

Klasifikasi Prestasi Akademik Siswa Berdasarkan Nilai Rapor dan Kedisiplinan dengan Metode K-Nearest Neighbor

Afif Muhaimin¹, Mokhamad Amin Hariyadi², M. Imamudin³

^{1,2,3} Magister Informatika, Fakultas Sains dan Teknologi, Universitas Islam Malik Maulana Ibrahim, Malang

¹200605210008@student.uin-malang.ac.id, ²adyt2002@uin-malang.ac.id, ³imamudin@ti.uin-malang.ac.id

Abstrak— Sekolah melaksanakan kegiatan pembelajaran melalui tahapan-tahapan dan proses, agar peserta didik mencapai prestasi belajar yang baik. Prestasi belajar adalah hasil yang diperoleh berupa kesan-kesan yang mengakibatkan perubahan dalam diri individu sebagai hasil dari aktivitas. Sebagai pedoman penentuan siswa yang berprestasi adalah dengan memiliki nilai akhir setiap semester yang baik serta nilai kedisiplinan selama mengikuti pembelajaran disekolah, seperti tidak memiliki poin pelanggaran yang tinggi. Belum adanya metode khusus yang digunakan untuk mengklasifikasikan siswa berdasarkan prestasinya dan banyak kemiripan data, dibutuhkan metode klasifikasi yang tepat dan akurat, salah satunya menggunakan ilmu di bidang data mining. Dalam artikel ini peneliti ingin mengklasifikasi prestasi siswa dengan metode *K-Nearest Neighbor* (*K-NN*) berdasarkan nilai akademik dan nilai kedisiplinan siswa dengan menggunakan data berjumlah 348 siswa di SMA Negeri 2 Batu Jawa Timur. Hasil eksperimen dan evaluasi model yang dilakukan, dengan pembagian data training dan data testing secara acak dengan beberapa percobaan diperoleh nilai akurasi tertinggi sebesar 91.39 %.

Kata Kunci: Siswa, Prestasi, Data Mining, Klasifikasi, *K-Nearest Neighbor*

Abstract— Schools carry out learning activities through stages and processes so that students achieve good learning achievements. Learning achievement is the result obtained in the form of impressions that result in changes in the individual as a result of activities. A guideline for determining students who excel is having good final grades for each semester as well as discipline scores while studying at school, such as not having high points of violation. There is no specific method used to classify students based on their achievements or the many similarities in the data. A precise and accurate classification method is needed, one of which uses knowledge in the field of data mining. In this article, researchers want to classify student achievement using the *K-Nearest Neighbor* (*K-NN*) method based on students' academic and disciplinary scores using data totaling 348 students at SMA Negeri 2 Batu, East Java. The results of the experiments and model evaluation carried out by randomly dividing the training data and test data with several experiments obtained an average accuracy value of 91.39%.

Keywords: Students, Achievement, Data Mining, Classification, *K-Nearest Neighbor*

1. PENDAHULUAN

Pendidikan merupakan kewajiban bagi warga Indonesia untuk mendapatkan ilmu dan mewajibkan masyarakat belajar selama 9 tahun. Yang tercantum pada Pasal 1 Undang-undang Sistem Pendidikan Nasional tahun 2003 yaitu pendidikan adalah usaha sadar dan terencana untuk mewujudkan suasana belajar dan proses pembelajaran agar peserta didik secara aktif mengembangkan potensi dirinya untuk memiliki kekuatan spiritual keagamaan, pengendalian diri, kepribadian, kecerdasan, akhlak mulia, serta keterampilan yang diperlukan dirinya, masyarakat, bangsa dan Negara (Undang-Undang No. 20 tahun 2003 Tentang Sistem Pendidikan Nasional, 2003)

Sekolah melaksanakan kegiatan pembelajaran melalui tahapan-tahapan dan proses, agar peserta didik mencapai prestasi belajar yang baik [1]. Proses belajar akan menghasilkan perubahan pengetahuan dan keterampilan siswa. Melalui proses belajar pengetahuan dan keterampilan yang dimiliki siswa akan bertambah. Acuan yang menjadi penentuan siswa yang berprestasi adalah dengan memiliki nilai rata-rata rapor sekolah yang baik serta disiplin selama mengikuti pembelajaran disekolah tersebut seperti tidak memiliki poin pelanggaran yang tinggi. Penilaian siswa berprestasi pada melibatkan beberapa kriteria yang harus diambil karena ada tiga ranah kompetensi yang dinilai yaitu kompetensi sikap, kompetensi pengetahuan, dan kompetensi keterampilan [2]. Nilai rata-rata yang didapatkan siswa pada setiap semester dapat diklasifikasi untuk mendapatkan informasi dengan menambahkan daftar nilai poin kedisiplinan sebagai penambahan penilaian siswa, sehingga mendapatkan hasil akurasi dari data yang diuji.

Belum adanya metode khusus yang digunakan untuk mengklasifikasikan siswa berdasarkan prestasinya dan banyak kemiripan data, menyebabkan sering terjadi salah pengklasifikasian. Oleh karena itu dibutuhkan metode klasifikasi yang tepat dan akurat, salah satunya menggunakan ilmu di bidang data mining. Data mining merupakan proses penggalian informasi dan berguna dari set data besar yang melibatkan konsep interdisipliner yang relatif baru yang melibatkan analisis data dan penemuan pengetahuan dari database dan menggunakan pendekatan multi-sisi yang mencakup analisis statistik, visualisasi data, penemuan pengetahuan, pengenalan pola dan manajemen basis data [3].

