



Sistem Prediksi Prestasi Akademik Siswa Menggunakan Algoritma K-Nearest Neighbor (KNN)

Febriansyah*, Siti Muntari

Program Studi Teknik Informatika, Institut Teknologi Pagar Alam, Pagar Alam, Indonesia

Email: ^{1,*}febriansyahh1213@gmail.com, ²muntariaza@gmail.com

Email Penulis Korespondensi: febriansyahh1213@gmail.com

Abstrak—Penelitian ini bertujuan untuk mengklasifikasikan tingkat prestasi siswa berdasarkan kemampuan akademik dalam menguasai materi pelajaran menggunakan metode K-Nearest Neighbor (K-NN). Penelitian ini dilatarbelakangi oleh keterbatasan sistem pengolahan data nilai siswa yang masih dilakukan secara manual menggunakan Microsoft Excel, di mana proses penjumlahan dan pengelompokan nilai ke dalam kategori rendah hingga tinggi memerlukan waktu yang cukup lama serta menyulitkan guru dalam mengidentifikasi tingkat prestasi siswa, seperti kategori Baik, Cukup, dan Kurang. Data yang digunakan terdiri dari 134 siswa dengan 11 atribut mata pelajaran sebagai variabel input dalam proses klasifikasi. Hasil penelitian menunjukkan bahwa dari 134 data siswa, sebanyak 90 siswa diklasifikasikan ke dalam kategori Baik, 20 siswa kategori Cukup, dan 24 siswa kategori Kurang. Pengujian menggunakan RapidMiner menunjukkan bahwa metode K-NN memperoleh tingkat akurasi sebesar 89,55%, yang menunjukkan bahwa metode ini efektif dalam mengelompokkan tingkat prestasi siswa. Evaluasi performa dilakukan menggunakan Confusion Matrix untuk membandingkan hasil klasifikasi dengan data actual. Pengujian dilakukan sebanyak 10 kali dengan total data yang diperoleh sebanyak 134 data siswa, algoritma K-Nearest Neighbor (KNN) menghasilkan akurasi mencapai 89,55% dan nilai kappa sebesar 0,763. Angka ini menunjukkan bahwa model memiliki kinerja klasifikasi yang sangat baik dan tingkat keandalan yang tinggi. Model yang dihasilkan selanjutnya dikembangkan menjadi system prediksi berbasis web yang diuji menggunakan metode expert system melalui pendekatan Black-box Testing dan memperoleh tingkat kelayakan sebesar 83,44%.

Kata Kunci: Data Mining; Klasifikasi; K-Nearest Neighbor (K-NN); Prestasi Akademik Siswa; Confusion Matrix

Abstract—This study aims to classify student achievement levels based on academic ability in mastering subject matter using the K-Nearest Neighbor (K-NN) method. This study is motivated by the limitations of the student grade data processing system which is still done manually using Microsoft Excel, where the process of adding and grouping grades into low to high categories takes a long time and makes it difficult for teachers to identify student achievement levels, such as Good, Sufficient, and Poor categories. The data used consists of 134 students with 11 subject attributes as input variables in the classification process. The results show that from 134 student data, 90 students are classified into the Good category, 20 students are classified into the Sufficient category, and 24 students are classified into the Poor category. Testing using RapidMiner shows that the K-NN method obtains an accuracy level of 89.55%, which indicates that this method is effective in grouping student achievement levels. Performance evaluation is carried out using a Confusion Matrix to compare the classification results with actual data. Testing was conducted 10 times with a total of 134 student data obtained, the K-Nearest Neighbor (KNN) algorithm produced an accuracy of 89.55% and a kappa value of 0.763. These figures indicate that the model has excellent classification performance and a high level of reliability. The resulting model was then developed into a web-based prediction system that was tested using the expert system method through the Black-box Testing approach and obtained a feasibility level of 83.44%.

Keywords: Data Mining; Classification; K-Nearest Neighbor (K-NN); Student Academic Achievement; Confusion Matrix

1. PENDAHULUAN

Prestasi belajar merupakan sarana untuk mengukur indikator kesuksesan dari proses belajar mengajar atau untuk menilai efektivitas program pendidikan di sekolah. Hasil belajar yang dicapai oleh siswa merupakan hasil dari interaksi berbagai faktor yang memengaruhi prestasi [1]. Seorang peserta didik akan berhasil dalam pelajarannya apabila dalam diri peserta didik itu ada keinginan untuk belajar [2]. Setiap individu yang ingin mencapai sesuatu pasti memiliki motivasi yang lebih untuk mewujudkan harapan atau impian yang diinginkan.[3][4]

Data Mining adalah disiplin ilmu yang relatif baru tetapi sudah memberikan manfaat yang signifikan saat ini. Keuntungan yang dirasakan dengan penggunaan *Data Mining* sudah tidak diragukan lagi. Banyak kasus telah berhasil diselesaikan dengan penerapan teknik data mining ini[5][6]. Data mining juga merupakan proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu[7][8][9].

Dalam penelitian ini diterapkan K-Nearest Neighbor (K-NN) merupakan salah satu algoritma klasifikasi yang bekerja dengan menentukan kelas suatu data berdasarkan kedekatannya terhadap sejumlah data pelatihan (training data) yang memiliki jarak paling dekat [10][11]. Prinsip kerja algoritma ini adalah menghitung jarak antara data uji dengan seluruh data pelatihan, kemudian menentukan kelas berdasarkan mayoritas dari sejumlah tetangga terdekat (nearest neighbors) yang telah ditentukan nilai **K**-nya [12][13].

