JTIF | Jurnal INOVATIF WIRA WACANA Vol. 04 No. 02 September 2025, hlm 123-135 **pISSN (**2961-7316**)**
eISSN (2962-5998**)**

PENGELOMPOKAN PERFORMA SISWA DALAM PEMBELAJARAN MATEMATIKA DENGAN ALGORITMA *K-NEAREST NEIGHBOR*

(*Student Performance Grouping In Learning Mathematics With K-Nearest Neighbor Algorithm*)

Afendi Alan Halakadu[1], Arini Aha Pekuwali [2], Murry Albert Agustin Lobo[3]

[1,2]Program Studi Teknik Informatika, Universitas Kristen Wira Wacana Sumba

[3]Program Studi Sistem Informasi, Universitas Kristen Wira Wacana Sumba

E-mail: [1]alanart249@gmail.com, [2]arini.pekuwali@unkriswina.ac.id, [3]albertlobo@unkriswina.ac.id

KEYWORDS:
Numeracy Literacy, K-Nearest Neighbor, Classification, Mathematics, Student Learning Performance.

ABSTRACT

Numeracy literacy is a crucial aspect of Mathematics education at the elementary school level, as it plays a significant role in developing students' logical thinking and problem-solving skills. However, differences in comprehension levels often lead to variations in learning performance that need to be identified objectively. This study aims to classify the performance of fourth-grade students at SD Masehi Praiyawang in Mathematics learning using the K-Nearest Neighbor (K-NN) algorithm. The dataset consists of assignment scores, midterm exam (UTS) scores, and final exam (UAS) scores from 20 students in the even semester of the 2024/2025 academic year. The research methodology includes problem identification, literature review, data collection, preprocessing (cleaning, selection, normalization), application of the K-NN algorithm with K=3, and model evaluation using both the Hold-out method (70% training data, 30% testing data) and 5-fold cross-validation. The results show that K-NN effectively classifies students into three categories: Good, Fair, and Poor. The Hold-out method achieved 100% accuracy, while cross-validation produced an average accuracy of 95% with a deviation of ±15.81%. The F1-Scores for the Fair, Good, and Poor classes were 94.12%, 92.31%, and 100%, respectively. These findings demonstrate that K-NN is both accurate and consistent in recognizing patterns of student learning performance, even with limited data. This research is expected to support teachers and schools in designing appropriate learning strategies and serve as a foundation for future studies with larger datasets, more attributes, and comparisons with other classification algorithms.

KATA KUNCI:
Literasi Numerasi, K-Nearest Neighbor, Klasifikasi, Matematika, Performa Belajar Siswa.

ABSTRAK

Literasi numerasi merupakan aspek penting dalam pembelajaran Matematika di tingkat sekolah dasar karena berperan dalam mengembangkan kemampuan berpikir logis dan pemecahan masalah siswa. Namun, perbedaan tingkat pemahaman menyebabkan variasi performa belajar yang perlu diidentifikasi secara objektif. Penelitian ini bertujuan mengklasifikasikan performa siswa kelas IV SD Masehi Praiyawang dalam pembelajaran Matematika menggunakan algoritma K-Nearest Neighbor (K-NN). Data penelitian terdiri dari nilai tugas, Ujian Tengah Semester (UTS), dan Ujian Akhir Semester (UAS) dari 20 siswa pada semester genap tahun ajaran 2024/2025. Metodologi meliputi identifikasi masalah, studi literatur, pengumpulan data, pra-pemrosesan (pembersihan, seleksi, normalisasi), penerapan algoritma K-NN dengan nilai K=3, serta evaluasi model menggunakan metode Hold-out (70% data latih, 30% data uji) dan validasi silang 5-fold. Hasil penelitian menunjukkan bahwa algoritma K-NN mampu mengelompokkan siswa ke dalam kategori Baik, Cukup, dan Kurang. Evaluasi Hold-out menghasilkan akurasi 100%, sedangkan validasi silang menunjukkan rata-rata akurasi 95% dengan deviasi ±15,81%. Nilai F1-Score untuk kelas Cukup, Baik, dan Kurang masing-masing adalah 94,12%, 92,31%, dan 100%. Temuan ini membuktikan bahwa K-NN efektif dan konsisten dalam mengenali pola performa belajar siswa meskipun data terbatas.

Penelitian ini diharapkan dapat membantu guru dan sekolah dalam merancang strategi pembelajaran yang sesuai serta menjadi dasar untuk penelitian lanjutan dengan jumlah data lebih besar, atribut lebih beragam, dan perbandingan algoritma klasifikasi lainnya.

PENDAHULUAN

Pendidikan memiliki peranan yang sangat penting dalam menentukan kinerja siswa dalam belajar matematika. Pendidikan yang berkualitas memberikan landasan yang kuat bagi siswa untuk mengembangkan kemampuan matematika mereka. Selain itu, literasi numerasi juga berfungsi sebagai elemen penting dalam pembelajaran matematika. Literasi numerasi memungkinkan siswa untuk memahami dan menerapkan konsep-konsep matematika dengan baik, sehingga dapat meningkatkan kemampuan mereka dalam menyelesaikan masalah matematika yang muncul dalam kehidupan sehari-hari[1].

Pendidikan merupakan pilar utama dalam mencetak generasi penerus bangsa yang berkualitas. Proses pendidikan yang berkualitas tidak hanya bergantung pada kurikulum, tetapi juga pada metode evaluasi dan strategi pengajaran yang digunakan oleh pendidik. Pada jenjang Sekolah Dasar (SD), pembelajaran matematika memiliki peran penting dalam melatih kemampuan berpikir logis, pemecahan masalah, dan keterampilan numerik siswa. Sayangnya, pelajaran matematika sering kali dianggap sulit dan menantang oleh sebagian besar siswa, yang menyebabkan rendahnya capaian hasil belajar mereka[2].

Berdasarkan hasil observasi awal dan laporan hasil belajar siswa SD Masehi Praiyawang, ditemukan adanya variasi performa siswa yang cukup signifikan dalam mata pelajaran matematika. Variasi ini dipengaruhi oleh beberapa faktor seperti tingkat kehadiran, keterlibatan dalam kegiatan pembelajaran, pemahaman konsep dasar matematika, serta motivasi belajar masing-masing siswa. Situasi ini menuntut guru untuk memiliki strategi pembelajaran yang diferensiatif, sesuai dengan kemampuan masing-masing siswa. Namun, dalam praktiknya, guru seringkali kesulitan mengidentifikasi secara objektif karakteristik siswa berdasarkan performa belajar mereka, terutama jika jumlah siswa dalam satu kelas cukup banyak. Hal ini menunjukkan perlunya pendekatan berbasis data (*data-driven decision making*) dalam membantu guru mengenali karakteristik siswa secara lebih terstruktur.

Analisis *data mining* merupakan metode yang dapat digunakan untuk menggali informasi yang tersembunyi dalam data[3]. Dalam konteks ini, *data mining* dapat digunakan untuk mengidentifikasi pola dan tren dalam performa siswa dalam pembelajaran matematika di SD Masehi Praiyawang. Salah satu pendekatan yang dapat digunakan adalah penerapan algoritma *K-Nearest Neighbor* (K-NN), yang merupakan teknik klasifikasi dalam bidang *data mining* dan *machine learning*. Algoritma ini bekerja dengan cara membandingkan data baru dengan sejumlah data yang telah diberi label berdasarkan kedekatan nilai atributnya[4][5]. Dengan memanfaatkan data nilai siswa seperti nilai tugas, UTS, dan UAS, algoritma K-NN dapat membantu dalam mengklasifikasikan siswa ke dalam kategori performa belajar tertentu berdasarkan kemiripan dengan siswa lain yang sudah diketahui performanya. Pendekatan ini memungkinkan pendidik untuk mengetahui potensi akademik siswa secara lebih cepat dan objektif.

