



Prediksi Kelulusan Siswa Menggunakan Algoritma K- Nearest Neighbors (K-NN) Di SMK Al Hasyimy Ibrahimy

Basmalia¹, Zaehol Fatah²

¹ Teknologi Informasi, Sains dan Teknologi, Universitas Ibrahimy

² Sistem Informasi, Universitas Ibrahimy

liabasma654@email.com, zaeholfatah@email.com

Abstrak

Model prediksi kelulusan siswa di SMK Al Hasyimy Ibrahimy dikembangkan dengan algoritma K-Nearest Neighbor (K-NN) yang diimplementasikan melalui aplikasi RapidMiner. Model ini memproses variabel seperti nilai Sumatif, UTS, PAS, UKK, dan PSAJ untuk mengklasifikasikan status kelulusan siswa sebagai "tepat waktu" atau "terlambat". Algoritma K-NN bekerja dengan menghitung kedekatan (Euclidean Distance) antara data baru dan data yang sudah memiliki hasil kelulusan. Hasil prediksi menunjukkan akurasi sebesar 95,51% memberikan manfaat bagi sekolah dalam mengenali siswa yang memerlukan dukungan tambahan agar lulus tepat waktu. Penggunaan pendekatan ini mendukung penerapan strategi pendidikan berbasis data untuk meningkatkan tingkat kelulusan siswa.

Kata Kunci: Prediksi kelulusan, K-Nearest Neighbor, Pendidikan

PENDAHULUAN

Pendidikan merupakan bagian esensial dari suatu struktur masyarakat yang membantu mereproduksi bentuk masyarakat yang dikehendaki oleh suatu bangsa. Kebijakan pendidikan menuntut upaya pembaharuan yang terus-menerus tentang fungsi dan tujuan pendidikan nasional, serta pelaksanaan, pemantauan, dan pengendaliannya. (Fajjin Amik, S.Pd., M.Si.; Hj. Een Nuraini, S.Pd.; Agis Sugiarti, 2016)

Salah satu faktor penting dalam membangun kualitas pendidikan adalah kualitas tenaga pendidik dalam merancang dan melaksanakan pembelajaran. Guru seharusnya memiliki keterampilan yang memadai untuk mendesain, mengembangkan, dan memanfaatkan media pembelajaran dalam upaya meningkatkan minat, perhatian, dan motivasi belajar peserta didik. Dengan meningkatnya motivasi dan minat belajar diharapkan dapat mencerna dan menerima pembelajaran dengan mudah. (Prof. Dr. Muhammad Yaumi, M.Hum., 2018)

Agar dapat menghasilkan sumber daya manusia (SDM) yang terampil, berpengetahuan, kompetitif, dan kreatif, lembaga pendidikan dituntut untuk menyediakan pendidikan yang berkualitas bagi para pesertanya. Untuk mewujudkannya, diadakan ujian sekolah maupun ujian nasional untuk menentukan kelulusan di setiap tingkat pendidikan serta sebagai tolok ukur dalam satuan pendidikan (Prof. Dr. Prayitno, M.Sc., 2021)

Kelulusan siswa ditentukan oleh kemampuan dalam menyelesaikan dan memenuhi persyaratan yang telah ditetapkan dalam rapat kelulusan dan disahkan oleh Kepala Sekolah melalui surat keputusan. Tingkat kelulusan yang dicapai digunakan untuk merumuskan strategi atau kebijakan baru agar hasil kelulusan tahun berikutnya dapat meningkat. Salah satu indikator kesiapan siswa adalah nilai rapor per semester dari beberapa mata pelajaran. Berdasarkan nilai-nilai ini, siswa dan guru dapat mengidentifikasi mata pelajaran yang perlu ditingkatkan, mendorong siswa untuk belajar lebih giat.