Dalam dunia pendidikan, prestasi belajar menjadi salah satu indikator utama untuk menilai kemampuan dan keberhasilan siswa dalam mengikuti proses pembelajaran di sekolah maupun lembaga pendidikan. Peningkatan prestasi belajar dipengaruhi oleh berbagai faktor, di antaranya minat belajar dan waktu belajar yang dimiliki siswa. Namun, masih terdapat beberapa hambatan yang sering ditemui, seperti rendahnya minat belajar serta keterbatasan waktu belajar yang efektif di kelas. Kondisi tersebut dapat menghambat peningkatan prestasi belajar siswa sehingga diperlukan suatu metode yang mampu membantu mengidentifikasi atau mengklasifikasikan tingkat prestasi belajar berdasarkan faktor-faktor yang memengaruhinya[14][15].

Berdasarkan permasalahan tersebut, penelitian ini menerapkan algoritma K-Nearest Neighbor (K-NN) untuk mengklasifikasikan prestasi akademik siswa berdasarkan hasil belajar yang diperoleh. Proses klasifikasi dilakukan dengan memanfaatkan data siswa, data nilai, dan data peringkat siswa di SMP Negeri 4 Pagar Alam sebagai data pelatihan dan data uji [16]. Hasil klasifikasi diharapkan dapat membantu pihak sekolah dalam mengelompokkan siswa sesuai dengan kemampuan akademiknya sehingga pendidik dapat menerapkan strategi pembelajaran yang lebih tepat dan sesuai dengan kebutuhan masing-masing siswa. Dengan demikian, proses pembelajaran diharapkan menjadi lebih efektif dalam meningkatkan kualitas pembelajaran dan prestasi akademik siswa [17][18]..

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian



Gambar 1. Tahapan Rapid Application Development (RAD)

Rapid Application Development (RAD) pada gambar 1 merupakan suatu model proses pengembangan perangkat lunak yang bersifat sekuensial dan linier, dengan penekanan pada siklus pengembangan yang sangat singkat (sekitar 60 hingga 90 hari)[19]

2.2 Metode Pengujian

Metode pengujian pada penelitian ini menerapkan pengujian *Confusion Matrix* dan *Black Box Testing*. *Confusion Matrix* digunakan untuk menilai seberapa efektif algoritma K-Nearest Neighbor dalam klasifikasi data, dengan penekanan pada akurasi algoritma yang diterapkan. Pengujian *Black Box* berfokus pada aspek-aspek aplikasi, seperti tampilan, fungsi-fungsi, dan kesesuaian alur fungsi dengan sistem yang diinginkan oleh pengembang. Pengujian Alpha menggunakan *Black Box Testing* mencakup pengujian antarmuka, pengujian algoritma, pengujian basis data, dan pengujian fungsionalitas sistem.

2.3 Confusion Matrix

Confusion matrix adalah suatu metode yang biasanya digunakan untuk melakukan perhitungan akurasi pada konsep *data mining*. *Confusion matrix* digambarkan dengan tabel yang menyatakan jumlah data uji yang benar diklasifikasikan dan jumlah data uji yang salah diklasifikasikan [7]

Tabel 1. Rumus Confusion Matrix

Confusion Matrix	Nilai Aktual	
	Positif	Negatif
Nilai Prediksi	Positif	True Positives
	Negatif	False Positives
Nilai Prediksi	Positif	False Negatives
	Negatif	True Negatives

Dimana *True Positive (TP)* adalah jumlah data yang diprediksi positif dan sesuai dengan kondisi sebenarnya (aktual positif). *False Positive (FP)* adalah jumlah data yang diprediksi positif, tetapi kondisi sebenarnya adalah negatif. *False Negative (FN)* adalah jumlah data yang diprediksi negatif, tetapi kondisi sebenarnya adalah positif dan *True Negative (TN)* adalah jumlah data yang diprediksi negatif dan sesuai dengan kondisi sebenarnya (aktual negatif).

Confusion Matrix digunakan untuk mengevaluasi kinerja model klasifikasi dengan membandingkan hasil prediksi terhadap data aktual. Berdasarkan nilai TP, FP, FN, dan TN, dapat dihitung metrik evaluasi seperti accuracy, precision, recall, dan F1-score untuk mengukur tingkat performa algoritma klasifikasi yang digunakan.



2.4 Tahapan K-Nearest Neighbor

Dataset penelitian diperoleh dari SMP Negeri 4 Kota Pagar Alam pada tahun 2024 sebanyak 134 data siswa. Data tersebut terdiri atas atribut identitas, yaitu NISN dan Nama, serta atribut akademik berupa nilai mata pelajaran. Dalam proses klasifikasi menggunakan algoritma K-Nearest Neighbor (K-NN), hanya atribut akademik yang digunakan sebagai variabel input, sedangkan NISN dan Nama digunakan sebagai identitas data dan tidak dilibatkan dalam proses perhitungan klasifikasi.

- Menetapkan Parameter K. yang tepat untuk data yang digunakan, Nilai K harus minimal 1 dan maksimalnya adalah jumlah dari data pelatihan.
- Mengukur Jarak antara data pelatihan dan data pengujian. Pengukuran jarak yang paling sering digunakan dalam perhitungan algoritma K-NN adalah dengan menggunakan jarak Euclidean. Kesimpulannya adalah sebagai berikut:

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

Dimana $d(p, q)$ adalah jarak Euclidean antara dua data. p_i adalah nilai atribut ke- i pada data pertama. q_i adalah nilai atribut ke- i pada data kedua. n adalah jumlah atribut yang digunakan dalam perhitungan. $\sum_{i=1}^n$ adalah penjumlahan dari atribut pertama hingga atribut ke- n .