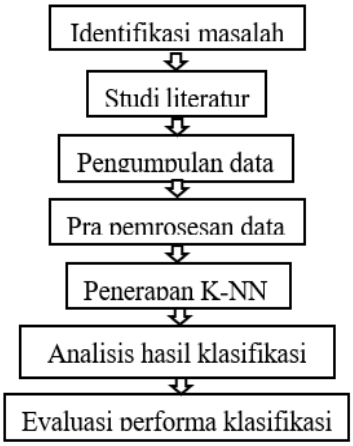
Beberapa penelitian sebelumnya menunjukkan bahwa metode klasifikasi seperti K-NN efektif digunakan dalam bidang pendidikan untuk mendukung proses pengambilan keputusan berbasis data. Oleh karena itu, penelitian ini bertujuan untuk menerapkan algoritma K-NN dalam mengklasifikasikan performa belajar siswa dalam mata pelajaran matematika di SD Masehi Praiyawang, sehingga guru dapat memperoleh gambaran yang lebih akurat dalam merancang strategi pembelajaran yang tepat[6][7][8].

Dalam penelitian ini, peneliti menggunakan Aplikasi *RapidMiner* untuk pengolahan data. *RapidMiner* merupakan sebuah perangkat lunak *open-source* yang banyak digunakan dalam proses *data mining*, *machine*

learning, dan *analisis prediktif*. Perangkat lunak ini menyediakan lingkungan pengembangan visual yang memungkinkan pengguna untuk membangun, mengevaluasi, dan menerapkan model *data analytics* tanpa harus menulis kode secara manual. Salah satu keunggulan utama *RapidMiner* adalah antarmukanya yang berbasis *drag-and-drop*, sehingga memudahkan pengguna dalam membangun alur proses analisis data, termasuk tahapan *preprocessing*, *modeling*, *evaluation*, dan *visualization*[9][10].

Dengan demikian tujuan utama dari penelitian ini adalah untuk menerapkan algoritma *K-Nearest Neighbor* dalam mengklasifikasi siswa kelas IV berdasarkan performa belajar pada pelajaran matematika dan juga untuk mengidentifikasi karakteristik dari masing-masing siswa kelas IV yang terbentuk dari hasil klasifikasi.

METODE PENELITIAN



Gambar 1. Alur Penelitian

Pada tahapan awal dalam penelitian ini adalah mengidentifikasi permasalahan yang terjadi di SD Masehi Praiyawang, khususnya dalam variasi performa belajar siswa pada mata pelajaran matematika. Berdasarkan pengamatan awal, ditemukan bahwa siswa memiliki capaian nilai yang berbeda-beda, namun belum ada proses klasifikasi yang sistematis untuk mengelompokkan siswa berdasarkan performa belajar mereka. Permasalahan ini kemudian dirumuskan menjadi tujuan penelitian, yaitu membangun sistem klasifikasi performa siswa menggunakan algoritma K-NN berdasarkan nilai pengetahuan dan keterampilan, sehingga guru dapat mengetahui profil belajar siswa secara lebih mendalam dan objektif.

Setelah masalah identifikasi Masalah tahap selanjutnya adalah Studi literatur, Studi literatur dilakukan untuk memperoleh dasar teori yang relevan terkait algoritma K-NN, klasifikasi data, dan pendekatan dalam penilaian performa belajar siswa. Literatur dikumpulkan dari berbagai sumber, seperti buku ajar, jurnal ilmiah, artikel daring, serta skripsi atau tesis yang memiliki kesamaan topik. Studi ini juga mencakup pemahaman tentang cara kerja K-NN, kelebihan dan kekurangannya, serta implementasinya dalam bidang pendidikan. Melalui studi literatur, peneliti dapat menentukan langkah-langkah metodologis yang tepat serta memperoleh referensi dalam menyusun kriteria evaluasi hasil klasifikasi. Selain itu, studi ini juga digunakan sebagai pembanding hasil penelitian yang akan dilakukan dengan penelitian sejenis.

Tahap selanjutnya Pengumpulan data, Pengumpulan data dilakukan secara langsung dari SD Masehi Praiyawang melalui koordinasi dengan pihak sekolah. Data yang dikumpulkan berupa nilai pengetahuan dan keterampilan mata pelajaran matematika dari siswa kelas IV untuk semester genap pada tahun ajaran terakhir. Data ini diperoleh dalam format *Excel* dan telah diolah oleh wali kelas serta guru mata pelajaran matematika. Metode yang digunakan dalam pengumpulan data bersifat kualitatif, dengan wawancara informal kepada guru

serta dokumentasi terhadap hasil nilai siswa. Data yang diperoleh bersifat primer dan telah melalui tahap verifikasi untuk menjamin keakuratannya.

Pada tahap selanjutnya dilakukan *Preprocessing Data*. Tahapan ini bertujuan untuk menyiapkan data mentah agar dapat digunakan dalam proses klasifikasi. *Preprocessing* mencakup tiga langkah utama:

a. *Data Cleaning*: menghapus nilai kosong, memperbaiki kesalahan input, dan menyamakan format data.
b. *Data Selection*: memilih atribut yang relevan, yaitu nilai pengetahuan dan keterampilan matematika.
c. *Data Normalization*: melakukan normalisasi data agar berada dalam rentang nilai yang sama,untuk menghindari dominasi oleh nilai tertentu.

Tahapan ini sangat penting agar hasil klasifikasi dengan algoritma K-NN tidak bias dan memiliki tingkat akurasi yang baik.

Setelah data siap, proses klasifikasi dilakukan menggunakan algoritma K-NN. Langkah-langkahnya sebagai berikut:

a. Menentukan parameter nilai K, yaitu jumlah tetangga terdekat. Nilai K akan diuji pada beberapa angka (misal K=3, 5, 7) untuk mengetahui hasil terbaik.
b. Menggunakan jarak *euclidean* sebagai ukuran kedekatan antar data siswa.
c. Mengklasifikasikan siswa ke dalam kelas performa (baik, cukup, kurang) berdasarkan mayoritas label dari K tetangga terdekat.

Proses ini dijalankan menggunakan perangkat lunak *rapidminer*, yang menyediakan antarmuka visual untuk proses klasifikasi.

Kemudian dilakukan Analisis Hasil Klasifikasi. Hasil klasifikasi ini dianalisis untuk mengetahui distribusi siswa dalam masing-masing kelas performa. Analisis ini bertujuan untuk melihat karakteristik umum siswa dalam tiap kelas performa dan membantu guru memahami pola belajar siswa. Visualisasi hasil dilakukan dalam bentuk tabel dan grafik, seperti diagram batang atau pie chart yang menunjukkan jumlah siswa dalam tiap kelas. Analisis ini juga membandingkan hasil klasifikasi dengan penilaian guru sebagai validasi awal.