Bagi pihak sekolah, khususnya para guru, tingkat kelulusan siswa menjadi prioritas utama karena berkaitan dengan akreditasi sekolah. Oleh karena itu, diperlukan strategi yang tepat untuk meningkatkan angka kelulusan siswa. Untuk memprediksi tingkat kelulusan siswa setiap tahun ajaran, dapat dilakukan analisis prediksi atau klasifikasi. Dalam klasifikasi ini terdapat beragam pilihan algoritma, salah satunya adalah Algoritma K-Nearest Neighbor (K-NN). Dengan cara ini, para guru dapat lebih fokus dalam meningkatkan proses belajar mengajar, terutama bagi siswa yang diprediksikan tidak lulus. (Prianto, 2020)

Algoritma K-Nearest Neighbor atau yang biasa disebut KNN merupakan metode pengklasifikasian data yang bekerja relatif dengan cara yang lebih sederhana dibandingkan dengan metode pengklasifikasian data lainnya. Algoritma ini berusaha mengklasifikasikan data baru yang belum diketahui class-nya dengan memilih data sejumlah k yang letaknya terdekat dari data baru tersebut. Class terbanyak dari data terdekat sejumlah k tersebut dipilih sebagai class yang diprediksikan untuk data yang baru. k umumnya ditentukan dalam jumlah ganjil untuk menghindari munculnya jumlah jarak yang sama dalam proses pengklasifikasian. (RAYMANA APRILIAN RONI HABIBI, S.T., M.Kom M. YUSRIL HELMI SETYAWAN, S.KOM., 2019)

Berdasarkan Hasil Penelitian sebelumnya keberhasilan pada IMPLEMENTASI ALGORITMA K-NEAREST NEIGHBORS DALAM MENENTUKAN KATEGORI PRESTASI SISWA BERDASARKAN NILAI AKADEMIK

memperoleh hasil untuk semester ganjil, akurasi cross-validation adalah 76.6%, precision, recall, dan f1-score masing-masing 77.6%. Sedangkan untuk semester genap, akurasi cross-validation adalah 75.2%, dengan precision, recall, dan f1-score masing-masing 75.2%. (Fadillah & Pradana, 2024) hasil penelitian lain PENERAPAN K-NEAREST NEIGHBOR UNTUK KLASIFIKASI TINGKAT KELULUSAN PADA SISWA memperoleh Hasil dari prediksi tingkat kelulusan siswa dengan metode KNN didapat rata-rata akurasi dengan nilai sebesar 96,49%. Pengolahan data dilakukan dengan menggunakan tools rapid miner. (Purwaningsih & Nurelasari, 2021)

SMK Al Hasyimy Ibrahimy adalah salah satu Sekolah Menengah Kejuruan yang berdiri sejak tahun 2015. Dalam konteks ini, sering kali terjadi kesalahan dalam pengolahan tingkat kelulusan siswa, di mana kesalahan dalam proses penilaian siswa dapat menyulitkan pihak sekolah dalam mengelola data kelulusan, sehingga menghasilkan informasi yang tidak akurat. Untuk mengatasi permasalahan tersebut digunakan metode K-Nearest Neighbor (K-NN) dengan memanfaatkan data yang sudah ada. Dengan demikian, pengolahan data tingkat kelulusan siswa di SMK Al Hasyimy Ibrahimy dapat dilakukan untuk memudahkan pihak sekolah dalam memprediksi tingkat kelulusan, baik untuk siswa yang lulus maupun yang tidak lulus.

Berdasarkan permasalahan di atas, penulis terdorong untuk mengambil judul “Prediksi Kelulusan Siswa Di SMK Al Hasyimy Ibrahimy Dengan Metode K-NN” Hal ini bertujuan untuk membantu pihak sekolah dalam memahami pola penerimaan dengan memanfaatkan nilai atau data siswa itu sendiri, sehingga dapat membantu pihak sekolah dalam menyusun strategi yang tepat untuk meningkatkan kualitas pendidikan.

Dengan aplikasi RapidMiner Algoritma K-NN dapat mengevaluasi model prediksi kelulusan siswa, serta mengidentifikasi faktor-faktor yang memengaruhi kinerja model tersebut. Diharapkan hasilnya dapat berkontribusi pada pengembangan sistem prediksi kelulusan siswa yang lebih efisien dan efektif. Selain itu, tujuan lainnya adalah memberikan wawasan tentang penerapan praktis Algoritma K-Nearest Neighbor (K-NN) dalam konteks prediksi kelulusan siswa.