- Mengatur urutan jarak yang terbentuk.
- Menetapkan jarak terdekat hingga urutan K.
- Mencocokkan kelas yang relevan
- Mencari jumlah kelas dari tetangga terdekat dan tetapkan kelas itu sebagai kelas data yang akan dianalisis

3. HASIL DAN PEMBAHASAN

3.1 Data Selection

Pada tahap ini, digunakan 13 atribut dari data siswa SMP Negeri 4 Kota Pagar Alam dengan total 134 *record*. Atribut tersebut meliputi dua data identitas, yaitu NISN dan nama siswa, serta 11 atribut nilai mata Pelajaran seperti terlihat pada Tabel 2. Selain itu, terdapat satu atribut tambahan berupa keterangan prestasi sebagai label klasifikasi dengan tiga kategori: "Baik", "Cukup", dan "Kurang".

Tabel 2. Data Selection

No	NISN	NAMA SISWA	Pai	Pkn	B. Indo	Mtk	Ipa	Ips	B. ing	Sbk	Pjok	Mulok	Tik
1	0109146851	ABI YAHYA	80	83	79	78	78	80	79	80	78	80	78
2	0099176074	Ade Saputra	71	72	69	67	72	70	75	67	65	67	75
3	115262217	Alisa Yorenza	87	78	84	82	80	80	75	80	78	85	79
4	0099277063	ANITA LERA AVENZA	74	67	71	65	68	68	69	71	71	75	71
5	0094126131	AZZAHARA AGATHA	87	81	85	86	83	82	80	84	84	82	78
6	0098985365	BINTANG ASATRISIYAH	69	67	65	69	74	71	71	75	73	74	62
7	0102383440	CHIANG CHI	65	75	69	65	67	65	69	70	80	70	68
8	0114069074	Dia Bunga Lestari	80	79	79	82	80	79	86	82	83	82	83
9	0103242674	DWI OKTAVIANI	87	79	79	84	79	80	80	85	85	85	84
11	Dst.....134												

3.2 Data Processing

Pada tahap data processing, dilakukan pemeriksaan menyeluruh terhadap data yang akan digunakan agar memastikan tidak ada missing value atau data kosong. Dari hasil pemeriksaan yang divisualisasikan pada gambar, terlihat bahwa ke-13 atribut yang digunakan memiliki nilai missing sebesar nol.

Hal ini menandakan bahwa seluruh data yang diperoleh lengkap dan tidak ada bagian yang hilang, sehingga dapat diproses lebih lanjut tanpa memerlukan teknik imputasi atau pengisian data. Dataset yang digunakan terdiri dari 134 data siswa SMP Negeri 4 Kota Pagar Alam, dengan total 14 atribut. Atribut-atribut tersebut mencakup data identitas siswa berupa NISN dan nama, serta 11 atribut nilai dari berbagai mata pelajaran yang diajarkan di sekolah. Selain itu, terdapat satu atribut keterangan prestasi yang berfungsi sebagai label klasifikasi.

Proses pengolahan data dilakukan menggunakan perangkat lunak RapidMiner Studio dengan metode klasifikasi *K-Nearest Neighbor* (KNN). Dengan data yang sudah bersih dan lengkap, proses pemodelan dapat berjalan optimal untuk mengklasifikasikan prestasi siswa berdasarkan nilai akademiknya. Tahap ini sangat penting untuk memastikan bahwa hasil klasifikasi nanti akurat dan dapat dipercaya

Name	Type	Missing	Statistics	Filter (14 / 14 attributes)	Search for Attributes
▼ NISN	Nominal	0	Least 10436067679 (1)	Most 00488551842 (1)	Values 00488551842 (1), 06019075007 (1),...
▼ NAMA	Nominal	0	Least adell aidi latif (1)	Most ABI YAHYA (1)	Values ABI YAHYA (1), AVEVYA AJGZAA (1),...
▼ NILAI PAI	Integer	0	Min 56	Max 88	Average 75,806
▼ NILAI PKN	Integer	0	Min 57	Max 90	Average 71,254
▼ NILAI B.INDONESIA	Integer	0	Min 56	Max 90	Average 77,015
▼ NILAI MTK	Integer	0	Min 58	Max 90	Average 78,754
▼ NILAI IPA	Integer	0	Min 56	Max 86	Average 78,388
▼ NILAI IPS	Integer	0	Min 60	Max 87	Average 75,433
▼ NILAI B.ING	Integer	0	Min 55	Max 88	Average 76,157

Gambar 2. Data Processing

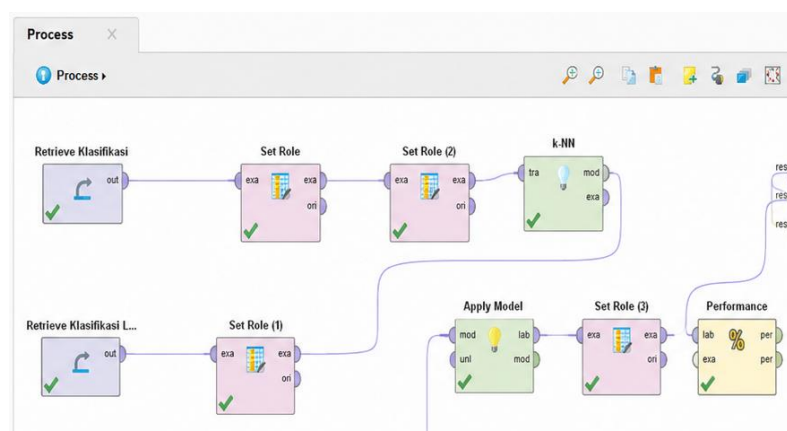
Pada penelitian ini, data siswa yang telah dikumpulkan diperiksa terlebih dahulu untuk memastikan tidak terdapat kesalahan format, data ganda, maupun nilai yang tidak sesuai. Selanjutnya dataset diimpor ke dalam aplikasi RapidMiner sehingga menghasilkan *ExampleSet* seperti yang ditunjukkan pada Gambar 2. Pada tahap ini RapidMiner secara otomatis mengidentifikasi tipe data setiap atribut, yaitu atribut NISN dan Nama bertipe Nominal, sedangkan atribut nilai mata pelajaran seperti Nilai PAI, Nilai PKN, Nilai Bahasa Indonesia, Nilai Matematika, Nilai IPA, Nilai IPS, Nilai Bahasa Inggris, dan atribut nilai lainnya bertipe Integer. Selain itu, sistem juga menampilkan informasi jumlah data sebanyak 134 record dengan 14 atribut serta ringkasan statistik berupa nilai minimum, maksimum, dan rata-rata (*average*) untuk setiap atribut numerik. Hasil preprocessing menunjukkan bahwa seluruh data telah berada dalam format yang sesuai sehingga dapat digunakan pada tahap transformasi data dan pemodelan menggunakan algoritma *K-Nearest Neighbor*.

