Jurnal Smart Teknologi
Vol. 1, No. 1, Agustus 2021, Halaman 100 – 102
ISSN: 2774-1702, http://jurnal.unmuhjember.ac.id/index.php/JST

Klasifikasi Siswa Berprestasi Menggunakan Metode *K-Nearest Neighbor* (KNN) Pada SMA Negeri 2 Situbondo

Classification Of Achieving Students Using K-Nearest Neighbor (KNN) Method At SMA Negeri 2 Situbondo

Verdi Alvian Dwi Hidayatullah[1], Agung Nilogiri[2], Habibatul Azizah Al Faruq[3]

[1] Mahasiswa Program studi Teknik Informatika, Fakultas Teknik, Universitas Muhammadiyah Jember
[1] Email : verdi.alvian@gmail.com
[2] Fakultas Teknik, Universitas Muhammadiyah Jember
Email : agungnilogiri@unmuhjember.ac.id
[3] Fakultas Teknik, Universitas Muhammadiyah Jember
Email : habibatulazizah@unmuhjember.ac.id

Abstrak

Lingkungan teman sekelas sangat berpengaruh terhadap perkembangan prestasi akademik siswa, jika hubungannya negatif maka akan berdampak negatif bagi siswa, dan juga sebaliknya. SMA Negeri 2 Situbondo merupakan salah satu sekolah di Situbondo yang menerapkan pengacakan pada pembentukan kelasnya, dimana siswa yang tergolong pintar disebar ke beberapa kelas seangkatan untuk membantu siswa lain yang memiliki masalah dalam belajar, penyebaran tersebut harus merata agar efektif. Tujuan penelitian ini adalah untuk mengukur tingkat akurasi, presisi dan recall dalam klasifikasi prestasi siswa kelas X jurusan IPA SMA Negeri 2 Situbondo. Data yang digunakan adalah siswa kelas X jurusan IPA tahun 2020 di SMA Negeri 2 Situbondo sebanyak 182. Dikarenakan belum adanya metode khusus yang digunakan untuk mengklasifikasikan siswa berdasarkan prestasinya dan banyak kemiripan data, maka digunakanlah teknik data mining yaitu metode klasifikasi. Metode klasifikasi yang digunakan pada penelitian kali ini adalah K-Nearest Neighbor (KNN). Hasil dari penelitian ini didapatkan akurasi sebesar 94%, presisi 91% dan recall 91%. Akurasi tesebut didaptkan pada data training fold K = 5 skenario 5 dengan KNN 7.

Kata Kunci : klasifikasi prestasi, *K-Nearest Neighbor* (KNN), prestasi siswa

Abstract

The environment of classmates is very influential on the development of students' academic achievement, if the relationship is negative, it will have a negative impact on students, and vice versa. SMA Negeri 2 Situbondo is one of the schools in Situbondo that applies randomization to class formation, where students who are classified as smart are distributed to several contemporaries to help other students who have problems in learning, the distribution must be evenly distributed to be effective. The purpose of this study was to measure the level of accuracy, precision and recall in the classification of achievement of class X students majoring in science at SMA Negeri 2 Situbondo. The data used are 182 students of class X majoring in Science in 2020 at SMA Negeri 2 Situbondo. Due to the absence of a special method used to classify students based on their achievements and many similarities in the data, data mining techniques are used, namely the classification method. The classification method used in this study is K-Nearest Neighbor (KNN). The results of this study obtained 94% accuracy, 91% precision and 91% recall. This accuracy is obtained from the training data fold K = 5 scenario 5 with KNN 7.

Keywords*: achievement classification, K-Nearest Neighbor (KNN), student achievement*

1. PENDAHULUAN

Prestasi akademik ialah perubahan dalam perihal kecakapan tingkah laku, maupun keahlian yang bisa meningkat sepanjang beberapa waktu serta tidak diakibatkan proses perkembangan, namun karena adanya suasana belajar (Sobur, 2009). Lingkungan teman sebaya sangat berpengaruh terhadap perkembangan prestasi akademik siswa, jika hubungannya negatif maka akan berdampak negatif bagi siswa, dan juga sebaliknya (Noviriandini, 2017). Namun, banyak sekolah yang menerapkan penggolongan dalam pembentukan kelas agar mempermudah guru dalam mengajar, dimana siswa yang pintar akan sekelas dengan siswa yang pintar dan juga sebaliknya, hal ini menyebabkan beberapa siswa belum mencapai prestasi akademik yang maksimal. Untuk menghindari masalah di atas, beberapa sekolah menerapkan pengacakan dalam pembentukan kelas.