Klasifikasi dalam data mining merupakan proses penempatan atau pemilihan atribut dan parameter yang tepat berdasarkan dataset yang akan diklasifikasikan dalam menentukan akurasi pada proses penghitungannya [4]. Salah satu metode teknik klasifikasi diantaranya adalah metode *K-Nearest Neighbor* bekerja berdasarkan asumsi bahwa suatu data akan memiliki kelas atau kategori yang sama dengan data yang berada disekitarnya. Konsep ini dikenal dengan konsep ketetanggaan [5].

Penelitian yang dilakukan oleh Denni Kurniawan dan Ade Saputra, tentang Penerapan *K-Nearest Neighbour* dalam Penerimaan Peserta Didik dengan Sistem Zonasi menyimpulkan bahwa penerapan metode *K-Nearest Neighbor* yang digunakan mampu sebagai acuan dalam menentukan peserta didik yang diterima berdasarkan zonasi dan non zonasi [6]. Dari percobaan perhitungan penentuan penerimaan peserta didik baru menggunakan perhitungan model *Confussion matrix* dan Rapidminer, Nilai akurasi yang diusulkan terletak pada K5 dengan tingkat akurasi sebesar 83,36%.

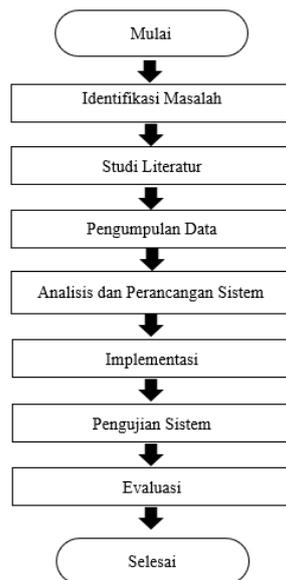
Esty Purwaningsih dan Ela Nurelasari dalam penelitiannya tentang penerapan *K-Nearest Neighbor* untuk klasifikasi tingkat kelulusan pada siswa [7]. Data yang digunakan dalam pengolahan adalah data pada tahun 2013 dan 2014 didapat dari salah satu instansi sekolah swasta di area Bekasi dengan jumlah siswa kelas VI sebanyak 104 siswa dengan atribut nilai PAS, Nilai US Teori, Nilai US Praktek, Nilai UTS, Nilai Prilaku Siswa. Pengujian Metode dengan 10-fold Cross Validation dihasilkan tingkat kelulusan siswa dengan metode KNN didapat rata-rata akurasi dengan nilai sebesar 96,49%.

Penelitian yang dilakukan oleh Verdi Alvian dkk, tentang klasifikasi siswa berprestasi menggunakan metode *K-Nearest Neighbor* (KNN) pada SMA Negeri 2 Situbondo [8] . Data nilai siswa sebanyak 182 akan displit menjadi data validasi sebanyak 32 dan unseen data sebanyak 150. Dimana data validasi akan masuk ke proses cross validation terlebih dahulu, dan unseen data akan langsung masuk ke proses *K-Nearest Neighbor* (KNN). Hasil akurasi tertinggi menggunakan data training pada 6 skenario dan data testing sebanyak 32 data, 11 didapatkan akurasi sebesar 94%, presisi 91% dan berdasarkan pengujian 32 unseen data terhadap data training dari 6 skenario, didapatkan akurasi terbaik sebesar 94%, presisi 91% dan recall 91% pada fold Kf = 5 skenario 5 dengan nilai k pada KNN = 7.

Tujuan dari penelitian ini untuk mengklasifikasi prestasi siswa berdasarkan nilai rapor dan kedisiplinan di SMA Negeri 2 Batu dengan menggunakan data berjumlah 348. Metode yang mendukung penelitian ini adalah metode *K-Nearest Neighbour* dengan menggunakan model *k-fold cross validation*. Eksperimen dan evaluasi model yang dilakukan, dengan pembagian data training dan data testing secara acak dengan beberapa percobaan untuk memperoleh nilai *Accuracy*, *Precision*, *Recall* dan *F1-Score* yang tertinggi.

2. METODOLOGI PENELITIAN

Desain penelitian ini meliputi tahapan-tahapan yang dilakukan dalam menyelesaikan penelitian. Adapun kerangka kerja penelitian yang dilakukan dengan identifikasi masalah, studi literatur, pengumpulan data, analisis dan perancangan system, implementasi, pengujian system, kesimpulan. Seperti pada Gambar 1. Alur Penelitian.



Gambar 1. Alur Penelitian

2.1 Mengidentifikasi masalah

Pada tahap ini dilakukan proses mendeskripsikan masalah, menganalisa masalah serta memahami masalah tersebut sehingga didapatkan solusi dan penyelesaian masalah.

2.2 Studi literatur

Pada tahap ini menjelaskan dasar teori yang berkaitan dengan kebutuhan penelitian yang terdapat teori-teori pendukung yang digunakan. Dasar-dasar teori dan informasi dari buku dan jurnal yang berkaitan dengan kebutuhan penelitian terkait mengenai studi kasus yang berkaitan dengan implementasi data mining menggunakan algoritma K-Nearest Neighbor.

2.3 Pengumpulan data

Data yang digunakan merupakan data yang bersifat primer yang dikarenakan data yang ada berdasarkan nilai rapor, kedisiplinan, ekstrakurikuler, presensi, dan bersifat sekunder dengan melakukan wawancara dengan pihak sekolah yang berkaitan dalam penentuan siswa berprestasi.