3.3 Data Transformation

Pada penelitian ini, data yang telah melalui proses seleksi atribut kemudian ditransformasikan dengan menyatukan atribut-atribut yang dipilih ke dalam satu dataset menggunakan aplikasi RapidMiner. Proses transformasi meliputi pengaturan tipe data setiap atribut, penentuan atribut yang berperan sebagai label (kelas), serta penyesuaian format data agar sesuai dengan kebutuhan algoritma K-Nearest Neighbor (K-NN). Hasil dari tahap ini adalah dataset yang telah siap digunakan pada proses pemodelan (modeling) dan pengujian klasifikasi di RapidMiner.

3.4 Pemodelan (Modelling)

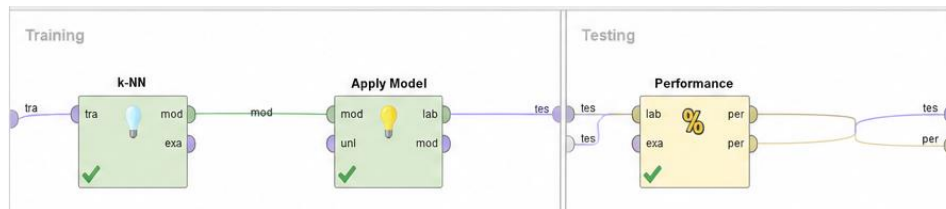
Pada tahap ini dilakukan proses klasifikasi data prestasi siswa menggunakan algoritma K-Nearest Neighbor (KNN) di aplikasi RapidMiner. Dataset dimuat menggunakan operator Retrieve, kemudian atribut Prestasi Siswa ditetapkan sebagai atribut target (label) menggunakan operator Set Role. Selanjutnya model KNN dibangun menggunakan operator K-NN dengan parameter $k = 5$ dan distance measurement menggunakan Euclidean Distance, karena metode ini sesuai untuk mengukur kedekatan antar data numerik berdasarkan jarak garis lurus antar atribut. Nilai $k = 5$ dipilih agar proses klasifikasi mempertimbangkan lima tetangga terdekat dalam menentukan kelas suatu data sehingga dapat mengurangi pengaruh data penculan (*outlier*).



Gambar 3. Pemodelan (Modelling)

Gambar 3 menunjukkan proses pemodelan menggunakan algoritma K-Nearest Neighbor (KNN) pada RapidMiner. Gambar menunjukkan rangkaian proses mulai dari pengambilan dataset (*Retrieve*), penentuan atribut label (*Set Role*), penggabungan data (*Join*), penerapan model klasifikasi (*Apply Model*), hingga evaluasi hasil klasifikasi menggunakan operator *Performance (Classification)*.

Untuk mengevaluasi kinerja model, digunakan operator Cross Validation dengan 10-fold cross validation, di mana dataset dibagi menjadi sepuluh bagian. Pada setiap iterasi, sembilan bagian digunakan sebagai data pelatihan (training) dan satu bagian sebagai data pengujian (testing) secara bergantian hingga seluruh data menjadi data uji. Pada bagian Training, model KNN dibentuk menggunakan data pelatihan. Selanjutnya pada bagian Testing, model diterapkan menggunakan operator Apply Model, kemudian hasil klasifikasi dievaluasi menggunakan operator Performance (Classification) untuk memperoleh nilai akurasi, presisi, recall, dan confusion matrix. Hasil evaluasi menunjukkan bahwa model mampu mengklasifikasikan prestasi siswa berdasarkan nilai akademik secara efektif sehingga dapat membantu pihak sekolah dalam proses pengambilan keputusan.



Gambar 4. Pemodelan Algoritma

Gambar 4 merupakan Proses evaluasi model menggunakan 10-Fold Cross Validation pada RapidMiner. Gambar menunjukkan bahwa model K-Nearest Neighbor (KNN) dievaluasi menggunakan metode 10-Fold Cross Validation, di mana dataset dibagi menjadi sepuluh bagian. Pada setiap iterasi, sembilan bagian digunakan sebagai data pelatihan, sedangkan satu bagian digunakan sebagai data pengujian. Hasil pengujian kemudian dievaluasi menggunakan operator Performance untuk memperoleh nilai performa model.

3.5 Hasil & Evaluasi Algoritma *K-NN*

Gambar X menunjukkan hasil klasifikasi data siswa SMP Negeri 4 Pagar Alam menggunakan algoritma K-Nearest Neighbor (KNN) yang diimplementasikan melalui perangkat lunak RapidMiner. Dataset yang digunakan terdiri atas 134 data siswa yang meliputi atribut identitas, yaitu NISN dan Nama, sebelas atribut nilai mata pelajaran sebagai variabel prediktor, serta satu atribut target berupa prestasi siswa yang diklasifikasikan ke dalam tiga kategori, yaitu Baik, Cukup, dan Kurang.