SMA Negeri 2 Situbondo merupakan salah satu sekolah di Situbondo yang menerapkan pengacakan pada pembentukan kelasnya, dimana siswa yang tergolong pintar atau mendapat predikat B akan disebar ke beberapa kelas untuk membantu siswa lain yang memiliki masalah dalam belajar, penyebaran tersebut harus merata agar efektif, siswa yang memliki prestasi akademik bagus biasanya adalah siswa yang aktif didalam kelas dan merupakan siswa dengan peringkat 5 besar dikelasnya masing-masing. Predikat atau dategori itu sendiri didapatkan dari perhitungan Nilai total mata pelajaran dengan beban belajar (SKS). Dikarenakan belum adanya metode khusus yang digunakan untuk mengklasifikasikan siswa berdasarkan prestasinya dan banyak kemiripan data, menyebabkan sering terjadi salah pengklasifikasian dan penyebaran siswa pintar yang tidak merata sehingga hasilnya menjadi kurang maksimal. Oleh karena itu dibutuhkan metode klasifikasi yang tepat dan akurat, salah satunya menggunakan ilmu di bidang Data Mining.

2. LANDASARAN TEORI

a. Prestasi Akademik

Prestasi adalah tes standar dalam mengukur pengetahuan seseorang didalam suatu pekerjaan atau belajar, prestasi sering disebut sesuatu yang telah dicapai (Purwodarminto, 1979). Belajar merupakan proses transformasi sikap berkat pengalaman serta latihan. Maksudnya tujuan aktivitas merupakan pergantian tingkah laku, baik yang menyangkut pengetahuan, keahlian ataupun perilaku terlebih lagi meliputi segenap aspek organisme ataupun individu (Djamarah & Zain, 2010). Apabila seseorang siswa tidak memperoleh sesuatu kenaikan mutu serta kuantitas keahlian di dalam proses pendidikan, sehingga bisa dikatakan jika siswa tersebut alami kegagalan di dalam proses pendidikan.

Teman memiliki pengaruh yang sangat besar terhadap prestasi siswa. Lingkungan teman positif akan berdampak positif dan sebaliknya. Agar siswa dapat memperoleh prestasi akademik yang maksimal, maka diperlukan juga lingkungan teman yang baik.

b. *Preprocessing Missing Value*

Missing value merupakan data yang tidak ada untuk suatu objek (permasalahan). Missing value terjadi sebab sesuatu data tentang objek tidak diberikan, susah dicari, ataupun memanglah data tersebut tidak terdapat. Salah satu tata cara untuk menangani permasalahan *missing value* adalah imputasi. Imputasi ialah Proses pengisian ataupun penggantian nilai-nilai pada sekumpulan data (data *set*) yang hilang (*missing values*) dengan nilai- nilai yang bisa jadi (*plausible values*) bersumber pada data yang didapatkan pada dataset tersebut (Myrtveit, dkk. 2001).

Salah satu pendekatan imputasi ialah *mean imputation*, *mean imputation* adalah dimana nilai yang lenyap ditukar oleh rata- rata (*mean*) dari kelompok ilustrasi unit terpaut.

c. *K-Nearest Neighbor*

Algoritma *K-Nearest Neighbor* (KNN) ialah suatu metode klasifikasi yang mengukur jarak antara data uji dengan data *training*. *K-Nearest Neighbor* (KNN) bekerja dengan cara mencari sebanyak *k* data training yang memiliki jarak

terdekat dengan data uji, setelah itu mengklsifikasikan kelas berdasarkan dengan mayoritas kelas *k* data terdekat (Suyanto, 2018). Dekat maupun jauhnya letak (jarak) dapa tdihitung dengan menggunakan rumus jarak seperti rumus jarak *Euclidean* serta jarak *Minkowski*. Dikarenakan memiliki tingkat akurasi yang besar, rumus jarak *Euclidean* sering digunakan dalam mengukur jarak antar data, dan berikut rumus jarak *Euclidean*:

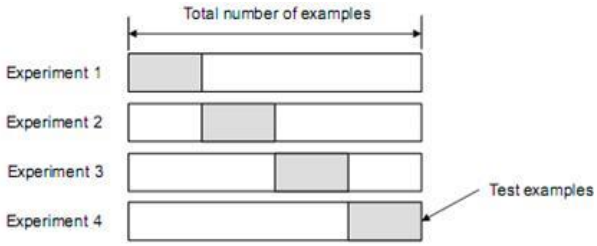
$$d(x_i, y_i) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2} \quad (12)$$

Sumber: Sikki, 2009.
Dimana:

d = Jarak antar data
x_i = sampel data
y_i = data uji
i = Variabel data
n = Dimensi data

d. *K-Fold Cross Validation*

K-Fold Cross validation ialah satu kategori *cross validation* yang berfungsi untuk mengukur serta memperkirakan kinerja model prediktif serta analisa statistik dengan teknik membagi data sample secara acak jadi *K sub-sample* sebanyak nilai *K*. Setelah itu, satu *sub-sampel* dijadikan sebagai data validasi untuk pengujian model, dan *sub-sample K¬1* yang tersisa dipakai sebagai data pelatihan. Proses ini diulang sebanyak *K* kali (Rohani, 2018). Dimisalkan dengan K_f = 4, dimana data hendak dipecah sebanyak 4 himpunan. Pada iterasi pertama berisi 3 himpunan data yaitu D_2, D_3, D_4 sebagai data latih serta satu himpunan D_1 untuk data uji. Pada iterasi kedua berisi 3 himpunan data yaitu: D_1, D_3, D_4 sebagai data latih serta satu himpunan D_2 untuk data uji. Demikian seterusnya untuk iterasi ketiga dan keempat hingga setiap himpunan data pernah menjadi data uji sebanyak satu kali. Contoh penerapan *K-Fold Cross Validation* dengan nilai K_f = 4 yang bisa dilihat pada gambar 1.



Gambar 1 4 *Fold Cross Validation*
Sumber: Slideplayer.info

e. *Confusion Matrix*

Confusion matrix ialah metode yang bisa dipakai memperkirakan kinerja prosedur klasifikasi. Hasil dari confusion matrix memiliki informasi yang bisa digunakan untuk membandingkan hasil klasifikasi yang dicoba oleh program dengan hasil klasifikasi yang sebenarnya (Prasetyo, 2012).

Dalam penggunaan *Confusion Matrix* untuk mengukur kinerja metode klasifikasi terdapat 4 istilah yang mereprestasikan hasil dari klasifikasi. Berikut keempat representasi hasil proses klasifikasi yang ditunjukan pada tabel 1.

Tabel 1 Representasi Hasil Proses Klasifikasi *Confusion Matrix*

True Positive (TP)	Total data positif yang terklasifikasi dengan benar oleh sistem
True Negative (TN)	Total data negatif yang terklasifikasi dengan benar oleh sistem
False Negative (FN)	Total data positif akan tetapi terklasifikasi sebagai data negatif oleh sistem
False Positive (FP)	Total data negatif akan tetapi terklasifikasi sebagai data positif oleh sistem

Sumber: Santra & Christy, 2012

Untuk mengevaluasi suatu model klasfikasi dapat menggunakan akurasi, presisi dan *recall*. Akurasi adalah perbandingan jumlah prediksi yang benar dengan total prediksi. Presisi adalah perbandingan prediksi positif yang benar dengan total prediksi positif. Sedangkan *Recall* adalah persentase prediksi positif dengan data positif (Santra & Christy, 2012). Berikut merupakan rumus akurasi, presisi dan *recall*:

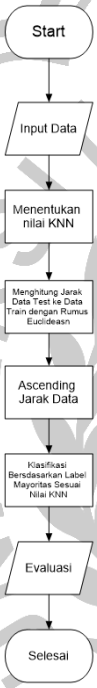
$$\text{Akurasi} = \frac{TP+TN}{TP+FP+TN+FN} \times 100\%$$

$$\text{Presisi} = \frac{TP}{FP+TP} \times 100\%$$

$$\textit{Recall} = \frac{TP}{TP+FN} X\ 100\%$$

3. METODE PENELITIAN

Pada penelitian ini akan dilakukan beberapa tahapan, berikut *flowchart* dari klasifikasi dengan memakai algoritma *K- Nearest Neighbor* serta *K-Fold Cross Validation* yang ditunjukkan pada gambar 2.