2.4 Analisis dan Perancangan Sistem

Penelitian ini bertujuan untuk melihat performance dari algoritma K-Nearest Neighbor dalam mengklasifikasikan data. Alur kerja klasifikasi yang diusulkan dalam penelitian ini meliputi:

1. Selection Data

Pada tahap ini dilakukan seleksi data nilai yang terdiri dari variabel-variabel prediktor dan satu target variabel. Variabel target yaitu klasifikasi nilai baik atau nilai lebih dari 85. Sedangkan variabel-variabel prediktor yaitu, nilai rata-rata semester 1, nilai rata-rata semester 2, nilai rata-rata semester 3, nilai rata-rata semester 4, nilai rata-rata semester 5, nilai rata-rata semester 6, dan jumlah poin kedisiplinan.

2. Preprocessing

Data yang diambil sesuai dengan jumlah siswa kelas XII jurusan Bahasa, Mipa dan IPS sebanyak 348. Dari data yang diambil dilakukan pembersihan data apabila terdapat data yang hilang, data ganda atau bersifat outlier.

3. Transformation

Setelah data bersih dari kesalahan, selanjutnya dilakukan normalization data yaitu dengan melakukan standarisasi data sehingga rentang data menjadi proporsional dengan menggunakan metode *Z-Transformation*. Sebenarnya *Z-Transformation* juga disebut sebagai normalisasi statistik yang bertujuan mengubah data menjadi distribusi normal pada rata-rata = 0 serta varians = 1 Namun, rentang distribusi normal standar tidak antara [0,1] tetapi sekitar -1 hingga +3. Berikut persamaannya (1).

$$Z = \frac{x-\mu}{\sigma} \tag{1}$$

4. Data Mining

Data Mining Pada tahap ini dilakukan pemilihan teknik data mining klasifikasi yang sesuai. Metode klasifikasi yang digunakan adalah algoritma *K-Nearest Neighbor*. Algoritma *K-Nearest Neighbor* (*K-NN*) adalah salah satu algoritma paling sederhana untuk memecahkan masalah klasifikasi dan sering menghasilkan hasil yang kompetitif dan signifikan [9]. Prinsip kerja KNN adalah mencari jarak terdekat antara data yang dievaluasi dengan k tetangga terdekatnya dalam data pelatihan. Persamaan penghitungan untuk mencari *Euclidean* dengan d adalah jarak dan p adalah dimensi data (Persamaan 2).

$$d_i = \sqrt{\sum_{i=1}^p (x_{1i} - x_{2i})^2} \tag{2}$$

Keterangan:

x_1 : sample data uji

x_2 : data uji

d : jarak

p : dimensi data

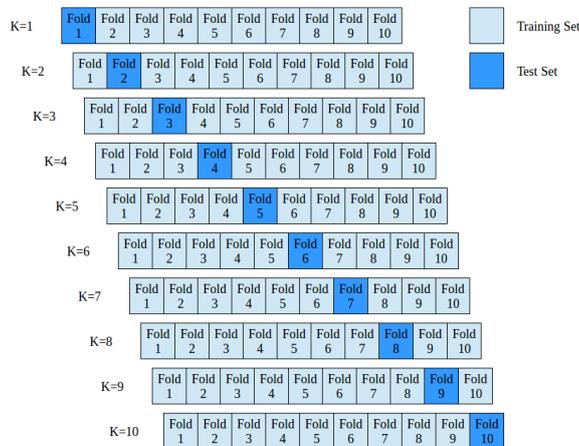
2.5 Implementasi

Implementasi dari penelitian ini sesuai dengan perancangan Algoritma K-Nearest Neighbor (KNN) sangatlah sederhana, bekerja berdasarkan jarak terpendek dari query instance ke training sample untuk menentukan KNN-nya. Langkah yang digunakan dalam metode K-Nearest Neighbor:

1. Tentukan parameter K (jumlah tetangga paling dekat).
2. Hitung kuadrat jarak euclidean masing –masing objek terhadap data sample yang diberikan.
3. Urutkan objek –objek kedalam kelompok yang memiliki jarak terkecil.
4. Kumpulkan kategori Y(Klasifikasi nearest neighbor).
5. Dengan kategori nearest neighbor yang paling banyak, maka dapat diprediksikan nilai query instance yang telah dihitung

2.6 Pengujian Sistem

Menjelaskan pengujian yang dilakukan untuk mendapatkan akurasi sistem yang telah dibuat. Dalam penelitian ini menggunakan metode *K-fold Cross Validation* untuk mengukur keberhasilan model klasifikasi dengan membagi dataset secara acak menjadi k himpunan bagian [10]. Setiap fold akan berfungsi sebagai data uji satu kali, sedangkan sisa fold digunakan sebagai data latih. Dalam setiap iterasi, model akan dilatih dengan menggunakan data latih dan data uji menggunakan data uji yang berbeda setiap iterasi. Gambar 2 menunjukkan *K-Fold Cross Validation* dengan menggunakan nilai k fold 10 [11].



Gambar 2. Diagram 10-fold cross validation

2.7 Evaluation

Evaluation Tahap ini digunakan untuk mengevaluasi klasifikasi algoritma yang dipilih menghasilkan nilai mendekati klasifikasi data sebenarnya. Evaluasi dilakukan dengan menggunakan metode *Confusion Matrix* dan Nilai performansi yang digunakan yaitu *Accuracy* dan *Precision*, *Recall* dan *F1-Score*. Akurasi akan menjelaskan tingkat efektifitas dari metode yang digunakan pada proses klasifikasi. *Recall* nantinya akan menunjukkan padanan rasio prediksi benar positif dengan keseluruhan jumlah data yang benar positif. Sedangkan, *Precision* menunjukkan padanan dari rasio prediksi benar positif dengan keseluruhan hasil yang diprediksi positif. *F1-score* kemudian akan melakukan perhitungan dengan membandingkan rata-rata nilai *Recall* dan *Precision* yang dibobotkan [12][13]. Berikut Persamaannya (3), (4), (5), (6).