Pada hasil klasifikasi ditampilkan dua kolom utama, yaitu prestasi sebagai label aktual dan prediction (prestasi) sebagai hasil prediksi yang dihasilkan oleh model KNN. Selain itu, RapidMiner juga menampilkan nilai confidence untuk setiap kategori prestasi yang menunjukkan tingkat keyakinan model terhadap hasil prediksi. Berdasarkan hasil yang diperoleh, sebagian besar data memiliki kesesuaian antara label aktual dan hasil prediksi, yang menunjukkan bahwa model KNN mampu mengenali pola nilai akademik siswa dengan baik.

Kinerja model tersebut didukung oleh hasil evaluasi menggunakan metode 10-Fold Cross Validation, yang menghasilkan akurasi sebesar 89,55% dengan nilai Kappa sebesar 0,763. Nilai akurasi tersebut menunjukkan bahwa model mampu mengklasifikasikan data dengan tingkat ketepatan yang tinggi, sedangkan nilai Kappa sebesar 0,763 termasuk dalam kategori substantial agreement, yang mengindikasikan adanya tingkat kesesuaian yang baik antara hasil

prediksi model dan kondisi sebenarnya. Dengan demikian, algoritma K-Nearest Neighbor (KNN) memiliki performa yang baik dan layak digunakan untuk mengklasifikasikan prestasi siswa berdasarkan nilai akademik

Row No.	prestasi	prediksitar...	confiden...	confidenc...	confidenc...	nisn	nama	nilai_pai	nilai_pkn	nilai_indone
9	Baik	Baik	1	0	0	10245874	DWI OKTAVIA	87	78	78
11	Baik	Baik	1	0	0	10495824	FIKRI HUDAW	65	78	82
12	Baik	Baik	1	0	0	97583032	Indria Lestari	68	79	85
16	Baik	Baik	1	0	0	10427539	NILAIN CAHY	68	63	88
27	Baik	Baik	1	0	0	99600119	DESYRIANTO	76	83	93
32	Baik	Baik	1	0	0	96532883	JUPITALEST	75	71	84
33	Baik	Baik	1	0	0	11401939	Laura Auren	60	73	78
37	Baik	Baik	1	0	0	96560374	MUHAMMO	82	71	65
38	Baik	Baik	1	0	0	96844829	NALA KIRAN	61	73	83
39	Baik	Baik	1	0	0	117568049	Panca Keksaa	61	71	84
42	Baik	Baik	1	0	0	10126964	ROZILA ULK	61	73	81
44	Baik	Baik	1	0	0	972448312	SELA APRIANI	73	73	75
45	Baik	Baik	1	0	0	101306865	THALIA FEBY	81	73	79
47	Baik	Baik	1	0	0	97878719	Yesi Amelia	82	73	84
56	Baik	Baik	1	0	0	112423676	AULIA HAMILI	61	73	84
58	Baik	Baik	1	0	0		Jargan Sahata	63	73	82

Gambar 5. Hasil Proses Algoritma K-NN

Gambar 5 adalah Hasil prediksi klasifikasi menggunakan algoritma K-Nearest Neighbor (KNN). Gambar menampilkan hasil klasifikasi setiap data siswa yang terdiri atas kelas aktual (prestasi), kelas hasil prediksi (prediction), serta nilai confidence yang menunjukkan tingkat keyakinan model terhadap hasil prediksi. Data dengan nilai prediksi yang sama dengan kelas aktual menunjukkan bahwa proses klasifikasi dilakukan dengan benar.

3.6 Deployment

Pada tahapan terakhir ini dapat model dengan akurasi tinggi selanjutnya di implementasikan pada sistem berbasis web untuk dapat mengklasifikasi prestasi akademik siswa. Dimana sistem ini selanjutnya dapat digunakan oleh user yaitu guru, siswa dan kepala sekolah. Berikut tampilan halaman sistem yang di implementasikan.

a. Halaman Data Siswa

Lembar yang berisi data siswa terdiri dari No, Nisn, nama, Nilai Pai, Nilai Pkn, Nilai B. Indonesia, Nilai Mtk, Nilai Ipa, Nilai Ips, Nilai B. Inggris, Nilai S.Budaya, Nilai Pjok, Nilai Mulok, Nilai Tik seperti pada gambar 6.