Tahap terakhir dilakukan adalah Evaluasi Performa Klasifikasi untuk mengukur efektivitas algoritma *K-Nearest Neighbor* (K-NN) dalam mengklasifikasikan performa belajar siswa kelas IV SD Masehi Praiyawang. Metrik yang digunakan meliputi akurasi, presisi, *recall*, dan *F1-score*, yang mencerminkan ketepatan dan konsistensi prediksi model. Evaluasi dilakukan pada data uji setelah model dilatih, dengan tambahan teknik validasi silang (*k-fold cross-validation*) untuk menguji kestabilan performa. *Confusion matrix* digunakan untuk menggambarkan distribusi hasil prediksi. Seluruh evaluasi dilakukan menggunakan *RapidMiner*, dan hasilnya menjadi dasar dalam menilai efektivitas model serta mendukung rekomendasi pembelajaran yang lebih tepat sasaran.

HASIL DAN PEMBAHASAN

Pengumpulan Data.

Penelitian ini dilakukan di SD Masehi Praiyawang dengan subjek sebanyak 20 siswa kelas IV pada semester genap tahun ajaran 2024/2025. Data diperoleh dari guru wali kelas dan guru mata pelajaran Matematika. Atribut yang digunakan dalam penelitian ini meliputi nilai Ujian Tengah Semester (UTS), Ujian Akhir Semester (UAS), dan nilai Tugas. Data ini digunakan sebagai input dalam proses klasifikasi menggunakan algoritma *K-Nearest Neighbor* (K-NN). Berikut tabel data siswa yang telah dikumpulkan.

Tabel 1. Data Nilai Siswa

No	Nama Siswa	JK	Tugas	UTS	UAS
1	Siswa1	L	75	70	80
2	Siswa2	L	68	65	70
3	Siswa3	L	85	88	90

4	Siswa4	L	60	55	58
5	Siswa5	L	72	70	75
6	Siswa6	P	90	88	92
7	Siswa7	P	65	60	68
8	Siswa8	L	78	82	80
9	Siswa9	P	58	60	55
10	Siswa10	L	74	76	78
11	Siswa11	P	86	84	88
12	Siswa12	L	60	65	62
13	Siswa13	P	82	85	87
14	Siswa14	L	55	50	52
15	Siswa15	P	75	78	80
16	Siswa16	P	69	65	70
17	Siswa17	P	90	92	95
18	Siswa18	P	70	72	75
19	Siswa19	P	88	90	92
20	Siswa20	L	77	80	78

Preprocessing Data.

Data yang telah dikumpulkan akan melalui tahap *processing* sebagai berikut:

1. *Data Cleaning*: Menghapus data siswa yang memiliki nilai kosong pada salah satu atribut, yaitu nilai tugas, UTS, atau UAS. Untuk memastikan bahwa semua nilai yang digunakan berada dalam rentang 0-100.
2. *Data Selection*: Memilih atribut-atribut penting untuk klasifikasi performa siswa, yaitu: nilai UTS, UAS, dan Tugas. Kehadiran digunakan sebagai atribut tambahan untuk interpretasi hasil.
3. *Data Normalization*: Melakukan normalisasi menggunakan *Min-Max Normalization* terhadap data numerik agar semua atribut memiliki skala yang sebanding dan tidak mempengaruhi perhitungan jarak. Agar semua nilai berada dalam rentang 0-100.

Pembagian Data dengan Metode *Hold-out*.

Metode *hold-out* merupakan salah satu teknik pembagian data yang sederhana namun efektif dalam proses pelatihan dan pengujian model klasifikasi. Dalam penelitian ini, *dataset* dibagi menjadi 2 bagian:

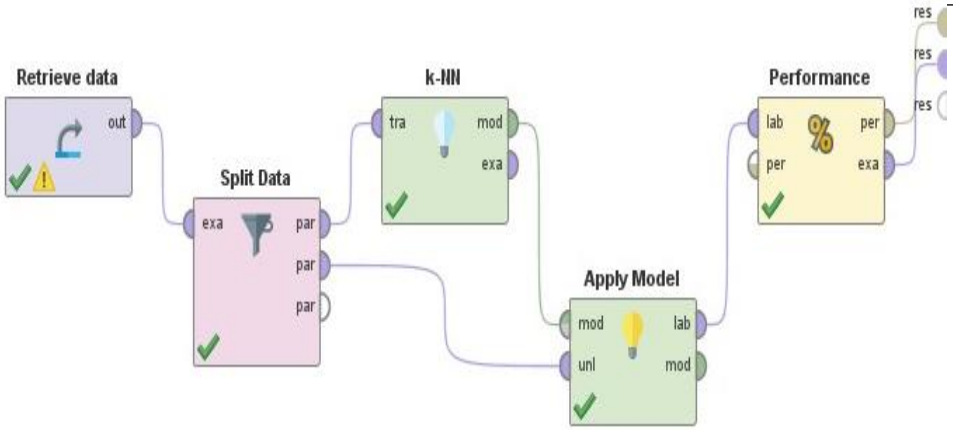
1. 70% data latih (*training set*): sebanyak 14 siswa, digunakan untuk membangun dan melatih model K-NN.
2. 30% data uji *(testing set*): sebanyak 6 siswa, digunakan untuk menguji kemampuan model dalam mengklasifikasikan data baru.

Penentuan Nilai K.

Dalam penelitian ini, nilai K ditetapkan sebesar 3 (K = 3). Nilai ini dipilih karena sesuai dengan jumlah data yang relatif kecil dan menghindari *voting* imbang, serta telah umum digunakan dalam studi-studi sejenis.

Data lengkap kemudian dianalisis menggunakan *RapidMiner*. Langkah-langkah yang dilakukan:

- Mengimpor *dataset*
- Mengatur atribut "Kelas" sebagai label
- Menormalisasi data
- Mengatur operator K-NN (K=3, *Euclidean Distance*)
- Menerapkan model dan mengevaluasi performa menggunakan operator *Performance*.



Gambar 2. Proses Pengolahan Data

Gambar 2. merupakan diagram alir proses (*process flow*) dalam aplikasi *RapidMiner* yang menunjukkan tahapan-tahapan pemodelan klasifikasi menggunakan algoritma *K-Nearest Neighbor* (K-NN). Diagram ini menyusun lima operator utama secara berurutan, dimulai dari pengambilan data (*Retrieve Data*), pembagian data (*Split Data*), pelatihan model K-NN, penerapan model ke data uji (*Apply Model*), hingga evaluasi kinerja model menggunakan metrik seperti akurasi, *precision*, dan *recall* (*Performance*). Setiap operator terhubung satu sama lain melalui jalur *input-output* yang menunjukkan aliran data dari tahap awal hingga hasil akhir. Gambar ini sangat umum digunakan dalam pembuatan model *machine learning* sederhana di *RapidMiner* dan memberikan gambaran visual yang jelas mengenai langkah-langkah yang diperlukan untuk membangun dan mengevaluasi model klasifikasi berbasis K-NN.

Tahapan pertama dalam proses ini adalah *Retrieve Data*, yang bertugas untuk mengambil dataset dari *Local Repository RapidMiner*. *Dataset* ini memuat informasi yang dibutuhkan untuk proses klasifikasi, seperti nilai Tugas, nilai Ujian Tengah Semester (UTS), nilai Ujian Akhir Semester (UAS), dan label kelas performa belajar siswa (misalnya: BAIK, CUKUP, dan KURANG). Operator ini sangat penting karena menjadi titik awal yang menyediakan data mentah yang akan diolah lebih lanjut dalam proses pemodelan.