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} \tag{3}$$

$$Precision = \frac{TP}{TP+FP} \tag{4}$$

$$Recall = \frac{TP}{TP+FN} \tag{5}$$

$$F1Score = 2 \times \frac{(Recall \times Precision)}{(Recall+Precision)} \tag{6}$$

3. HASIL DAN PEMBAHASAN

Pada bab ini akan diuraikan hasil penelitian dan pembahasan mengenai penelitian yang telah dilakukan meliputi skenario pengujian dari penerapan algoritma K-Nearest Neighbor pada dataset yang digunakan untuk menilai performa model klasifikasi termasuk *Accuracy*, *Precision*, *Recall* dan *F1-Score* terbaik. Skenario pengujian data dengan menggunakan dua data yaitu memisahkan dataset menjadi dua bagian yaitu data training dan data testing secara acak. Data testing dilakukan dengan menentukan nilai k=10, hal ini dikarenakan dengan pengambilan nilai k yang tertinggi. Tujuan pengujian ini adalah untuk mengetahui bagaimana jumlah data training dan testing memengaruhi kemampuan klasifikasi oleh model K-Nearest Neighbor. Pengujian dilakukan dengan memanfaatkan metrik yang dikenal sebagai *Confusion Matrix*. Detail dari model pengujian ini terlihat dalam Tabel 1.

Tabel 1. Pembagian Data Training dan Data Testing

Training %	Testing %
20	80
30	70
40	60
50	50

60	40
70	30
80	20

3.1 Pemilihan Data

Dataset yang digunakan pada penelitian ini yaitu nilai akademik, nilai ekstrakurikuler dan poin alpa siswa SMA Negeri 2 Batu angkatan 2020/2021. Data yang diperoleh 348 data set siswa dengan 11 atribut atau variabel. Variabel yang digunakan anatara lain nama, jenis kelamin, jurusan, nilai semester 1, nilai semester 2, nilai semester 3, nilai semester 4, nilai semester 5, nilai semester 6, poin kedisiplinan dan predikat nilai. Adapun contoh data yang digunakan dapat dilihat pada Tabel 2.

Tabel. 2 Data Nilai Per Semester

No	nama	L/P	jurusan	smt 1	smt 2	smt 3	smt 4	smt 5	smt 6	rata2	predikat
1	Ade Krisna Putra	L	MIPA 1	87.78	83.07	85.69	83.07	85.92	86.15	85.28	BAIK
2	Afrizal Bramantara P	L	MIPA 1	79.07	80.43	80.76	79.38	78.92	80.84	79.90	KURANG
3	Amartya Sota Tantular	P	MIPA 1	88.78	88.64	80.15	86	84.46	82.92	85.16	BAIK
4	Andika Maulana	L	MIPA 1	77.5	79.93	78.92	78.46	79	82.23	79.34	KURANG
5	Ardian Septevano W	L	MIPA 1	79.85	80.36	80.23	81.46	82.23	81.84	81.00	KURANG
...
344	Tria Ayu Khoirunnisa	P	IBBu	77.5	79.5	83.14	82	84.14	85.21	81.92	KURANG
345	Wanda Zellsa Billa	P	IBBu	88.44	89.06	85.71	88.86	86.86	87.57	87.75	BAIK
346	Wassy Assyahadat	L	IBBu	84.94	86.38	85.5	83.5	87.57	88.71	86.10	BAIK
347	Wenda Ferdian Syah	L	IBBu	87.13	89.94	86.43	85.21	84.79	87.43	86.82	BAIK
348	Yeshinta Pangestika	P	IBBu	82.75	84.81	86.36	86.71	87.79	87.57	86.00	BAIK

3.2 Preprocessing

Pra pemrosesan data perlu dilakukan untuk mendapatkan data set dengan kualitas baik. Cara yang dilakukan antara lain *validation* data yaitu untuk menghilangkan duplikasi data dan *missing value*. Untuk mendapatkan data dengan kualitas baik beberapa teknik yang dapat dilakukan antara lain *validation, integration and transformation, size reduction/discritization*. Pentingnya *preproccesing* data sebelum data set diproses menggunakan teknik data mining. *Preproccesing* meliputi: memeriksa dan membuang data yang inkonsisten, data ganda, data yang perlu diperbaiki dan atau menambah data sesuai dengan kebutuhan.

Hasil dari pengumpulan data didapatkan record sebanyak 348 data set siswa dengan 11 variabel. Adapun rincian 11 variabel yang dilakukan terlihat seperti dalam Tabel 3. Berikut.

Tabel 3. Variabel Pemrosesan Data

No	Variabel	Jenis Data
1	Nama	Karakter
2	Jenis Kelamin	Kategorial
3	Jurusan	Karakter
4	Nilai Semester 1	Numerik
5	Nilai Semester 2	Numerik
6	Nilai Semester 3	Numerik
7	Nilai Semester 4	Numerik
8	Nilai Semester 5	Numerik
9	Nilai Semester 6	Numerik
10	Poin Kedisiplinan	Numerik
11	Predikat	Kategorial

3.3 Transformation

Tranformasi Data, digunakan untuk mengubah nilai semester 1, nilai semester 2, nilai semester 3, nilai semester 4, nilai semester 5, nilai semester 6 dan poin kedisiplinan. Data *normalization* yaitu dengan melakukan standarisasi data sehingga rentang data menjadi proporsional dengan menggunakan metode *Z-Transformation*. Hasil implementasi normalisasi data menggunakan persamaan rumus *Z-Score* dapat dilihat pada Tabel 4.