METODE

Metode Pengambilan Data

Bagian ini menguraikan metode yang diterapkan untuk mengumpulkan data dalam penelitian yang sedang berlangsung. Peneliti mengumpulkan data melalui studi literatur menggunakan data testing dan data training berupa data siswa yang diambil dari data Sekolah SMK Al Hasyimy Ibrahimy. Data ini kemudian disaring dan akan diproses untuk memprediksi kelulusan siswa pada tahun ini dengan bantuan aplikasi RapidMiner. Beberapa atribut yang dianalisis mencakup Nilai Sumatif, UTS, PAS, UKK, dan PSAJ. Dataset ini disajikan dalam gambar berikut:

Tabel 1. Dataset Training

NAMA	JENIS KELAMIN	SUMATI F	UTS	PAS	UKK	PSAJ	RATA- RATA	STATUS KELULUSAN
ANIK WIDAYANTI	PEREMPUAN	2,76	2,8	3,2	3,17	2,98	2,98	TERLAMBAT
DWI HESTYNA	PEREMPUAN	3	3,3	3,14	3,14	2,84	3,08	TERLAMBAT
MURYA ARIEF BASUKI	PEREMPUAN	3,5	3,3	3,7	3,29	3,53	3,46	TERLAMBAT
NANIK SUSANTI	PEREMPUAN	3,17	3,41	3,61	3,36	3,48	3,41	TERLAMBAT
RIFKA ISTIQFARINA	PEREMPUAN	2,9	2,89	3,3	2,85	2,98	2,98	TERLAMBAT
SUHARYONO	LAKI - LAKI	2,95	2,82	3,09	3,1	2,78	2,95	TERLAMBAT
FARIKHATUN NAZLI	PEREMPUAN	2,76	3,14	2,6	2,95	3,23	2,94	TEPAT
FIFI SUNALISA	PEREMPUAN	2,62	2,89	2,32	2,5	2,5	2,57	TEPAT

Tabel 2. Dataset Testing

NAMA	JENIS KELAMIN	SUMATIF	UTS	PAS	UKK	PSAJ	RATA- RATA	STATUS KELULUSAN
UNAMA	LAKI - LAKI	3,17	2,7	3,23	2,41	3	2,902	
LEYLA TRIYANA PRATIWI	PEREMPUAN	3,6	3,5	3,42	2,85	3,31	3,336	
VERIS SOFIYAN PRAYOGA	LAKI - LAKI	2,67	2,66	2,93	3,14	2,92	2,864	
ADITYA AKBAR NUGRAHA	LAKI - LAKI	2,48	2,86	2,09	2,55	2,55	2,506	
ERNA EKA RIYANTI	PEREMPUAN	3,19	3,08	3,31	2,83	3,36	3,154	
FARID DWI NORYANTO	LAKI - LAKI	3,1	2,98	3,17	3,25	3,41	3,182	
DAFIK HADI WINOTO	LAKI - LAKI	2,98	2,68	2,23	2,86	2,25	2,6	
WAHYU FITRIYANTO	LAKI - LAKI	3,45	3,15	3,54	3,78	3,42	3,468	

Klasifikasi

Klasifikasi adalah proses mengevaluasi data untuk menemukannya ke dalam salah satu kelas yang ada. Proses ini melibatkan pembuatan model berdasarkan data latih, yang kemudian digunakan untuk mengklasifikasikan data baru. Klasifikasi bisa diartikan sebagai pembelajaran atau pelatihan terhadap fungsi target yang memetakan setiap kumpulan atribut (fitur) ke salah satu label kelas yang tersedia. Sistem klasifikasi diharapkan mampu mengklasifikasikan semua data dengan akurat, meskipun tidak selalu 100% benar. Oleh karena itu, kinerja sistem klasifikasi perlu dievaluasi, dan biasanya dilakukan menggunakan matriks konfusi. (Andarsyah, 2023)

Prediksi

Prediksi mirip dengan klasifikasi dan estimasi, namun bedanya adalah hasil dari prediksi akan terjadi di masa depan. Contoh penerapan prediksi dalam bisnis dan penelitian antara lain: memprediksi harga beras dalam tiga bulan mendatang dan memperkirakan persentase kenaikan kecelakaan lalu lintas pada tahun depan jika batas kecepatan minimum dinaikkan. Beberapa metode dan teknik yang dipakai dalam klasifikasi dan estimasi juga bisa diterapkan untuk prediksi, asalkan dalam kondisi yang sesuai. (Kusrini, 2023)

Data Mining

Data mining didefinisikan sebagai satu set teknik yang digunakan secara otomatis untuk mengeksplorasi secara menyeluruh dan membawa ke permukaan relasi-relasi yang kompleks pada set data yang sangat besar. Set data yang dimaksud di sini adalah set data yang berbentuk tabulasi, seperti yang banyak diimplementasikan dalam teknologi manajemen basis data relasional. Akan tetapi, teknik-teknik data mining dapat juga diaplikasikan pada representasi data yang lain, seperti domain data spasial, berbasis teks, dan multimedia (citra). (Adam Puspabuana, S. Kom., M. Kom., Amril Mutoi siregar, S. Kom., M., 2020)