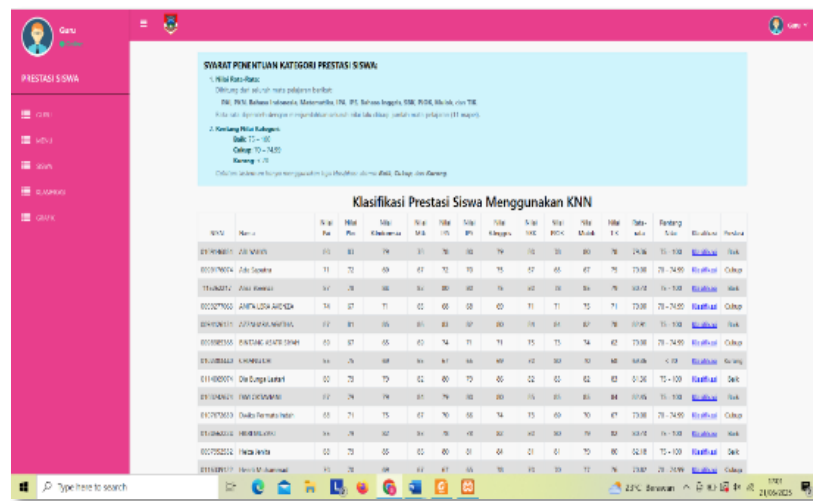
No	Nisn	Nama	Nilai Pai	Nilai Pkn	Nilai B. Indonesia	Nilai Mtk	Nilai Ipa	Nilai Ips	Nilai B. Inggris	Nilai S. Budaya	Nilai Pjok	Nilai Mulok	Nilai Tik	Action
1	013384890	MUHAMMAD ABBAS PUTRA ADITYA	84	79	87	87	83	86	84	85	85	88	88	
2	009857883	MUHAMMAD ALIF HANZAH	82	79	90	89	84	82	90	81	84	83	80	
3	013433406	Muhammad Agung Saputero	83	79	84	85	84	88	83	84	84	85	85	
4	01817848	Muhammad Abdir Rozzaq	81	79	78	78	82	81	81	82	82	84	84	
5	01368122	MUHAMMAD BEN ANTONI	82	79	90	89	84	82	90	81	84	83	80	
6	3134029673	MUHAMMAD RALZAN ANGAR	81	79	84	85	82	81	81	82	82	88	85	

Gambar 6. Halaman Data Siswa

b. Halaman Hasil Klasifikasi

Halaman ini adalah halaman yang tampil pada saat admin mengklik menu klasifikasi, pada halaman ini kepala sekolah melihat hasil. Data yang tersimpan pada halaman ini berupa Nisn, Nama, Nilai Pai, Nilai B. Indonesia, Nilai Pkn,

Nilai Mtk, Nilai Ipa, Nilai Ips, Nilai B. Inggris, Nilai S.Budaya, Nilai Pjok, Nilai M.Lokal, Nilai Tik, Prestasi. Hasil klasifikasi terdiri dari Baik, Cukup, Kurang seperti ditunjukkan pada gambar 7.



SIKAT PENENTUAN KATEGORI PRESTASI SISWA

1. **Nilai Mata Pelajaran**
 Masukkan data nilai mata pelajaran berikut:
 IPK, IPS, Bahasa Indonesia, Matematika, IPA, Pj. Sosial, Inggris, SBK, PKN, Musik, dan TIK.

2. **Penentuan Hasil Keluaran**
 Baik (75 - 100)
 Cukup (50 - 74,99)
 Kurang (< 50)

3. **Nilai dan Prestasi**
 Masukkan data nilai dan prestasi siswa berikut ke dalam tabel.

Klasifikasi Prestasi Siswa Menggunakan KNN

NO	Nama	Nilai IPA	Nilai IPS	Nilai Mat	Nilai Pj. S.	Nilai Inggris	Nilai SBK	Nilai PKN	Nilai Musik	Nilai TIK	Prestasi	Ranking	Kategori
00000001	Adi Wicak	85	80	75	85	80	85	80	85	80	80	75	Baik
00000002	Adi Wicak	75	70	65	75	70	75	70	75	70	70	65	Cukup
00000003	Adi Wicak	65	60	55	65	60	65	60	65	60	60	55	Kurang
00000004	Adi Wicak	55	50	45	55	50	55	50	55	50	50	45	Kurang
00000005	Adi Wicak	45	40	35	45	40	45	40	45	40	40	35	Kurang
00000006	Adi Wicak	35	30	25	35	30	35	30	35	30	30	25	Kurang
00000007	Adi Wicak	25	20	15	25	20	25	20	25	20	20	15	Kurang
00000008	Adi Wicak	15	10	5	15	10	15	10	15	10	10	5	Kurang
00000009	Adi Wicak	5	0	0	5	0	5	0	5	0	0	0	Kurang
00000010	Adi Wicak	0	0	0	0	0	0	0	0	0	0	0	Kurang
00000011	Adi Wicak	0	0	0	0	0	0	0	0	0	0	0	Kurang
00000012	Adi Wicak	0	0	0	0	0	0	0	0	0	0	0	Kurang
00000013	Adi Wicak	0	0	0	0	0	0	0	0	0	0	0	Kurang
00000014	Adi Wicak	0	0	0	0	0	0	0	0	0	0	0	Kurang
00000015	Adi Wicak	0	0	0	0	0	0	0	0	0	0	0	Kurang
00000016	Adi Wicak	0	0	0	0	0	0	0	0	0	0	0	Kurang
00000017	Adi Wicak	0	0	0	0	0	0	0	0	0	0	0	Kurang
00000018	Adi Wicak	0	0	0	0	0	0	0	0	0	0	0	Kurang
00000019	Adi Wicak	0	0	0	0	0	0	0	0	0	0	0	Kurang
00000020	Adi Wicak	0	0	0	0	0	0	0	0	0	0	0	Kurang
00000021	Adi Wicak	0	0	0	0	0	0	0	0	0	0	0	Kurang
00000022	Adi Wicak	0	0	0	0	0	0	0	0	0	0	0	Kurang
00000023	Adi Wicak	0	0	0	0	0	0	0	0	0	0	0	Kurang
00000024	Adi Wicak	0	0	0	0	0	0	0	0	0	0	0	Kurang
00000025	Adi Wicak	0	0	0	0	0	0	0	0	0	0	0	Kurang
00000026	Adi Wicak	0	0	0	0	0	0	0	0	0	0	0	Kurang
00000027	Adi Wicak	0	0	0	0	0	0	0	0	0	0	0	Kurang
00000028	Adi Wicak	0	0	0	0	0	0	0	0	0	0	0	Kurang
00000029	Adi Wicak	0	0	0	0	0	0	0	0	0	0	0	Kurang
00000030	Adi Wicak	0	0	0	0	0	0	0	0	0	0	0	Kurang

Gambar 7. Halaman Hasil Klasifikasi

3.7 Pembahasan

Berdasarkan hasil pengujian menggunakan 10-Fold Cross Validation, algoritma K-Nearest Neighbor (KNN) memperoleh tingkat akurasi sebesar 89,55% dengan nilai Kappa sebesar 0,763, yang menunjukkan bahwa model memiliki kemampuan klasifikasi yang baik dalam mengelompokkan prestasi akademik siswa. Tingkat akurasi tersebut menunjukkan bahwa atribut nilai mata pelajaran yang digunakan mampu merepresentasikan karakteristik prestasi siswa sehingga proses klasifikasi dapat dilakukan secara efektif.