Gambar 2 *Flowchart* Alur Klasifikasi *K-Nearest Neighbor* (KNN)
Sumber: Hasil Pengamatan

a. *Pre-Processing*

Pre-processing dilaksanakan guna mempersiapkan data agar tidak ada *missing value* (nilai hilang) yang dapat mempengaruhi proses klasifikasi. Salah satu metode untuk mengatasi *missing value* adalah metode mean imputation. Berikut contoh *missing value* yang terdapat pada atribut sejarah indonesia (SJI).

Tabel 2 Data Nilai Siswa Sebelum *Pre-processing*

No	Nama	SJI
1	Abdul Muqid	82
2	Cherly Audytia	85
3	Findi Fita Deliya	84
4	Gusti Praja Wibowo	80
5	Novika Widianingsih	?
6	Luvia Destari	87
7	Ibnul Mubarok	79
8	Taufiqur Rahman	79
9	Suryatiningsih	?
…..	……...	…….
182	Zulaeha	84

Sumber: SMA Negeri 2 Situbondo

Berikut contoh perhitungan *mean imputation* dengan menghitung nilai rata pada atribut SJI (Sejarah Indonesia) untuk mengisi nilai pada data yang mengalami *missing value.*

$$\text{Rata} - \text{Rata} = \frac{\text{Jumlah Nilai}}{\text{Jumlah Data}} \quad (1)$$

Sumber: Ilham, 2020.

Contoh :

$$\text{Rata} - \text{Rata}_{(SJI)} = \frac{82+85+84+80+87+89+\ldots+84}{180}$$

$$\text{Rata} - \text{Rata}_{(SJI)} = \frac{14952}{180}$$

$$\text{Rata} - \text{Rata}_{(SJI)} = 83{,}07$$

Setelah ditemukan hasil dari *pre-processing*, selanjutnya pembulatan untuk menyeragamkan nilai hasil *pre-processing* dengan nilai yang lain dan data setelah melalu *pre-processing* bisa dilihat pada tabel 3

Tabel 3 Data Nilai Siswa Setelah *Pre-processing*

No	Nama	SJI
1	Abdul Muqid	82
2	Cherly Audytia	85
3	Findi Fita Deliya	84
4	Gusti Praja Wibowo	80
5	Novika Widianingsih	83
6	Luvia Destari	87
7	Ibnul Mubarok	79
8	Taufiqur Rahman	79
9	Suryatiningsih	83
…..	……...	…….
182	Zulaeha	84

Sumber: SMA Negeri 2 Situbondo

b. Menghitung Jarak Menggunakan *Euclidean*

Pada sesi ini hendak dicoba perhitungan jarak antara data training serta data uji guna memperoleh pemodelan di dalam teknik

klasifikasi ke dalam 2 kelas. Untuk mendapatkan jarak *Euclidean* antar data, dilakukan dengan menghitung kedekatan jarak antara data indeks pertama (x_1) dengan data uji indeks ke 182 (x_{182}) sebagai berikut :

$$d(x_1,x_{182}) = \sqrt{(84-84)^2+(79-85)^2+(82-83)^2+\ldots+(78-80)^2}$$

$$d(x_1,x_{182}) = 15{,}198684$$

Selanjutnya menghitung jarak *Euclidean* pada data kedua sampai. Berikut merupakan nilai jarak *Euclidean* terhadap data indeks ke 2 sampai 5 :

$$d(x_2,x_{182}) = 17{,}606817$$

$$d(x_3,x_{182}) = 18{,}248288$$

$$d(x_4,x_{182}) = 13{,}892444$$

$$d(x_5,x_{182}) = 10{,}392305$$

c. Pengujian Menggunakan KNN

Setelah menghitung jarak antara data uji serta *data training* dengan rumus jarak *Euclidean*, berikutnya mengurutkan jarak *Euclidean* dari nilai terkecil hingga nilai terbesar dengan nilai *k* pada *KNN* = 3, dimana data indeks ke 5 (D5) menjadi data dengan jarak *euclidean* terkecil. Berikut tabel *rangking* nilai jarak *Euclidean* yang bisa dilihat pada tabel 4.