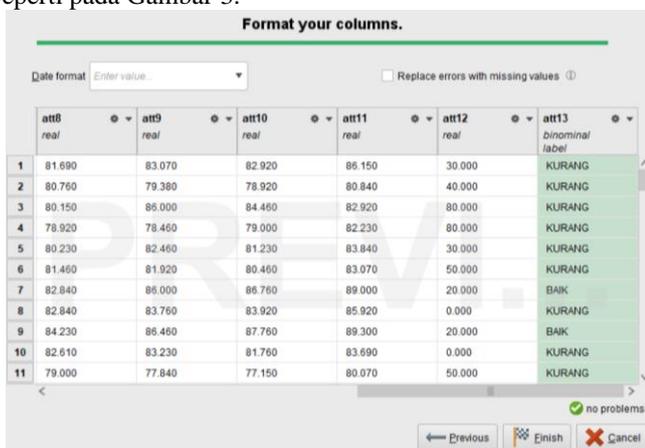
Tabel 4. Hasil Normalisasi Data

NO	X1	X2	X3	X4	X5	X6	X7
1	0.17386016	0.60173925	-0.33353335	-0.28626611	-0.24748991	0.14099554	-0.28332161
2	-0.69973966	-0.74498398	-0.79036492	-1.74026516	-1.34572686	-1.81371097	0.02968131
3	0.17386016	-0.63785827	-1.09000713	0.86826430	0.17533132	-1.04802556	1.28169298
4	-1.50181669	-1.00004520	-1.69420373	-2.10277983	-1.32376212	-1.30202697	1.28169298
5	-0.30125553	0.23955232	-1.05070979	-0.52662910	-0.71149502	-0.70935701	-0.28332161
....
344	1.05256875	0.82109190	0.37873093	0.86826430	0.91115008	0.89931859	1.59469590
345	-0.51071514	-0.93372928	-0.81492575	0.41906134	0.28515501	0.29560509	0.34268423
346	0.25560049	-0.26036766	-0.42686453	-0.51086759	-0.89270411	-0.38909436	1.90769882
347	0.35266714	0.02530090	-0.46124970	-0.62513852	-0.00862337	-0.12404941	0.34268423
348	1.18028802	1.48935229	1.96044885	1.14803105	1.08961358	0.66372308	-0.90932744

3.4 Klasifikasi *K-Nearest Neighbor*

a. Penentuan Atribut Data

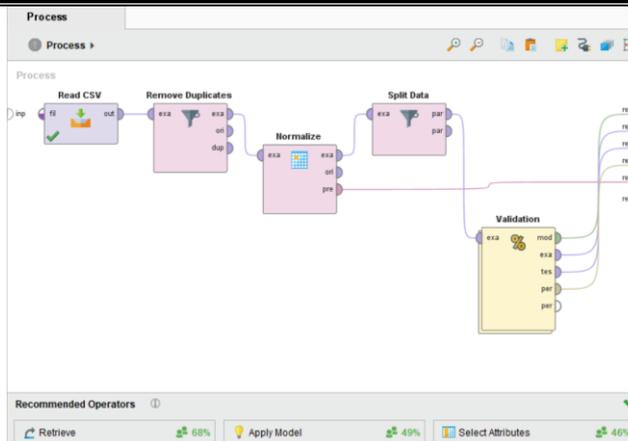
Pada tahap ini, Penentuan atribut dan label pada algoritma *K-Nearest Neighbor*, data-data tersebut menggunakan tipe data numerik pada setiap atribut-atributnya. Pada satu data menggunakan tipe data binominal pada satu atribut label. Seperti pada Gambar 3.



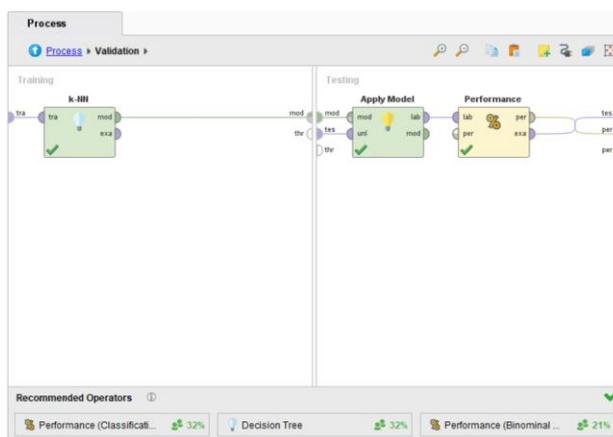
Gambar 3. Penentuan Atribut *K-Nearest Neighbor*

b. Desain Rancangan System dengan Rapid Minner

Gambar 4 dan 5 Rancangan model *K-Nearest Neighbor* untuk menguji dataset. Pada data tersebut dilakukan proses cleaning dengan membuang duplikasi data apabila adanya data tersebut terdapat data duplikasi. Kemudian dilanjutkan menormalisasikan data terlebih dahulu untuk membuat mengubah data menjadi distribusi normal. Penggunaan split data serta *performance classification* untuk mengetahui performa dari algoritma yang di uji ketika proses klasifikasi dan mendapatkan hasil akurasi.



Gambar 4. Desain Model System



Gambar 5. Desain Model *K-Nearest Neighbor*

c. Pengujian Data

Pengujian data dilakukan dengan memasukkan dataset yang telah terformat ke excel (.csv) lalu menghapus data yang kemungkinan adanya duplikasi setelah itu melakukan normalisasi terhadap data, kemudian Split Data dengan memisahkan dataset menjadi dua bagian yaitu data training dan data testing secara acak. Pada tahap selanjutnya dengan menggunakan *cross validation* untuk masukkan model *K-Nearest Neighbor* dengan menggunakan metode *10 fold Cross Validation* dilakukan percobaan sebanyak 10 kali dengan memberikan nilai k 10 menggunakan dataset yang berbeda. Split Data dilakukan untuk data training dan *apply model* sebagai operator *Performance*.