Hasil penelitian ini sejalan dengan penelitian Muhaimin dkk.[16], yang juga menerapkan algoritma K-Nearest Neighbor untuk klasifikasi prestasi akademik siswa berdasarkan nilai rapor. Penelitian tersebut menunjukkan bahwa algoritma KNN mampu menghasilkan performa klasifikasi yang baik dalam membantu proses pengambilan keputusan di bidang pendidikan. Persamaan penelitian ini dengan penelitian Muhaimin dkk. terletak pada penggunaan algoritma KNN untuk mengklasifikasikan prestasi siswa, sedangkan perbedaannya terdapat pada atribut yang digunakan, jumlah data, serta implementasi hasil penelitian yang pada penelitian ini dikembangkan menjadi sistem berbasis web.

Selain itu, hasil penelitian ini juga mendukung penelitian yang dilakukan oleh Putra [13] serta Novayanti dkk. [17], yang menyatakan bahwa algoritma KNN memiliki kemampuan yang baik dalam menyelesaikan permasalahan klasifikasi karena proses penentuan kelas dilakukan berdasarkan kedekatan jarak antar data menggunakan Euclidean Distance. Pada penelitian ini digunakan nilai $k = 5$ sehingga proses klasifikasi menjadi lebih stabil terhadap data pencilan dan menghasilkan tingkat akurasi yang tinggi. Dengan demikian, penelitian ini memperkuat temuan-temuan sebelumnya bahwa algoritma KNN merupakan salah satu metode yang efektif untuk menyelesaikan permasalahan klasifikasi, khususnya pada bidang pendidikan.

4. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa penerapan algoritma K-Nearest Neighbor (KNN) dengan pendekatan CRISP-DM berhasil digunakan untuk mengklasifikasikan prestasi siswa di SMP Negeri 4 Kota Pagar Alam berdasarkan nilai akademik. Proses klasifikasi dilakukan melalui tahapan pemahaman data, persiapan data, pemodelan, evaluasi, dan implementasi sistem berbasis web. Hasil pengujian menggunakan 10-Fold Cross Validation pada RapidMiner menunjukkan bahwa model mampu mengklasifikasikan data dengan tingkat akurasi sebesar 89,55% dan nilai Kappa sebesar 0,763, yang menunjukkan tingkat kesesuaian yang baik antara hasil prediksi dan data aktual. Model klasifikasi menghasilkan tiga kategori prestasi, yaitu 90 siswa berkategori Baik, 20 siswa berkategori Cukup, dan 24 siswa berkategori Kurang. Selanjutnya, hasil klasifikasi tersebut berhasil diimplementasikan ke dalam sistem berbasis web yang dapat dimanfaatkan oleh guru dan kepala sekolah sebagai pendukung pengambilan keputusan dalam mengidentifikasi prestasi siswa secara lebih cepat, objektif, dan terstruktur. Berdasarkan hasil pengujian sistem menggunakan kuesioner kepada pakar, sistem memperoleh nilai kelayakan sebesar 83,44%, sehingga termasuk dalam kategori sangat layak untuk diterapkan sebagai media pendukung dalam proses klasifikasi prestasi siswa.

REFERENCES

- [1] D. F. Firdaus, "Pengaruh Persepsi Siswa Tentang Kompetensi Profesional Guru Dan Kompetensi Sosial Guru Terhadap Prestasi Belajar Siswa Pada Mata Pelajaran Ekonomi Di Sman 1 Kuningan," *J. Syntax Idea*, vol. 2, no. 3, pp. 12–17, 2020, doi: <https://doi.org/10.46799/syntax-idea.v2i3.153>.
- [2] P. Astuti, S., "Pengaruh Kemampuan Awal dan Minat Belajar terhadap Prestasi Belajar Fisika," *Form. J. Ilm. Pendidik. MIPA*, vol. 5, no. 1, pp. 68–75, 2015, doi: <http://dx.doi.org/10.30998/formatif.v5i1.167>.