K-Nearest Neighbors

K-Nearest Neighbor (KNN) adalah salah satu algoritma dalam machine learning yang digunakan untuk klasifikasi dan regresi. (Ahmad Yogiarto, Ahmad Homaidi, 2024) Dengan kata lain, KNN melakukan prediksi berdasarkan data baru yang didekatkan dengan tetangga terdekatnya. Metode Nearest Neighbor adalah pendekatan yang digunakan untuk menemukan kasus dengan cara menghitung kedekatan antara data baru (data pengujian) dan data lama (data pelatihan). Algoritma KNN adalah metode klasifikasi yang dapat mengelompokkan data baru berdasarkan jarak antara data baru tersebut dan tetangga terdekat dalam data pelatihan. (Esi Vidia Rahemadani, Syafriah Fachri Pane, 2020)

Kesamaan data dengan label yang dipakai adalah menggunakan rumus Euclidean distance sebagai berikut:

$$Euc = d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

Keterangan:

x_i = data latih (data training)

y_i = data uji (data testing)

i = variabel data

$euc = d(x, y)$ = jarak

n = jumlah data

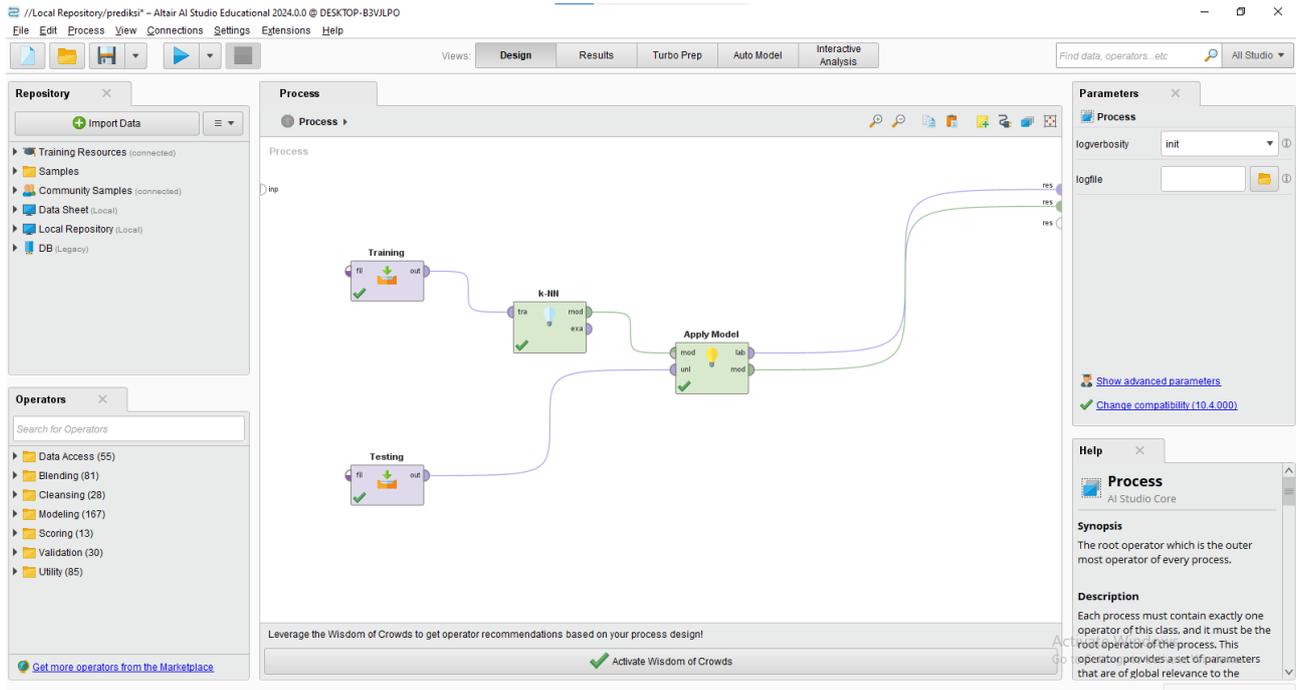
Euclidean Distance sering digunakan untuk menghitung jarak. Euclidean Distance berfungsi menguji ukuran yang bisa digunakan sebagai interpretasi kedekatan jarak antara dua objek.