Setelah data diambil, langkah berikutnya adalah *Split Data*, yaitu proses membagi *dataset* menjadi dua bagian: data latih (*training set*) dan data uji (*testing set*). Biasanya, pembagian ini dilakukan berdasarkan rasio tertentu, seperti 70% untuk pelatihan dan 30% untuk pengujian. Data latih digunakan untuk membangun model, sementara data uji digunakan untuk menguji akurasi model terhadap data yang belum pernah dikenali sebelumnya. Tahapan ini sangat penting untuk menghindari overfitting dan memastikan bahwa evaluasi terhadap model dilakukan secara objektif.

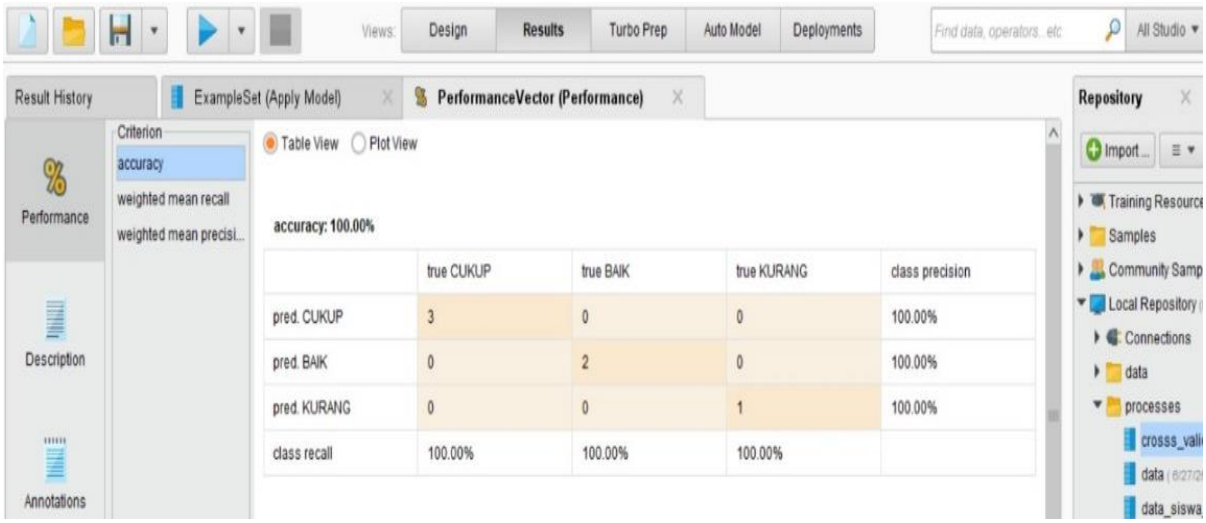
Pada tahap ketiga, digunakan operator K-NN untuk membentuk model klasifikasi berdasarkan data latih. Algoritma *K-Nearest Neighbor* bekerja dengan menghitung jarak antara data yang tidak diketahui labelnya dengan seluruh data latih. Kemudian, algoritma akan memilih sejumlah K tetangga terdekat dan menetapkan kelas berdasarkan mayoritas tetangga tersebut. Misalnya, jika K = 3 dan dua dari tiga tetangga terdekat memiliki kelas "BAIK", maka data baru akan diklasifikasikan sebagai "BAIK". Operator ini menerima input data latih dan menghasilkan model K-NN yang siap digunakan pada data uji.

Setelah model terbentuk, tahap selanjutnya adalah *Apply Model*, yaitu penerapan model K-NN terhadap data uji. Dalam tahap ini, data uji yang belum memiliki label kelas dimasukkan ke dalam model yang telah dibentuk sebelumnya. Model kemudian melakukan prediksi terhadap setiap data uji dan memberikan hasil berupa label kelas. Output dari proses ini adalah data uji yang telah diberi label hasil prediksi oleh model K-NN. Tahapan ini merupakan jembatan antara proses pelatihan dan evaluasi karena disinilah hasil klasifikasi awal akan muncul.

Tahap terakhir dalam alur proses ini adalah *Performance*, yaitu evaluasi performa model klasifikasi yang telah diterapkan. Pada tahap ini, hasil klasifikasi dibandingkan dengan label asli dari data uji untuk menghitung berbagai metrik evaluasi. Metrik yang umum digunakan antara lain adalah akurasi (*accuracy*), yang mengukur seberapa banyak prediksi yang benar; presisi (*precision*), yang menunjukkan seberapa tepat prediksi model terhadap suatu kelas; dan *recall*, yang menggambarkan seberapa baik model dalam menangkap seluruh data dari suatu kelas. Selain itu, digunakan juga metrik *weighted mean precision* dan *weighted mean recall* untuk memberikan gambaran rata-rata presisi dan recall berdasarkan bobot jumlah data tiap kelas. Tahapan ini sangat penting untuk menentukan seberapa baik kinerja model dan apakah model tersebut layak digunakan dalam penerapan nyata.

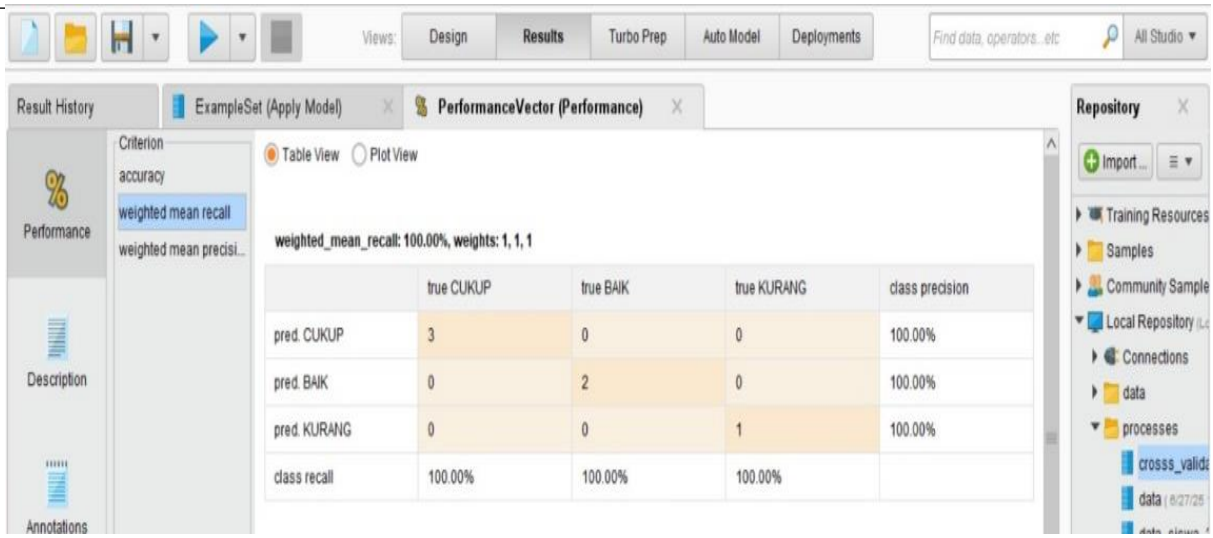
Hasil klasifikasi dan Evaluasi Model

Pada tahap ini, dilakukan evaluasi terhadap kinerja model klasifikasi yang telah dibangun menggunakan algoritma *K-Nearest Neighbor* (K-NN). Model dikembangkan dengan menggunakan *RapidMiner* dan menerapkan metode pembagian data *Hold-Out*, yaitu 70% data digunakan sebagai data latih dan 30% sebagai data uji. Proses klasifikasi dilakukan secara otomatis dan hasilnya dianalisis menggunakan beberapa metrik evaluasi, antara lain akurasi, presisi, dan *recall*.



