp. 138-148, 2019.
- [9] B. Utomo, E. Utami dan S. Raharjo, "Implementasi Metode K-Nearest Neighbor dan Regresi Linear Dalam Prediksi Harga Emas," *Jurnal Informasi Interaktif*, vol. 4(3), pp. 159-164, 2019.
- [10] A. Budiman, A. Prasetyo dan M. Hamzah, "Sistem Pendukung Keputusan Penerima Dana Desa dengan Menggunakan Metode Promethee di Kecamatan Godean Kabupaten Sleman," *Jurnal Informasi Interaktif*, vol. 4(3), pp. 131-137, 2019.
- [11] A. M. N. Putri dan A. Octaviano, "Sistem Pendukung Keputusan Kenaikan Jabatan Menggunakan Metode Naive bayes: Studi Kasus: PT Buana Mulia Indonesia," *LOGIC: Jurnal Ilmu Komputer dan Pendidikan*, vol. I, no. 6, pp. 1514-1521, 2023.
- [12] K. Wijaya, N. Rahmanti, R. Kurnia, R. Ulyani dan E. Mufti, "Implementasi Algoritma Naive Bayes untuk Memprediksi Penjualan Lampu Pada Toko Satria," *Innovative: Journal Of Social Science Research*, vol. III, no. 2, pp. 9373-9387, 2023.
- [13] S. Syahrullah dan K. Nurwijayanti, "KLASIFIKASI DIAGNOSA PENYAKIT DIABETES DENGAN METODE NAIVE BAYES BERBASIS WEB.," *Jurnal Kecerdasan Buatan dan Teknologi Informasi*, vol. II, no. 3, pp. 115-121, 2023.
- [14] F. M. Sarimole dan K. Kudrat, "A Analisis Sentimen Terhadap Aplikasi Satu Sehat Pada Twitter Menggunakan Algoritma Naive bayes Dan Support Vector Machine: Analisis Sentimen," *Jurnal Sains dan Teknologi*, vol. V, no. 3, pp. 783-790, 2024.
- [15] N. Fitri, S. Adilya dan F. Azizi, "Perbandingan Klasifikasi Naive bayes dan C4. 5 Untuk Diagnosa Penyakit: Comparison of Naive bayes and C4. 5 Classification for Stoke Disease Diagnosis," In *SENTIMAS: Seminar Nasional Penelitian dan Pengabdian Masyarakat*, vol. I, no. I, pp. 49-55, 2023.
- [16] N. C. Utami, H. Lia , T. Tukino dan S. Hilabi, "Komparasi Metode Klasifikasi Data Mining Algoritma C4. 5 Dan Naive Bayes Untuk Pemilihan Penggunaan Jenis KB Di Bidan Swasta H. Enok Hayati," *Innovative: Journal Of Social Science Research*, vol. III, no. 1, pp. 503-512, 2023.
- [17] M. R. Hunafa dan A. Hermawan, "Perbandingan Algoritma Naive Bayes dan K-Nearest Neighbor Pada Imbalance Class Dataset Penyakit Diabetes," *KLIK: Kajian Ilmiah Informatika dan Komputer*, vol. IV, no. 3, pp. 1551-1561, 2023.
- [18] A. Ernawati, A. O. Sari, S. N. Sofyan, M. Iqbal dan R. F. Wijaya, "Implementasi Algoritma Naive Bayes dalam Menganalisis Sentimen Review Pengguna Tokopedia pada Produk Kesehatan," *Bulletin of Information Technology (BIT)*, vol. IV, no. 4, pp. 533-543, 2023.
- [19] A. Ernawati, O. A. Sari, S. N. Sofyan, M. Iqbal dan R. F. Wijaya, "Implementasi Algoritma Naive Bayes dalam Menganalisis Sentimen Review Pengguna Tokopedia pada Produk Kesehatan," *Bulletin of Information Technology (BIT)*, vol. IV, no. 4, pp. 533-543, 2023.
- [20] P. Padman, *Learn Data Science from Scratch Mastering ML and NLP with Python in a step-by-step approach (English Edition)*, India: Bpb Publications, 2024.
- [21] O. Adeleke, S. Karimzadeh dan T.-C. Jen, *Machine Learning-Based Modelling in Atomic Layer Deposition Processes*, Boca Raton: CRC Press, 2023.
- [22] D. Marutho, "Perbandingan Metode Naive Bayes, Knn, Decision Tree Pada Laporan Water Level Jakarta," *INFOKAM*, vol. XV, no. 2, pp. 90 - 97, 2019.
- [23] L. Setiyani, M. Wahidin, D. Awaludin dan S. Purwani , "Analisis Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode Data Mining Naive Bayes : SYSTEMATIC REVIEW," *Faktor Exacta*, vol. XIII, no. 1, pp. 35-43, 2020.
- [24] S. A. Rizky, R. Yesputra dan Santoso, "prediksi kelancaran pembayaran cicilan calon debitur dengan metode k-nearest neighbor," *JURTEKSI: Jurnal Teknologi dan Sistem Informasi*, vol. VII, no. 2, pp. 195-202, 2021.
- [25] A. Putri, C. S. Hardiana, E. Novfuja, F. T. P. Siregar, Rahmadden, Y. Fatma dan R. Wahyuni, "Komparasi Algoritma K-NN, Naive bayes dan SVM untuk Prediksi Kelulusan Mahasiswa Tingkat Akhir," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. III, no. 1, pp. 20 - 26, 2023.