HASIL DAN PEMBAHASAN

Model klasifikasi yang diterapkan adalah K-Nearest Neighbors (K-NN), menggunakan algoritma Machine Learning RapidMiner versi 10.4. Dataset yang digunakan di ambil dari SMK Al Hasyimy Ibrahimy. Dataset ini mencakup 7 atribut, termasuk klasifikasi apakah siswa lulus tepat waktu atau tidak.

Penerapan Data Mining Menggunakan RapidMiner

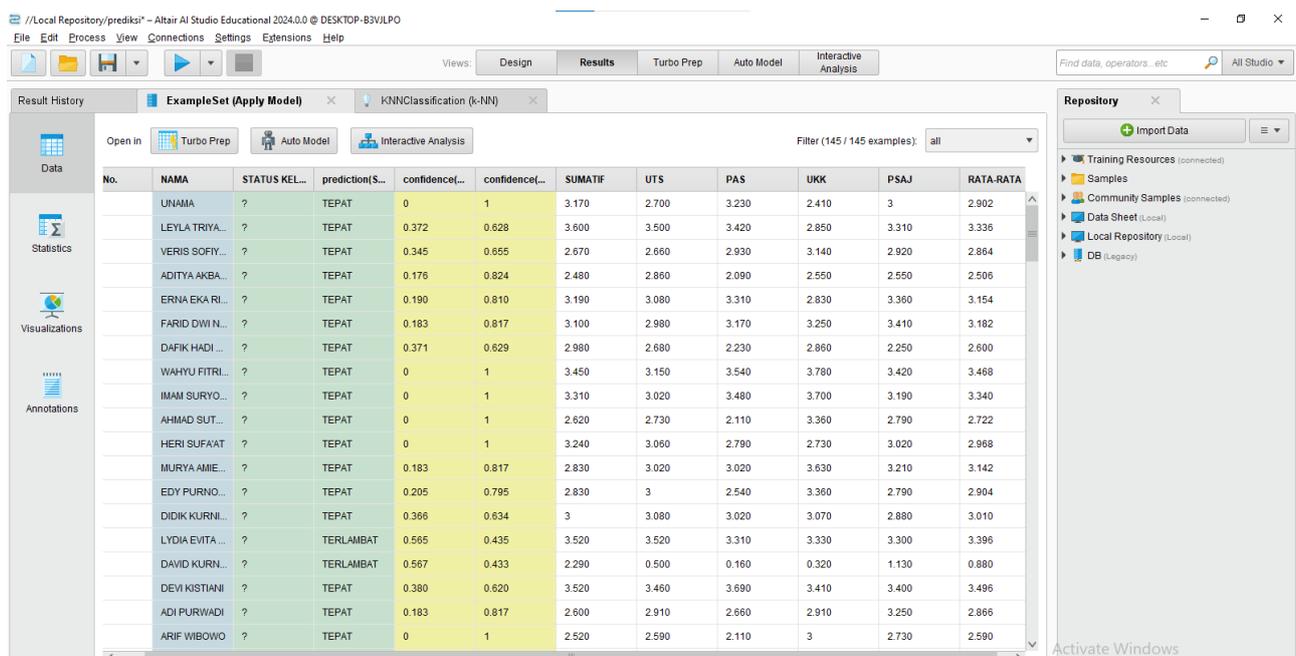
Pada Gambar 1 menjelaskan tentang penginputan proses data melalui Rapid Miner



Gambar 1 Proses penghitungan Tools Rapidminer dengan Metode K-NN

Penerapan data mining menggunakan algoritma K-Nearest Neighbors (K-NN) dilakukan dengan aplikasi RapidMiner versi 10.4 sebagaimana ditunjukkan pada Gambar 1. Langkah selanjutnya adalah menghubungkan semua operator yang telah disiapkan. Operator Apply Model dalam RapidMiner adalah salah satu operator utama yang digunakan untuk menerapkan model yang telah dilatih ke dataset baru, sehingga memungkinkan prediksi atau klasifikasi pada data yang belum dikenal berdasarkan model yang dibuat sebelumnya melalui operator pelatihan. Setelah melatih model K-NN dengan data pelatihan menggunakan operator K-NN, dapat menggunakan operator Apply Model untuk menerapkannya pada data uji dan memperoleh prediksi terkait Kelulusan siswa pada dataset tersebut.