3.5 Evaluation

a. Skenario Pengujian 1

Skenario Pengujian 1 dilakukan untuk mengetahui performa *K-Nearest Neighbor* dengan menggunakan seluruh atribut yang ada pada dataset yang dipisahkan menjadi 20% untuk training dan 80% untuk testing, dengan 70 untuk training dan 278 untuk testing. Detail hasil *Confusion Matrix* dapat ditemukan dalam Gambar 6.

	true KURANG	true BAIK	class precision
pred. KURANG	54	9	85.71%
pred. BAIK	1	5	83.33%
class recall	98.18%	35.71%	

Gambar 6. *Confusion Matrix* Pengujian 1

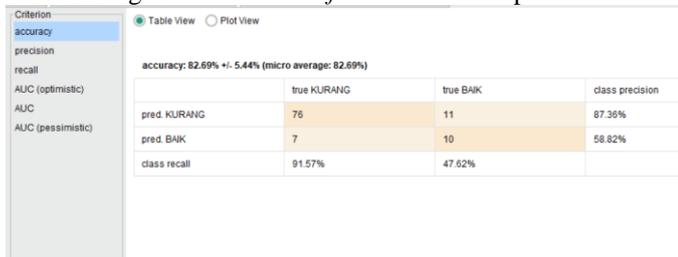
Gambar 6. menunjukkan hasil dari *Confusion Matrix* dalam ujicoba 1, yang mencantumkan nilai-nilai Akurasi, Presisi, Recall dan F1-Score yang akan diuraikan secara lebih detail dalam Tabel 5.

Tabel 5. Accuracy, Precision, Recall, F1-Score pada Ujicoba 1

Accuracy	Precision	Recall	F1-Score
85.67	83.33	50	62.50

b. Skenario Pengujian 2

Skenario Pengujian 2 dataset yang dipisahkan menjadi 30% untuk training dan 70% untuk testing, dengan 104 untuk training dan 244 untuk testing. Detail hasil *Confusion Matrix* dapat ditemukan dalam Gambar 7.



Gambar 7. Confusion Matrix Pengujian 2

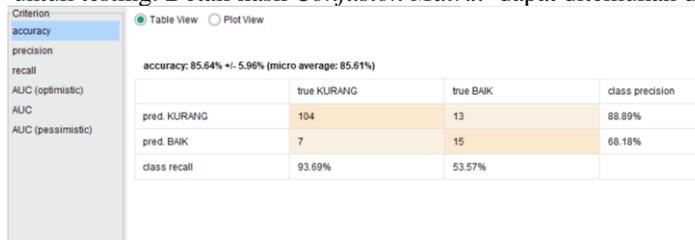
Gambar 7. menunjukkan hasil dari *Confusion Matrix* dalam ujicoba 2, yang mencantumkan nilai-nilai Akurasi, Presisi, Recall dan F1-Score yang akan diuraikan secara lebih detail dalam Tabel 6.

Tabel 6. Accuracy, Precision, Recall, F1-Score pada Ujicoba 2

Accuracy	Precision	Recall	F1-Score
82.69	75	60.71	67.10

c. Skenario Pengujian 3

Skenario Pengujian 3 dataset yang dipisahkan menjadi 40% untuk training dan 60% untuk testing, dengan 139 untuk training dan 209 untuk testing. Detail hasil *Confusion Matrix* dapat ditemukan dalam Gambar 8.



Gambar 8. Confusion Matrix Pengujian 3

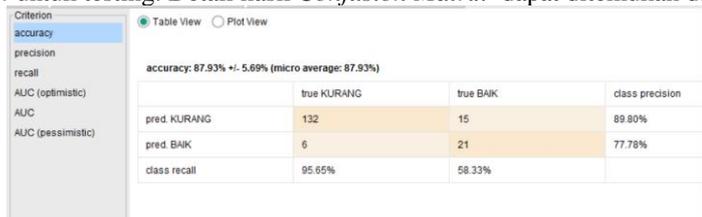
Gambar 8. menunjukkan hasil dari *Confusion Matrix* dalam ujicoba 3, yang mencantumkan nilai-nilai Akurasi, Presisi, Recall dan F1-Score yang akan diuraikan secara lebih detail dalam Tabel 7

Tabel 7. Accuracy, Precision, Recall, F1-Score pada Ujicoba 3

Accuracy	Precision	Recall	F1-Score
85.64	78.12	65.79	71.43

d. Skenario Pengujian 4

Skenario Pengujian 4 dataset yang dipisahkan menjadi 50% untuk training dan 50% untuk testing, dengan 174 untuk training dan 174 untuk testing. Detail hasil *Confusion Matrix* dapat ditemukan dalam Gambar 9.



Gambar 9. Confusion Matrix Pengujian 4

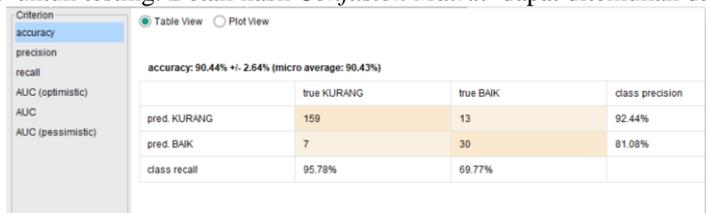
Gambar 9. menunjukkan hasil dari *Confusion Matrix* dalam ujicoba 4, yang mencantumkan nilai-nilai Akurasi, Presisi, Recall dan F1-Score yang akan diuraikan secara lebih detail dalam Tabel 8.