- [3] M. Makkawaru, "Pentingnya Pendidikan Bagi Kehidupan dan Pendidikan," *J. Konsepsi*, vol. 8, no. 3, pp. 116–119, 2019.
- [4] A. P. Albarado and V. Eminita, "Pengaruh kebiasaan belajar terhadap prestasi belajar siswa di mts khazanah kebajikan," *FIBONACCI J. Pendidik. Mat. dan Mat.*, vol. 6, no. 2, pp. 167–174, 2020, doi: <https://doi.org/10.24853/fbc.6.2.167-174>.
- [5] R. Dewi, Obert, and R. Gusmana, "Implementasi Metode K-Nearest Neighbor (KNN) dalam Pengelompokan Status Ekonomi Warga," *J. Big Data Anal. Artif. Intell.*, vol. 4, no. 1, pp. 1–8, 2018, [Online]. Available: <https://media.neliti.com/media/publications/308845-implementasi-metode-k-nearest-neighbor-k-ef5ce754.pdf>
- [6] F. Febriansyah and S. Muntari, "Penerapan Algoritma K-Means untuk Klasterisasi Penduduk Miskin pada Kota Pagar Alam," *JISKA (Jurnal Inform. Sunan Kalijaga)*, vol. 8, no. 1, pp. 66–77, 2023, doi: <https://doi.org/10.14421/jiska.2023.8.1.66-77>.
- [7] Febriansyah and Efan, "Implementasi Algoritma K-Means Untuk Klasterisasi Besarnya Penggunaan Listrik Rumah Tangga Di Pagar Alam," *J. Sist. Inf. Triguna Dharma (JURSI TGD)*, vol. 4, no. 3, pp. 437–445, 2025, doi: 10.53513/jursi.v4i3.10955.
- [8] T. K. Titus and M. Jajuli, "Clustering Data Kecelakaan Lalu Lintas di Kecamatan Cileungsi Menggunakan Metode K-Means," *Gener. J.*, vol. 6, no. 1, pp. 1–12, 2022, doi: 10.29407/gj.v6i1.16103.
- [9] C. S. Octiva, T. I. Fajri, E. B. Sulistiarini, S. Suharjo, and U. W. Nuryanto, "Penggunaan Teknik Data Mining untuk Analisis Perilaku Pengguna pada Media Sosial," *J. Minfo Polgan*, vol. 13, no. 1, pp. 1074–1078, 2024, doi: 10.33395/jmp.v13i1.13936.
- [10] S. N. Bakri and L. S. Harahap, "Analisis klasifikasi Algoritma K-Nearest Neighbor (K-NN) pada struktur Daerah di Kota Medan," *J. Ilmu Komput. dan Sist. Inf.*, vol. 4, no. 2, pp. 182–193, 2025, doi: 10.70340/jirsi.v4i2.165.
- [11] F. A. Nazhifa, M. S. Hasibuan, and Suhardi, "Klasifikasi Jenis Tanaman Anggrek Berdasarkan Daun Menggunakan Metode Principal Component Analysis Dan K-Nearest Neighbor I," *J. Ilm. Tek. dan Ilmu Komput.*, vol. 5, no. Vol. 5 No. 1 (2026): Februari, pp. 1–8, 2026, doi: <https://doi.org/10.55123/storage.v5i1.7086>.
- [12] M. B. Billah, I. Jangcik, Sasmita, and Yuhelmi, "Penerapan Metode K- Nearest Neighbor Klasifikasi Jenis Buah Semangka Dengan Image Processing," *Zo. J. Sist. Inf.*, vol. 5, no. 3, pp. 535–548, 2023, doi: <https://doi.org/10.31849/zn.v5i3.16301>.
- [13] I. L. Putra, "Implementasi Algoritma Particle Swarm Optimization(Pso) Dan K-Nearest Neighbor(K-Nn) Dalam Memprediksi Keberhasilan Anak Smk Mendapatkan Kerja," *Technol. J. Ilm.*, vol. 13, no. 4, p. 339, 2022, doi: 10.31602/tji.v13i4.8167.
- [14] R. Darajaad and D. Fitriyati, "Pengaruh Minat Belajar dan Jam Belajar Terhadap Prestasi Belajar Siswa pada Mata Pelajaran Ekonomi di Siswa Kelas XI IPS 3 SMA Negeri 1 Kesamben," *J. Pendidik. Ekon.*, vol. 4, no. 3, pp. 1–6, 2016, doi: <https://doi.org/10.26740/jupe.v4n3.p%25p>.
- [15] E. P. K. Wardany and H. A. Rigianti, "Pengaruh Kinerja Guru Terhadap Hasil Belajar Siswa di Sekolah Dasar," *Attadrib J. Pendidik. Guru Madrasah Ibtidaiyah*, vol. 6, no. 2, pp. 250–261, 2023, doi: 10.54069/attadrib.v6i2.541.
- [16] A. Muhaimin, M. Amin Hariyadi, and M. I. Imamudin, "Klasifikasi Prestasi Akademik Siswa Berdasarkan Nilai Rapor dan Kedisiplinan dengan Metode K-Nearest Neighbor," *J. Ilmu Komput. dan Sist. Inf.*, vol. 7, no. 1, pp. 193–202, 2024, doi: 10.55338/jikomsi.v7i1.2865.
- [17] R. H. N. L. G. P. S. I. G. N. A. K. I. G. A. D. S. P. D. Novayanti, "Analisis Perbandingan Algoritma K-Nearest Neighbor Klasifikasi Data Perbankan," *J. Inform. dan Komput.*, vol. 19, no. 1, pp. 167–173, 2025, doi: <https://doi.org/10.35959/jik.v13i01.693>.
- [18] R. N. Sesilia, H. Prihastuti, and Q. Arie, "Penerapan K-Nearest Neighbor (KNN) untuk Klasifikasi Anggrek Berdasarkan Karakter Morfologi Daun dan Bunga," *Komputasi*, vol. 15, no. 1, pp. 118–125, 2018, doi: <https://doi.org/10.33751/komputasi.v15i1>.
- [19] T. Pricillia and Zulfachmi, "Perbandingan Metode Pengembangan Perangkat Lunak (Waterfall, Prototype, RAD)," *J. Bangkit Indones.*, vol. 10, no. 1, pp. 6–12, 2021, doi: 10.52771/bangkitindonesia.v10i1.153.
- [20] Oon Wira Yuda, Darmawan Tuti, Lim Sheih Yee, and Susanti, "Penerapan Penerapan Data Mining Untuk Klasifikasi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode Random Forest," *SATIN - Sains dan Teknol. Inf.*, vol. 8, no. 2, pp. 122–131, 2022, doi: 10.33372/stn.v8i2.885.