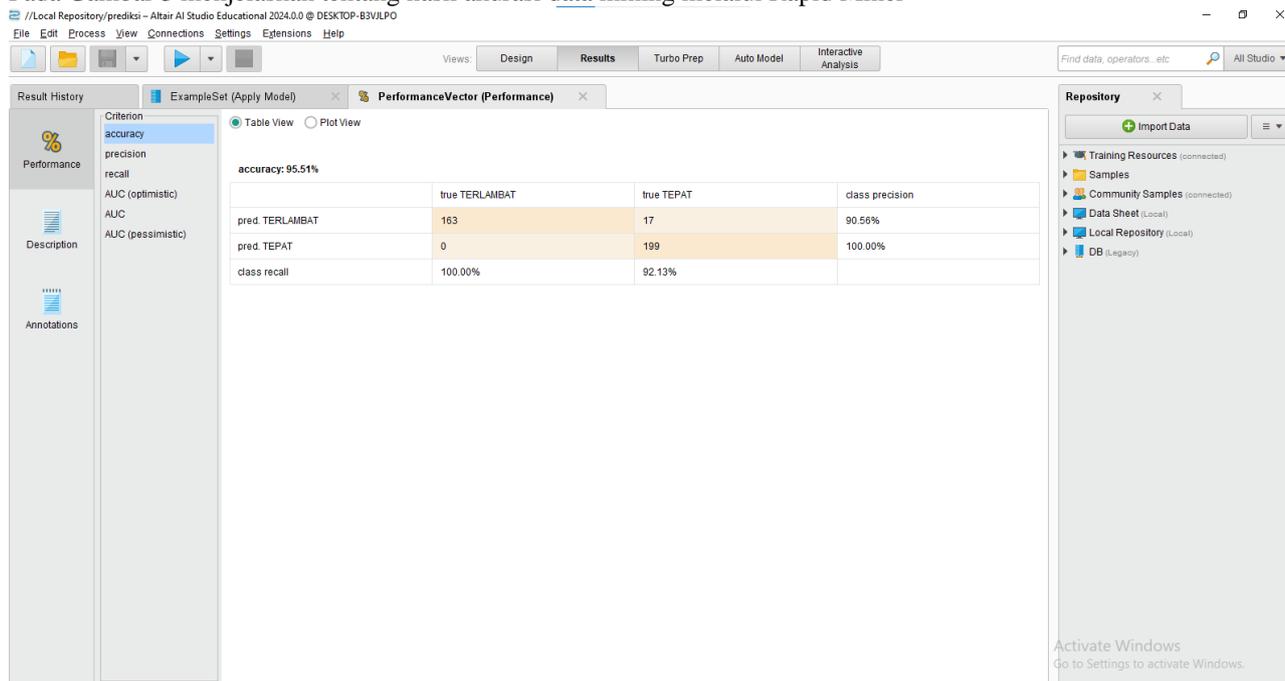
Pada Gambar 2 menjelaskan tentang hasil klasifikasi data mining melalui Rapid Miner



Gambar 2 Klasifikasi

Berdasarkan hasil klasifikasi pada Gambar diatas, penerapan metode K-NN menggunakan tools rapidminer dapat menghasilkan prediksi Siswa yang lulus tepat waktu dan siswa yang lulus terlambat.

Pada Gambar 3 menjelaskan tentang hasil akurasi data mining melalui Rapid Miner



Gambar 3 Akurasi

Dari hasil pengujian pada Gambar 3 didapat hasil akurasi sebesar 95,51% Terdapat Perhitungan kedekatan kasus lama pada data training dengan kasus baru pada data testing, diketahui dari 379 record data, 17 data yang tidak sesuai seharusnya lulus terlambat. Dari hasil pengujian dan analisa dapat dilihat bahwa dari 379 data yang digunakan terdapat 332 data siswa yang sesuai dengan prediksi, dan ada 17 data siswa yang prediksinya tidak sesuai dengan prediksi. Tingkat akurasi penerapan algoritma k-nearest neighbor ini sebesar 95,51% jika K nya 3.