Tabel 8. Accuracy, Precision, Recall, F1-Score pada Ujicoba 4

Accuracy	Precision	Recall	F1-Score
87.93	81.25	67.45	73.71

e. Skenario Pengujian 5

Skenario Pengujian 5 dataset yang dipisahkan menjadi 60% untuk training dan 40% untuk testing, dengan 209 untuk training dan 139 untuk testing. Detail hasil *Confusion Matrix* dapat ditemukan dalam Gambar 10.



Gambar 10. *Confusion Matrix* Pengujian 5

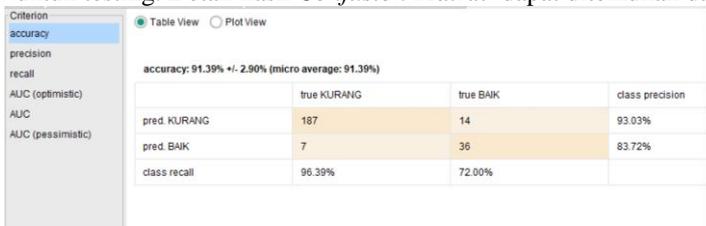
Gambar 10. menunjukkan hasil dari *Confusion Matrix* dalam ujicoba 5, yang mencantumkan nilai-nilai Akurasi, Presisi, Recall dan F1-Score yang akan diuraikan secara lebih detail dalam Tabel 9.

Tabel 9. *Accuracy, Precision, Recall, F1-Score* pada Ujicoba 5

Accuracy	Precision	Recall	F1-Score
90.44	85.42	73.22	78.85

f. Skenario Pengujian 6

Skenario Pengujian 6 dataset yang dipisahkan menjadi 70% untuk training dan 30% untuk testing, dengan 244 untuk training dan 104 untuk testing. Detail hasil *Confusion Matrix* dapat ditemukan dalam Gambar 11.



Gambar 11. *Confusion Matrix* Pengujian 6

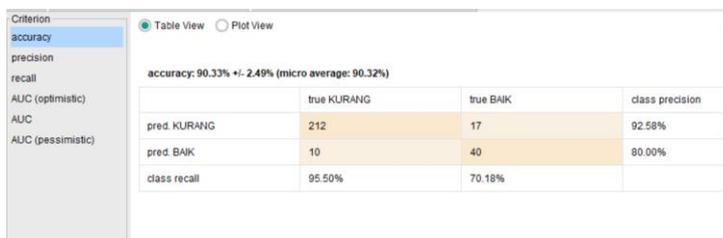
Gambar 11. menunjukkan hasil dari *Confusion Matrix* dalam ujicoba 6, yang mencantumkan nilai-nilai Akurasi, Presisi, Recall dan F1-Score yang akan diuraikan secara lebih detail dalam Tabel 10.

Tabel 10. *Accuracy, Precision, Recall, F1-Score* pada Ujicoba 6

Accuracy	Precision	Recall	F1-Score
91.39	86	76.4	80.92

g. Skenario Pengujian 7

Skenario Pengujian 6 dataset yang dipisahkan menjadi 80% untuk training dan 20% untuk testing, dengan 278 untuk training dan 70 untuk testing. Detail hasil *Confusion Matrix* dapat ditemukan dalam Gambar 12.



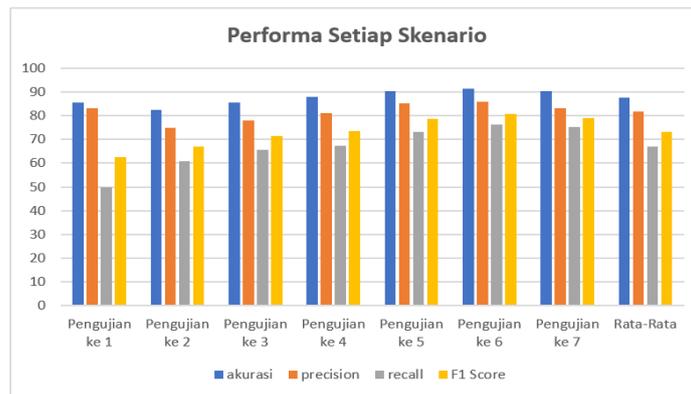
Gambar 12. *Confusion Matrix* Pengujian 7

Gambar 12. menunjukkan hasil dari *Confusion Matrix* dalam ujicoba 7, yang mencantumkan nilai-nilai Akurasi, Presisi, Recall dan F1-Score yang akan diuraikan secara lebih detail dalam Tabel 11.

Tabel 11. *Accuracy, Precision, Recall, F1-Score* pada Ujicoba 7

Accuracy	Precision	Recall	F1-Score
90.33	83.33	75.4	79.17

Dari ke tujuh skenario pengujian di atas menghasilkan nilai rata-rata *Accuracy* 87,73%, rata-rata *Precision* 81.78%, rata-rata *Recall* 67.00% dan rata-rata *F1-Score* 73.38%. Grafik akurasi terbaik pada setiap skenario dapat dilihat pada Gambar 13.