Tabel 1 *Rank* Kedekatan Jarak *Euclidean*

Data	*Euclidean*	Kategori
D5	10,392305	B
D4	13,892444	C
D1	15,198684	C
D2	17,606817	C
D3	18,248288	C

Sumber: Perhitungan Excell

Berdasarkan tabel diatas dengan label mayoritas pada nilai *k* pada *KNN* = 3, maka data uji dapat diklasifikasikan masuk ke dalam kelas C.

d. Evaluasi

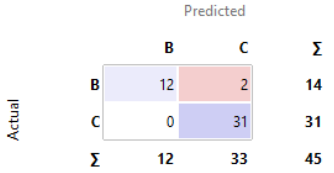
Proses evaluasi menggunakan *Confusion Matrix* yang bermanfaat guna menganalisis kualitasi *classifier* dalam mengidentifikasi tuple-tuple dari setiap kelas yang ada. Berikut represesntasi hasil klasifikasi pada penelitian ini :

Tabel 5 *Representasi* Confusion Matrix *Pada Penelitian Ini*

True Positive (TP)	Total data label B yeng terprediksi benar oleh sistem
True Negative (TN)	Total data label C yeng terprediksi benar oleh sistem
False Postive (FP)	Total data label C yang terprediksi sebagai label B oleh sistem
False Negative (FN)	Total data label B yang terprediksi sebagai label C oleh sistem

Sumber: Hasil Pengamatan

Berikut contoh perhitungan *Confusion Matrix* dengan data *training* 75% dan data *testing* 25% dengan KNN 5 :



Gambar 3 *Confusion Matrix KNN* 5
Sumber: Orange Penelitian Sendiri

Akurasi $= \frac{TP+TN}{TP+FP+TN+FN} \times 100\% = \frac{12+31}{31+2+0+12} \times 100\% = 96\%$

Presisi $= \frac{TP}{FP+TP} \times 100\% = \frac{12}{12+0} \times 100\% = 100\%$

Recall $= \frac{TP}{TP+FN} X\ 100\% = \frac{12}{12+2} \times 100\% = 86\%$

4. HASIL DAN PEMBAHASAN

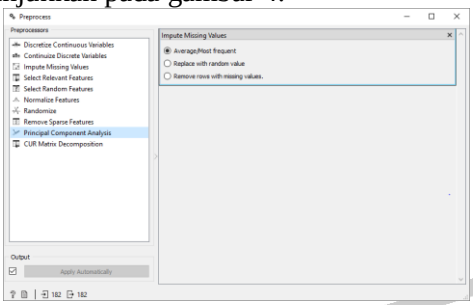
a. *Split* data

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). Unseen data merupakan data yang belum terklasifikasi labelnya (Widaningsih, 2019).

b. *Preprocessing* Data

Preprocessing data nilai siswa dilakukan dengan menggunakan teknik *mean imputation*. *Mean imputation* dilakukan dengan menggunakan *widget preprocess* dan menu *missing values*. Berikut proses *preprocessing* menggunakan *software orange* yang

ditunjukkan pada gambar 4.



Gambar 4 *Preprcessing Missing Value*
Sumber: Orange Penelitian Sendiri

c. *K-fold Cross Validation*

Setelah melewati proses *preprocessing*, selanjutnya akan dilakukan *Cross Validation* dengan metode *K-Nearest Neighbor* (KNN). Proses *Cross Validation* dengan metode *K-Nearest Neighbor* (KNN) dilakukan untuk menentukan skenario terbaik yang kemudian akan diuji dengan *unseen* data.