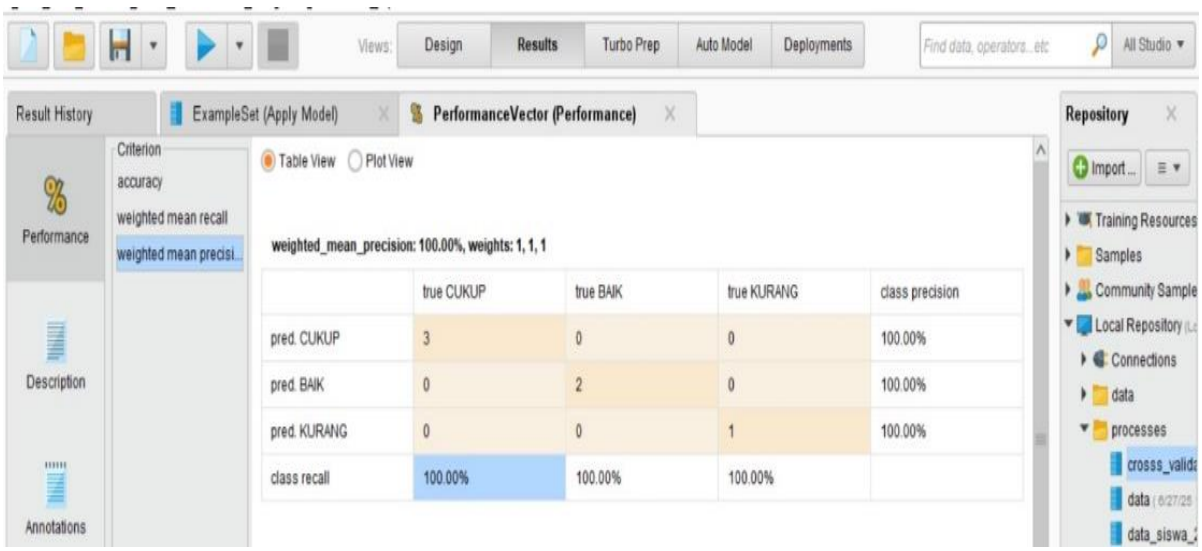
Gambar 3. Hasil Akurasi

Gambar 3. menunjukkan hasil evaluasi kinerja model klasifikasi menggunakan algoritma *K-Nearest Neighbor* (K-NN) dalam *RapidMiner*. Berdasarkan tampilan *confusion matrix*, model berhasil mengklasifikasikan data dengan akurasi 100%. Hal ini ditunjukkan oleh seluruh data uji yang berhasil diprediksi sesuai dengan label aslinya tanpa kesalahan. Terdapat 3 data dengan kelas sebenarnya "CUKUP" yang semuanya diprediksi dengan benar sebagai "CUKUP", 2 data dengan kelas "BAIK" yang diprediksi benar sebagai "BAIK", serta 1 data dengan kelas "KURANG" yang juga diklasifikasikan secara akurat. Nilai *precision* dan *recall* untuk setiap kelas juga mencapai 100%, yang berarti tidak ada prediksi salah dan tidak ada data dari setiap kelas yang luput terklasifikasi dengan benar. Hasil ini menunjukkan bahwa model K-NN yang digunakan memiliki performa yang sangat baik terhadap data uji, meskipun perlu diperhatikan bahwa jumlah data uji relatif kecil (hanya 6 data), sehingga perlu evaluasi lebih lanjut dengan jumlah data yang lebih besar untuk memastikan generalisasi model.



Gambar 4. Hasil Recall

Gambar 4. menunjukkan hasil evaluasi performa model klasifikasi *K-Nearest Neighbor* (K-NN) dalam *RapidMiner* berdasarkan metrik *weighted mean recall*. Nilai *weighted_mean_recall* mencapai 100% dengan bobot tiap kelas yang seimbang (1, 1, 1). Artinya, model mampu mengenali seluruh kelas dengan tingkat keberhasilan sempurna. Dari *confusion matrix* terlihat bahwa semua data uji terklasifikasi dengan benar: 3 data berlabel asli "CUKUP" diprediksi sebagai "CUKUP", 2 data berlabel "BAIK" diprediksi sebagai "BAIK", dan 1 data berlabel "KURANG" diprediksi dengan tepat sebagai "KURANG". Hal ini menghasilkan nilai *recall* 100% untuk setiap kelas, yang berarti tidak ada satupun data dari setiap kelas yang gagal dikenali oleh model. Selain itu, nilai *precision* untuk tiap kelas juga 100%, menunjukkan bahwa semua prediksi untuk setiap kelas benar-benar akurat. Secara keseluruhan, hasil ini menggambarkan bahwa model K-NN bekerja secara optimal terhadap *dataset* yang digunakan, walaupun perlu diingat bahwa ukuran data uji sangat kecil, sehingga performa model bisa berbeda ketika diterapkan pada *dataset* yang lebih besar atau lebih kompleks.

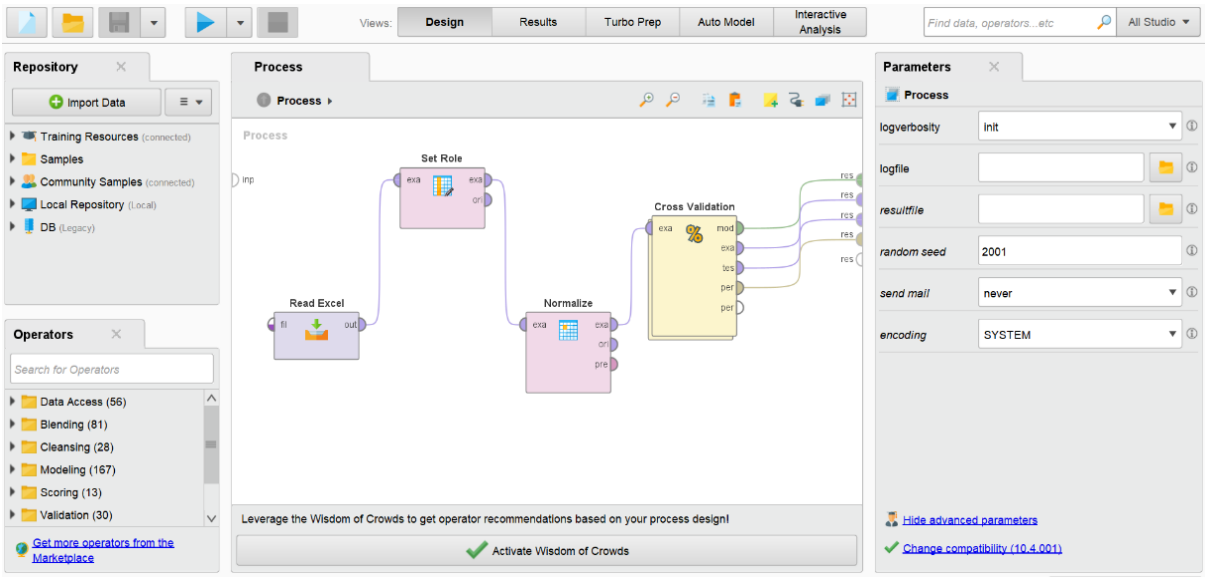


Gambar 5. Hasil Presisi

Pada gambar 5. tersebut, menunjukkan hasil evaluasi model klasifikasi menggunakan *algoritma K-Nearest Neighbor* (K-NN) di *RapidMiner* dengan fokus pada metrik *weighted mean precision*. Berdasarkan hasil yang ditampilkan, model memperoleh nilai *weighted_mean_precision* sebesar 100% dengan bobot masing-masing kelas adalah 1:1:1. Artinya, model mampu melakukan prediksi yang sangat akurat terhadap