Gambar 13. Grafik Performa Setiap Skenario

4. KESIMPULAN

Berdasarkan hasil penelitian yang sudah dilakukan, dapat diambil kesimpulan bahwa klasifikasi prestasi akademik siswa berdasarkan nilai rapor dan kedisiplinan dengan metode *K-Nearest Neighbor* dengan menggunakan data berjumlah 348 siswa SMA Negeri 2 Batu, dapat di klasifikasi dengan baik. Hal ini dapat dilihat dari analisis hasil pengujian sebanyak tujuh kali menghasilkan nilai rata-rata *Accuracy* 87,73%, rata-rata *Precision* 81,78%, rata-rata *Recall* 67,00% dan rata-rata *F1-Score* 73,38%. Pada pengujian ke-6 dengan split data 70% data training sebanyak 244 dan 30% data testing sebanyak 104 menghasilkan nilai yang tertinggi dengan *Accuracy* 91,39%, *Precision* 86,00%, *Recall* 76,40% dan *F1-Score* 80,92%. Penelitian ini memberikan kontribusi penting dalam konteks prestasi siswa, yang dapat digunakan sebagai dasar sekolah untuk mengambil keputusan siswa berprestasi berdasarkan nilai rapor dan nilai kedisiplinan. Pada penelitian ini masih terdapat kekurangan, maka perlu adanya pengembangan untuk ke depannya dan disarankan untuk melakukan perbandingan dengan berbagai data dan metode klasifikasi yang berbeda.

REFERENCES

- [1] P. Adinoto, "Pengaruh Kegiatan Awal Pembelajaran, Disiplin Belajar Dan Motivasi Belajar Terhadap Prestasi Belajar," *J. Imiah Pendidik. dan Pembelajaran*, vol. 3, no. 1, p. 53, 2019, doi: 10.23887/jipp.v3i1.17110.
- [2] S. Melati and G. Triyono, "Pemodelan Sistem Pendukung Keputusan Penentuan Siswa Terbaik Menggunakan Metode Simple Addictive Weighting (Saw)," *IDEALIS Indones. J. Inf. Syst.*, vol. 3, no. 2, pp. 574–580, 2020, doi: 10.36080/idealism.v3i2.2748.
- [3] S. Jambekar, S. Nema, and Z. Saquib, "Prediction of Crop Production in India Using Data Mining Techniques," in *2018 Fourth International Conference on Computing Communication Control and Automation (ICCCUBEA)*, Aug. 2018, pp. 1–5. doi: 10.1109/ICCCUBEA.2018.8697446.
- [4] A. Sudrajat and I. Budi, "Analisis Kinerja Algoritma Support Vector Machine (Svm) Pada Data Seleksi Penerima Beasiswa Menggunakan Particle Swarm Optimization (Pso) (Studi Kasus: Politeknik Tede Bandung)," *J. TEDC*, vol. 13, no. 1, pp. 1–11, 2019, [Online]. Available: <http://ejournal.poltektedc.ac.id/index.php/tedc/article/view/15>
- [5] S. Adinugroho and Y. Arum Sari, *Implementasi Data Mining Menggunakan Weka*. 2018.
- [6] D. Kurniawan and A. Saputra, "Penerapan K-Nearest Neighbour dalam Penerimaan Peserta Didik dengan Sistem Zonasi," *J. Sist. Inf. Bisnis*, vol. 9, no. 2, p. 212, 2019, doi: 10.21456/vol9iss2pp212-219.
- [7] E. Purwaningsih and E. Nurelasari, "Penerapan K-Nearest Neighbor Untuk Klasifikasi Tingkat Kelulusan Pada Siswa," *Syntax J. Inform.*, vol. 10, no. 01, pp. 46–56, 2021, doi: 10.35706/syji.v10i01.5173.
- [8] V. Alvian, A. Nilogiri, and H. Azizah, "Klasifikasi Siswa Berprestasi Menggunakan Metode K-Nearest Neighbor (KNN) Pada SMA Negeri 2 Situbondo," *J. Smart Teknol.*, vol. 3, no. 6, pp. 602–611, 2022, [Online]. Available: <http://jurnal.unmuhjember.ac.id/index.php/JST>
- [9] D. A. Adeniyi, Z. Wei, and Y. Yongquan, "Automated web usage data mining and recommendation system using K-Nearest Neighbor (KNN) classification method," *Appl. Comput. Informatics*, vol. 12, no. 1, pp. 90–108, 2016, doi: <https://doi.org/10.1016/j.aci.2014.10.001>.
- [10] I. A. Nikmatun and I. Waspada, "Implementasi Data Mining Untuk Klasifikasi Masa Studi Mahasiswa Menggunakan Algoritma K-Nearest Neighbor," *J. SIMETRIS*, vol. 10, no. 2, pp. 421–432, 2019.
- [11] J. Andrade, L. Goliatt, M. Farage, and G. Marques, "PREDICTION OF THE PERFORMANCE OF BITUMINOUS MIXES USING ADAPTIVE NEURO-FUZZY INFERENCE SYSTEMS," *Rev. Mundi Eng. Tecnol. e Gestão (ISSN 2525-4782)*, vol. 5, 2020, doi: 10.21575/25254782rmetg2020vol5n61367.
- [12] D. Yunita, P. Rosyani, and R. Amalia, "Analisa Prestasi Siswa Berdasarkan Kedisiplinan, Nilai Hasil Belajar, Sosial Ekonomi dan Aktivitas Organisasi Menggunakan Algoritma Naïve Bayes," *J. Inform. Univ. Pamulang*, vol. 3, no. 4, p. 209, 2018, doi: 10.32493/informatika.v3i4.2032.
- [13] H. Yuliansyah, R. A. P. Imaniati, A. Wirasto, and M. Wibowo, "Predicting Students Graduate on Time Using C4.5 Algorithm," *J. Inf. Syst. Eng. Bus. Intell.*, vol. 7, no. 1, p. 67, 2021, doi: 10.20473/jisebi.7.1.67-73.