Untuk *fold* K_f = 2, data akan dibagi menjadi 2 himpunan . dimana setiap himpunan akan berisi 75 data. Skenario terbaik *fold* K_f = 2 ditemukan pada skenario 2 dengan nilai *k* pada *KNN* = 5. Berikut akurasi dari tiap-tiap skenario *fold* K_f = 2 yang dapat dilihat pada tabel 6.

akan dibagi menjadi 3 himpunan . dimana setiap himpunan akan berisi 50 data. Terdapat 2 skenario terbaik *fold* K_f = 3, yiatu pada skenario 3 dengan nilai *k* pada *KNN* 3 dan skenario 3 dengan *KNN* 5. Berikut akurasi dari tiap-tiap skenario *fold* K_f = 3 yang dapat dilihat pada tabel 6.

Untuk *fold* K_f = 5, data akan dibagi menjadi 5 himpunan . dimana setiap himpunan akan berisi 30 data. Terdapat 4 skenario terbaik *fold* K_f = 5, yiatu pada skenario 4 dengan nilai *k* pada *KNN* = 3, skenario 5 dengan nilai *k* pada *KNN* = 3, skenario 5 dengan nilia k pada *KNN* = 5 dan skenario 5 dengan nilai *k* pada *KNN* = 7. Berikut akurasi dari tiap-tiap skenario *fold* K_f = 5 yang dapat dilihat pada tabel 6.

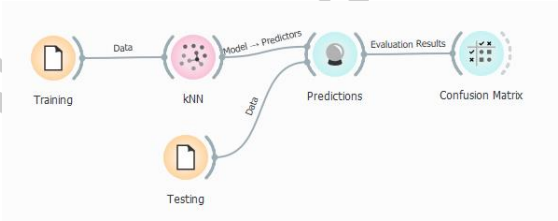
Tabel 6 Hasil *K-Fold Cross Validation*

KNN		3NN			5NN			7NN		
KFold		Akurasi	Presisi	Recall	Akurasi	Presisi	Recall	Akurasi	Presisi	Recall
2Fold	Skenario 1	91%	95%	77%	87%	86%	73%	88%	90%	73%
	Skenario 2	95%	91%	91%	97%	95%	95%	95%	95%	86%
3Fold	Skenario 1	96%	100%	88%	94%	89%	94%	94%	89%	94%
	Skenario 2	86%	80%	75%	88%	81%	81%	90%	82%	88%
	Skenario 3	100%	100%	100%	100%	100%	100%	98%	100%	93%
5Fold	Skenario 1	97%	100%	92%	97%	92%	100%	97%	92%	100%
	Skenario 2	97%	100%	86%	97%	100%	86%	97%	100%	86%
	Skenario 3	87%	100%	71%	90%	100%	79%	80%	83%	71%
	Skenario 4	100%	100%	100%	93%	60%	100%	93%	60%	100%
	Skenario 5	100%	100%	100%	100%	100%	100%	100%	100%	100%

Sumber: Hasil Penelitian

d. *K-nearest Neighbor*

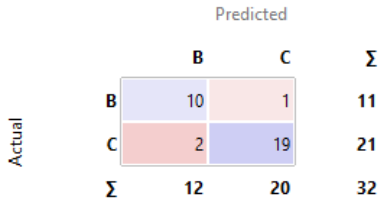
Tahap berikutnya adalah pengujian 6 skenario terbaik dengan metode *K-Nearest Neighbor* (KNN) dimana data *testing* dari setiap skenario akan diganti dengan 32 *unseen* data yang telah dipisah diawal. Berikut proses *K-Nearest Neighbor* (KNN) menggunakan software orange pada gambar 5.



Gambar 5 *K-Nearest Neighbor* (KNN)
Sumber: Orange Penelitian Sendiri

Pengujian data *training fold* K_f = 3 skenario 3 dengan *unseen* data menggunakan nilai k

pada *KNN* = 3, didapatkan akurasi sebesar 91%, presisi 83% dan *recall* 91%.