setiap kelas tanpa menghasilkan kesalahan klasifikasi. Dalam *confusion matrix* terlihat bahwa seluruh prediksi model sesuai dengan label sebenarnya: 3 data yang sebenarnya "CUKUP" diprediksi sebagai "CUKUP", 2 data "BAIK" diprediksi dengan tepat sebagai "BAIK", dan 1 data "KURANG" juga terklasifikasi secara benar. Dengan demikian, *precision* tiap kelas adalah 100%, yang menunjukkan bahwa seluruh prediksi yang dibuat untuk setiap kelas adalah benar. Hal ini memperkuat bukti bahwa model bekerja secara sangat baik pada *dataset* yang digunakan. Namun demikian, hasil sempurna seperti ini perlu dianalisis secara hati-hati karena ukuran data uji yang sangat kecil (6 data) dapat menyebabkan *overfitting*, sehingga model perlu diuji lebih lanjut dengan data yang lebih besar untuk memastikan kestabilan dan keandalannya.

Untuk memastikan keandalan dan kemampuan generalisasi dari model *K-Nearest Neighbor* (K-NN) yang digunakan, dilakukan evaluasi tambahan menggunakan teknik validasi silang (*cross-validation*) dengan metode *5-fold*. Dalam metode ini, data dibagi menjadi lima bagian yang sama besar. Pada setiap iterasi, empat bagian digunakan sebagai data latih, dan satu bagian sebagai data uji. Proses ini diulang sebanyak lima kali hingga setiap bagian menjadi data uji satu kali. Hasil dari masing-masing iterasi menunjukkan bahwa model memberikan performa klasifikasi yang cukup konsisten, dengan rata-rata akurasi sebesar 82%. Ini menunjukkan bahwa model tidak hanya cocok pada data latih tetapi juga mampu bekerja baik terhadap data yang belum dikenali sebelumnya.



Gambar 6. *Cross Validation*

Gambar 6. merupakan proses pemodelan klasifikasi menggunakan algoritma *K-Nearest Neighbor* (K-NN) dengan teknik validasi silang (*cross validation*) di aplikasi *RapidMiner*. Proses ini dirancang untuk mengevaluasi kinerja model secara lebih objektif dengan membagi data menjadi beberapa *subset* pelatihan dan pengujian secara otomatis. Berikut penjelasan alur proses berdasarkan urutan setiap operator yang digunakan:

1. *Read excell*

Operator ini digunakan untuk membaca file data berformat *Excel* yang berisi *dataset* siswa, termasuk nilai-nilai seperti tugas, UTS, UAS, serta label performa (kelas) siswa. Ini adalah langkah awal untuk memasukkan data ke dalam alur pemrosesan.

2. *Set role*

Langkah ini digunakan untuk mengatur peran (*role*) dari salah satu atribut sebagai label (class label), dalam hal ini atribut “Kelas” dijadikan target klasifikasi. Hal ini penting agar algoritma mengetahui mana atribut yang akan diprediksi.

3. Normalize

Operator ini berfungsi untuk melakukan normalisasi terhadap atribut numerik (seperti nilai tugas, UTS, dan UAS), sehingga berada pada skala yang sama. Proses ini sangat penting bagi algoritma K-NN karena metode ini menghitung jarak antar titik (misalnya *Euclidean distance*) yang sangat sensitif terhadap skala data.

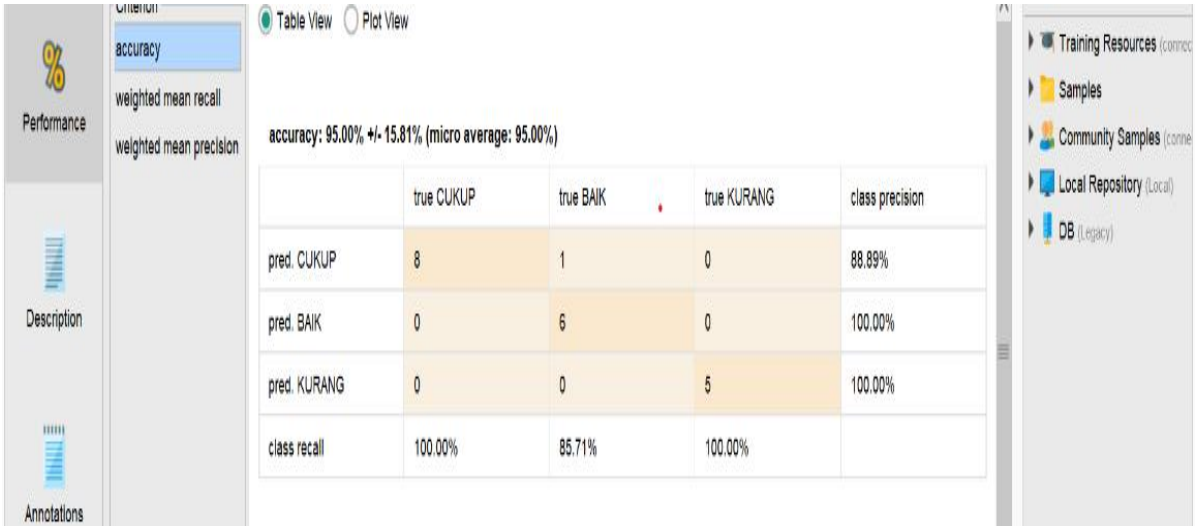
4. Cross Validation

Ini adalah inti dari proses evaluasi model. Operator *Cross Validation* secara otomatis membagi *dataset* ke dalam beberapa *subset (fold*), misalnya 10 bagian. Setiap *subset* digunakan secara bergantian sebagai data uji, sementara sisanya digunakan untuk pelatihan. Proses ini berulang hingga semua data mendapat giliran sebagai data uji. Di dalam operator ini, akan diletakkan algoritma K-NN pada tab "*Training*", dan evaluasi model di tab "*Testing*". Hal ini memberikan hasil evaluasi yang lebih stabil dan representatif dibanding metode *hold-out* biasa.