KESIMPULAN

Metode K-Nearest Neighbor (K-NN) terbukti efektif untuk memprediksi kelulusan siswa di SMK Al Hasyimy Ibrahimy dengan aplikasi RapidMiner. Algoritma K-NN bekerja dengan menganalisis kesamaan data, menggunakan variabel seperti nilai Sumatif, UTS, PAS, UKK, dan PSAJ. Data siswa dipisahkan menjadi data pelatihan dan pengujian, sehingga model dapat dilatih untuk mengklasifikasikan status kelulusan siswa.

Hasil prediksi menunjukkan akurasi yang tinggi dalam membedakan siswa yang lulus tepat waktu dan terlambat. Dengan adanya prediksi ini, sekolah dapat merancang strategi pendidikan yang lebih efektif, memberikan perhatian lebih kepada siswa yang membutuhkan, dan meningkatkan angka kelulusan demi menjaga akreditasi.

UCAPAN TERIMA KASIH

Disampaikan Terimakasih kepada semua pihak yang telah mendukung dan membantu terlaksananya Jurnal ini. Terutama ditujukan kepada SMK Al Hasyimy Ibrahimy atas pemberian data dan akses yang diperlukan, serta kepada dosen pembimbing di Universitas Ibrahimy atas bimbingan, arahan, dan saran berharga yang telah diberikan. Ucapan terima kasih juga untuk keluarga dan rekan-rekan yang memberikan dukungan moral serta semangat yang sangat berarti. Semoga hasil yang didapat bisa bermanfaat bagi peningkatan mutu pendidikan dan kualitas kelulusan di tingkat pendidikan menengah kejuruan.

DAFTAR PUSTAKA

- : Fajjin Amik, S.Pd., M.Si.; Hj. Een Nuraini, S.Pd.; Agis Sugiarti, M. P. . dkk. (2016). *No Title MENUJU GURU DAN SISWA CERDAS*.
- Adam Puspabuana, S. Kom., M. Kom, Amril Mutoi siregar, S. Kom., M, K. (2020). *No Title Data Mining*.
- Ahmad Yogiarto,Ahmad Homaidi, Z. F. (2024). *G-Tech : Jurnal Teknologi Terapan*. 8(3), 1720–1728.
- Andarsyah, M. R. R. (2023). *No Title KOMPARASI PERFORMA MODEL TERHADAP KLASIFIKASI SINYAL MIT-BIH ARRHYTHMIA DATABASE*.
- Esi Vidia Rahemadani, Syafriah Fachri Pane, N. H. H. (2020). *No Title Algoritma C4.5 dan K-Nearest Neighbors (KNN) Untuk Memetakan Matakuliah dan Keterlambatan Kelulusan Mahasiswa*.
- Fadillah, J. R., & Pradana, R. (2024). *IMPLEMENTASI ALGORITMA K-NEAREST NEIGHBORS IMPLEMENTATION OF THE K-NEAREST NEIGHBORS ALGORITHM IN DETERMINING STUDENT ACHIEVEMENT CATEGORIES BASED*. 3(September), 426–434.

- Kusrini, emha taufiq luthfi. (2023). *ALGORITMA DATA MINING*. penerbit andi.
https://www.google.co.id/books/edition/Algoritma_Data_Mining/-Ojclag73O8C?hl=id&gbpv=0
- Prianto, E. A. N. W. I. R. C. (2020). *No Title SISTEM PERBANDINGAN ALGORITMA K-MEANS DAN NAIVE BAYES UNTUK MEMPREDIKSI PRIORITAS PEMBAYARAN TAGIHAN RUMAH SAKIT BERDASARKAN TINGKAT KEPENTINGAN*.
- Prof. Dr. Muhammad Yaumi, M.Hum., M. A. (2018). *No Title MEDIA DAN TEKNOLOGI PEMBELAJARAN*.
- Prof. Dr. Prayitno, M.Sc., E. (2021). *No Title DASAR TEORI dan PRAKSIS PENDIDIKAN*.
- Purwaningsih, E., & Nurelasari, E. (2021). *Penerapan K-Nearest Neighbor Untuk Klasifikasi Tingkat Kelulusan Pada Siswa*. 10(01), 46–55.
- RAYMANA APRILIAN RONI HABIBI, S.T., M.Kom M. YUSRIL HELMI SETYAWAN, S.KOM., M. K. (2019). *No Title ALGORITMA KNN DALAM MEMPREDIKSI CUACA UNTUK MENENTUKAN TANAMAN YANG COCOK SESUAI MUSIM*.