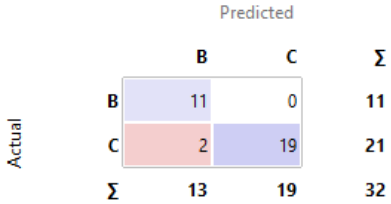
Gambar 6 *Confusion Matrix Fold* K_f = 3 Skenario 3 dengan Nilai *k* Pada KNN = 3
Sumber: Orange Penelitian Sendiri

$$\text{Akurasi} = \frac{10+19}{10+1+2+19} \; x\; 100\% = 91\%$$

$$\text{Presisi} = \frac{10}{10+2} \; x\; 100\% = 83\%$$

$$\textit{Recall} = \frac{10}{10+11} \; x\; 100\% = 91\%$$

Pengujian data *training fold* K_f = 3 skenario 3 dengan *unseen* data menggunakan KNN 5, didapatkan akurasi sebesar 94%, presisi 85% dan *recall* 100%.



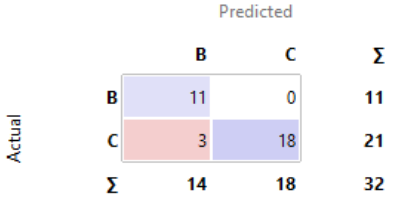
Gambar 7 *Confusion Matrix Fold* K_f = 3 Skenario 3 dengan Nilai *k* Pada KNN 5
Sumber: Orange Penelitian Sendiri

$$\text{Akurasi} = \frac{11+19}{11+0+2+19} \; x\; 100\% = 94\%$$

$$\text{Presisi} = \frac{11}{11+2} \; x\; 100\% = 85\%$$

$$\textit{Recall} = \frac{11}{11+0} \; x\; 100\% = 100\%$$

Pengujian data *training fold* K_f = 5 skenario 4 dengan *unseen* data menggunakan nilai *k* pada *KNN* = 3, didapatkan akurasi sebesar 91%, presisi 79% dan *recall* 100%.



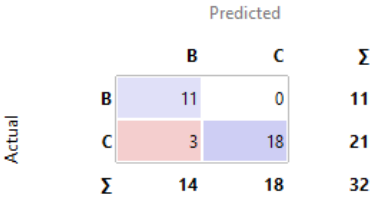
Gambar 8 *Confusion Matrix Fold* K_f = 5 Skenario 4 dengan Nilai *k* Pada *KNN* = 3
Sumber: Orange Penelitian Sendiri

$$\text{Akurasi} = \frac{11+18}{11+0+3+18} \; x\; 100\% = 91\%$$

$$\text{Presisi} = \frac{11}{11+3} \; x\; 100\% = 79\%$$

$$\textit{Recall} = \frac{11}{11+0} \; x\; 100\% = 100\%$$

Pengujian data *training fold* K_f = 5 skenario 5 dengan *unseen* data menggunakan KNN 3, didapatkan akurasi sebesar 91%, presisi 79% dan *recall* 100%.



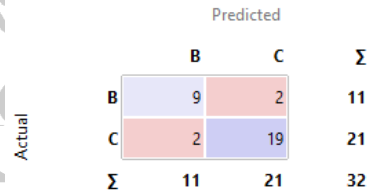
Gambar 9 *Confusion Matrix Fold* K_f = 5 Skenario 5 dengan Nilai *k* Pada *KNN* = 3
Sumber: Orange Penelitian Sendiri

$$\text{Akurasi} = \frac{11+18}{11+0+3+18} \; x\; 100\% = 91\%$$

$$\text{Presisi} = \frac{11}{11+3} \; x\; 100\% = 79\%$$

$$\textit{Recall} = \frac{11}{11+0} \; x\; 100\% = 100\%$$

Pengujian data *training fold* K_f = 5 skenario 5 dengan *unseen* data menggunakan nilai *k* pada *KNN* = 5, didapatkan akurasi sebesar 87%, presisi 82% dan *recall* 82%.