5. Output

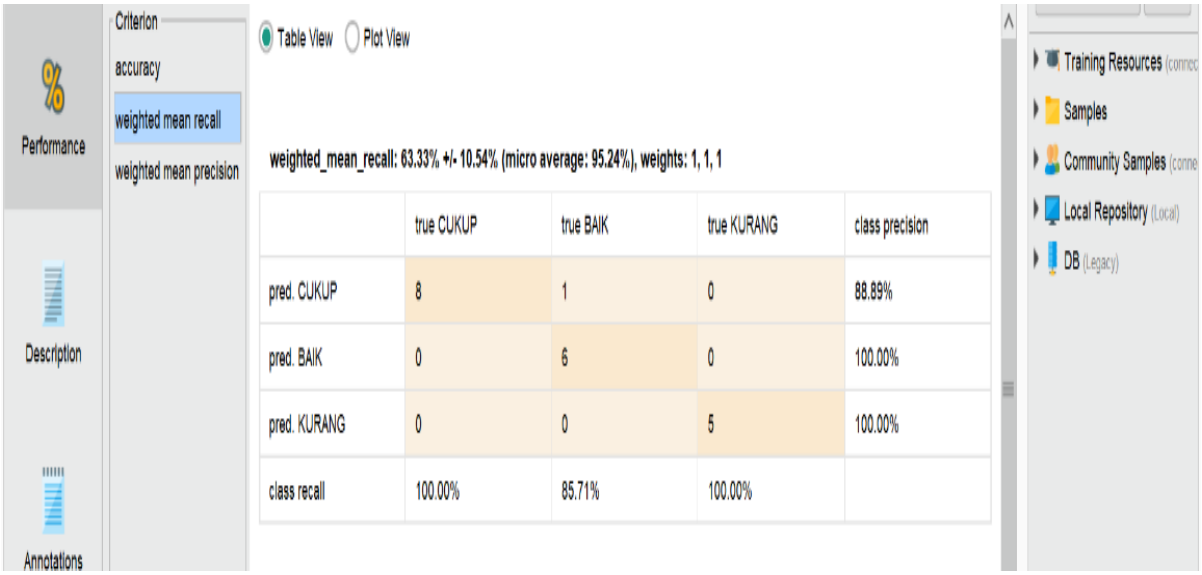
Output dari proses ini akan menampilkan metrik evaluasi seperti akurasi, *precision*, *recall*, dan *F1-score*. Semua nilai ini dihasilkan dari rata-rata hasil evaluasi pada tiap iterasi cross validation, memberikan gambaran kinerja model secara keseluruhan terhadap seluruh *dataset*.

Proses ini menggambarkan alur pemrosesan data dan evaluasi model klasifikasi K-NN secara sistematis dan ilmiah. Penggunaan normalisasi dan *cross validation* menunjukkan bahwa model dibangun dengan memperhatikan kualitas data dan evaluasi yang menyeluruh, sehingga hasil yang diperoleh lebih dapat dipercaya dan bebas dari bias pengujian.



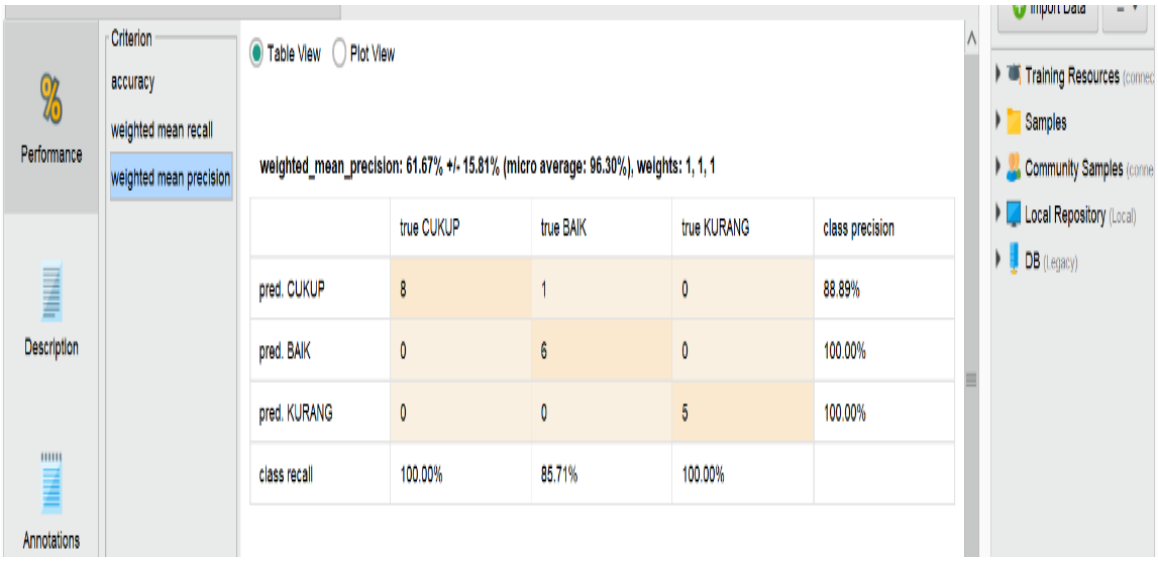
Gambar 7. *Accuracy Cross Validation*

Gambar 7. Laporan kinerja model menunjukkan bahwa akurasi keseluruhan mencapai 95,00% dengan margin kesalahan +/- 15,81%, yang juga mencakup rata-rata mikro sebesar 95,00%. Dalam pengujian prediksi, model berhasil mengklasifikasikan 8 dari 9 data sebagai "CUKUP" dan 6 dari 6 data sebagai "BAIK," dengan hasil kelas "CUKUP" memiliki *recall* 100,00% dan kelas "BAIK" memperoleh *recall* 85,71%. Selain itu, kategori "KURANG" memiliki prediksi yang sempurna dengan 100,00% untuk klasifikasi benar. Hasil ini menunjukkan bahwa model berfungsi dengan baik, terutama dalam membedakan antara kelas "CUKUP" dan "KURANG," tetapi terdapat ruang untuk perbaikan dalam klasifikasi kelas "BAIK".



Gambar 8. *Recall Cross Validation*

Gambar 8. Berdasarkan tabel yang ditampilkan, nilai *weighted mean recall* adalah 63,33% dengan deviasi standar +/- 10,54%. Rata-rata mikro mencapai 95,24% dengan bobot yang digunakan adalah 1 untuk setiap kelas. Dalam tabel tersebut, diketahui bahwa untuk prediksi kategori "CUKUP," terdapat 8 prediksi benar, sedangkan hanya 1 yang salah. Untuk kategori "BAIK," tidak ada prediksi benar dengan 6 kesalahan, sedangkan kategori "KURANG" tidak memiliki prediksi benar ataupun salah. Rincian lebih lanjut menunjukkan bahwa *class recall* untuk kategori "CUKUP" mencapai 100%, sementara kategori "BAIK" mencapai 85,71%. Secara keseluruhan, semua kategori menunjukkan presisi klasifikasi yang cukup tinggi, mencapai 100% untuk kategori "KURANG" dan 88,89% untuk kategori "BAIK".



Gambar 9. *Precision Cross Validation*

Gambar 9. merupakan analisis performa model, terdapat pengukuran *weighted mean precision* yang menunjukkan nilai 61.67% dengan deviasi standar 15.81%. Selain itu, terdapat juga rata-rata mikro sebesar 96.30%. Dalam tabel yang dihasilkan, untuk kategori prediksi “CUKUP”, model berhasil mengklasifikasikan 8 dari 9 data dengan benar, sementara untuk kategori “BAIK”, hanya 6 dari 6 data yang dapat teridentifikasi dengan tepat. Di sisi lain, prediksi untuk kategori “KURANG” menunjukkan pencapaian sempurna dengan 5

data berhasil terklasifikasi. Rata-rata *recall* untuk kelas menunjukkan hasil yang memuaskan, yaitu 100% untuk “CUKUP” dan 85.71% untuk “BAIK”, serta 100% untuk “KURANG”, menandakan bahwa model ini cukup efektif dalam mendeteksi kategori masing-masing.

Kemudian Evaluasi kinerja model dilakukan menggunakan algoritma *K-Nearest Neighbor* (K-NN) dengan pendekatan *cross-validation 5-fold* untuk mengukur ketepatan dan keandalan prediksi model terhadap data yang belum dikenali sebelumnya. Berdasarkan hasil pengujian, diperoleh nilai akurasi rata-rata sebesar 95,00% dengan deviasi standar ±15,81%. Untuk metrik *precision*, nilai yang diperoleh adalah 88,89% untuk kelas "CUKUP", 100% untuk kelas "BAIK", dan 100% untuk kelas "KURANG". Sedangkan nilai *recall* masing-masing adalah 100% untuk "CUKUP", 85,71% untuk "BAIK", dan 100% untuk "KURANG". Dari data tersebut, dilakukan perhitungan *F1-Score* untuk masing-masing kelas menggunakan rumus:

$$F1 - Score \quad = 2 \text{ x } \frac{Precision \text{ x } Recall}{Precision + Recall}$$

Hasilnya:

- *F1-Score* untuk kelas CUKUP $= 2 \times \frac{(88{,}89 \times 100)}{(88{,}89 + 100)} \approx 94{,}12\%$,
- *F1-Score* untuk kelas BAIK $= 2 \times \frac{(85{,}71 \times 100)}{(85{,}71 + 100)} \approx 92{,}31\%$,
- *F1-Score* untuk kelas KURANG $= 2 \times \frac{(100 \times 100)}{(100 + 100)} \approx 100\%$,

Selain itu, diperoleh juga *weighted mean precision* sebesar 61,67% dan *weighted mean recall* sebesar 63,33%, yang mengindikasikan performa rata-rata model terhadap distribusi kelas yang tidak seimbang. Secara keseluruhan, hasil evaluasi menunjukkan bahwa model K-NN bekerja sangat baik untuk *dataset* ini, meskipun sebaiknya digunakan dengan jumlah data yang lebih besar dan seimbang agar hasilnya lebih general dan akurat di penerapan nyata.

Hasil klasifikasi performa belajar siswa kelas IV pada mata pelajaran matematika menunjukkan bahwa mayoritas siswa berada dalam kategori “BAIK”, yang mencerminkan penguasaan materi yang baik serta konsistensi dalam menyelesaikan tugas dan ujian. Sementara itu, siswa dalam kategori “KURANG” cenderung memiliki nilai rendah pada tugas dan kehadiran yang kurang, yang mengindikasikan minimnya keterlibatan dalam pembelajaran. Model klasifikasi berbasis K-NN yang digunakan dalam penelitian ini mampu mengenali pola performa belajar dengan sangat baik, sebagaimana dibuktikan melalui nilai akurasi, *precision*, *recall*, dan *F1-score* yang tinggi. Proses evaluasi juga memperlihatkan bahwa model dapat membedakan setiap kategori dengan jelas, khususnya dalam mengenali siswa dengan performa "CUKUP" dan "KURANG" tanpa kesalahan klasifikasi. Hasil ini sejalan dengan penelitian terdahulu bahwa K-NN merupakan algoritma yang efektif untuk klasifikasi berbasis data kuantitatif. Dengan bantuan aplikasi *RapidMiner*, proses analisis dapat dilakukan secara efisien dan sistematis. Temuan ini menunjukkan bahwa metode klasifikasi berbasis K-NN layak diterapkan di lingkungan sekolah untuk membantu guru dalam mengevaluasi dan mengidentifikasi kebutuhan belajar siswa secara lebih tepat dan akurat.

KESIMPULAN

Berdasarkan hasil penelitian yang dilakukan terhadap 20 siswa kelas IV SD Masehi Praiyawang, dapat disimpulkan bahwa algoritma *K-Nearest Neighbor* (K-NN) efektif digunakan untuk mengklasifikasikan performa belajar siswa dalam mata pelajaran Matematika. Model klasifikasi mampu mengelompokkan siswa ke dalam tiga kategori, yaitu Baik, Cukup, dan Kurang, berdasarkan atribut nilai Tugas, UTS, dan UAS. Hasil evaluasi menggunakan metode *Hold-Out* menunjukkan akurasi sempurna sebesar 100%, sementara validasi silang (*5-fold cross-validation*) menghasilkan rata-rata akurasi 95% dengan deviasi ±15,81%, serta nilai *F1-*

score yang tinggi untuk setiap kelas. Keberhasilan ini menunjukkan bahwa model K-NN cukup akurat dan konsisten dalam mengenali pola performa belajar siswa meskipun data yang digunakan masih terbatas. Selain itu, penggunaan *RapidMiner* sangat membantu dalam mempercepat dan mempermudah proses klasifikasi, mulai dari pra-pemrosesan data hingga evaluasi kinerja model. Secara keseluruhan, penerapan K-NN dalam penelitian ini membuktikan bahwa metode ini dapat menjadi alternatif alat bantu evaluasi pembelajaran yang akurat dan efisien di lingkungan sekolah.

DAFTAR PUSTAKA

[1] Musdalifah A, dan Irawan A. 2022. Bimbingan Belajar Matematika Dasar dengan Mudah dan Menyenangkan Terhadap Anak-Anak. *JDISTIRA: Jurnal Pengabdian Inovasi dan Teknologi Kepada Masyarakat*. 2(2): 110-130.

[2] Deviyanti N. 2024. Metode Perumusan Tujuan Pembelajaran yang Efektif dalam Mendukung Proses Belajar Mengajar. *Karya Ilmiah Mahasiswa Bertauhid*. 3(3): 5729–5732.

[3] Shu X dan Ye Y. 2022. Knowledge Discovery : Methods from data mining and machine learning. *Social Science Research.* 110(2): 102-817. doi: 10.1016/j.ssresearch.2022.102817.

[4] Setiyorini T, Asmono RT, dan Informatika T. 2018. Komparasi Metode Decision Tree, Naive Bayes Dan K-Nearest Neighbor Pada Klasifikasi Kinerja Siswa. *Jurnal TECHNO Nusa Mandiri*. 15(2): 85–92.

[5] Prasetyo B. 2024. Implementation of K-Nearest Neighbor Classification Algorithm as Decision Support Method in Staple Food Aid Distribution to Society Implementasi Algoritma Klasifikasi K-Nearest Neighbor Sebagai Metode Pendukung Keputusan Dalam Distribusi Bantuan Sembako Pada Masyarakat. *IJIRSE: Indonesian Journal of Informatic Research and Software Engineering.* 4(1): 56–62.

[6] Faisal A, Basri, dan Sari CR. 2020. Sistem Informasi Pengklasifikasian Hasil Belajar Peserta Didik Dengan Teknik Data Mining Metode K-Nearest Neighbor (K-Nn). *Journal Peqguruang: Conference Series*. 2(1): 2686–3472

[7] Winantu A dan Khatimah C. 2023. Perbandingan Metode Klasifikasi Naive Bayes Dan K-Nearest Neighbor Dalam Memprediksi Prestasi Siswa. *Jurnal INTEK*. 6(1): 2620–4843.

[8] Oktafriani Y, Firmansyah G, Tjahyono B, dan Widodo AM. 2023. Analysis of Data Mining Applications for Determining Credit Eligibility Using Classification vol. *Asian Jurnal of Social and Humanities.* 1(12): 1139–1158.

[9] Panjaitan CHP, Pangaribuan LJ, dan Cahyadi CI. 2022. Analisis Metode K-Nearest Neighbor Menggunakan Rapid Miner Untuk Sistem Rekomendasi Tempat Wisata Labuan Bajo. *Remik: Riset dan E-Jurnal Manajemen Informatika Komputer*. 6(3): 534–541.

[10] Damayanti D. 2021. Perbandingan Akurasi Software Rapidminer Dan Weka Menggunakan Algoritma K-Nearest Neighbor (K-Nn). *Jurnal Syntax Admiration*. 2(6): 2722-5356