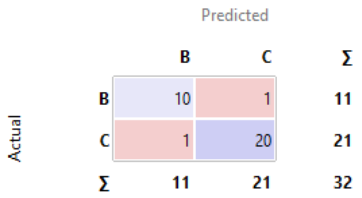
Gambar 10 *Confusion Matrix Fold* K_f = 5 Skenario 5 dengan Nilai *k* Pada *KNN* = 5
Sumber: Orange Penelitian Sendiri

$$\text{Akurasi} = \frac{9+19}{19+2+2+19} \; x\; 100\% = 87\%$$

$$\text{Presisi} = \frac{9}{9+2} \; x\; 100\% = 82\%$$

$$\textit{Recall} = \frac{9}{9+2} \; x\; 100\% = 82\%$$

Pengujian data *training fold* K_f = 5 skenario 5 dengan *unseen* data menggunakan nilai *k* pada *KNN* = 7, didapatkan akurasi sebesar 94%, presisi 91% dan *recall* 91%.



Gambar 11 *Confusion Matrix Fold* K_f = 5 Skenario 5 dengan Nilai *k* Pada *KNN* = 7
Sumber: Orange Penelitian Sendiri

$$Akurasi = \frac{10 + 20}{10 + 1 + 1 + 20} \; x\; 100\% = 94\%$$

$$Presisi = \frac{10}{10 + 1} \; x\; 100\% = 91\%$$

$$Recall = \frac{10}{10 + 1} \; x\; 100\% = 91\%$$

Berdasarkan pengujian 32 *unseen* data terhadap data *training* dari 6 skenario menggunakan algoritma *K-Nearest Neighbor* (KNN), maka didapatkan akurasi terbaik sebesar 94%, presisi 91% dan *recall* 91% pada *fold* K_f = 5 skenario 5 dengan nilai *k* pada *KNN* = 7.

Tabel 7 Perbandingan Akurasi *K-Fold Cross Validation* Dan *K-Nearest Neighbor*

Data *Training*	*K-Fold Cross Validation*			KNN		
	Akurasi	Presisi	*Recall*	Akurasi	Presisi	*Recall*
Fold K_f = 3 Skenario 3 dengan *KNN* 5	100%	100%	100%	94%	85%	100%
Fold K_f = 5 Skenario 5 dengan KNN 7	100%	100%	100%	94%	91%	91%

Sumber: Hasil Penelitian

5. KESIMPULAN DAN SARAN

a. Kesimpulan

Berdasarkan penelitian yang telah di lakukan, dapat di ambil kesimpulan sebagai berikut**:**

1. Hasil akurasi tertinggi klasifikasi siswa berprestasi dengan menggunakan algoritma *K-Fold Cross Validation* dan *K-Nearest Neighbor* (KNN) sebesar 100%, dengan presisi 100% dan *recall* 100%. Akurasi tertinggi tersebut didapatkan pada 6 skenario, yaitu *fold* K_f = 3 skenario 3 dengan nilai *k* pada *KNN* = 3, *fold* K_f = *3* skenario 3 dengan nilai *k* pada *KNN* = 5, *fold* K_f = 5 skenario 4 dengan nilai *k* pada *KNN* = 3, *fold* K_f = 5 skenario 5 dengan nilai *k* pada *KNN* = *3*, *fold* K_f = 5 skenario dengan nilai *k* pada *KNN* = 5 dan *fold* K_f = 5 skenario 5 dengan nilai *k* pada *KNN* = 7.
2. Hasil akurasi tertinggi menggunakan algoritma *K-Nearest Neighbor* (KNN) dengan menggunakan data *training* pada 6 skenario dan data *testing* berupa data final sebanyak 32 data, didapatkan akurasi sebesar 94%, presisi 91% dan *recall* 91%. Akurasi tesebut didaptkan pada data *training fold* K_f = 5 skenario 5 dengan nilai *k* pada *KNN* = 7.
3. Kesimpulannya, secara keseluruhan kinerja lagoritma *K-Nearest Neighbor* (KNN) dalam mengklasifikasikan data siswa berprestasi sangat bagus, walau akurasinya mengalami penurunan sebesar 6% setelah data *testing*-nya diganti dengan data *testing* final.

b. Saran

Berdasarkan penelitian yang telah dilakukan, beberapa saran yang dapat dikembangkan untuk peneliti selanjutnya adalah sebagai berikut:

1. Peneliti selanjutnya dapat mencoba menggunakan algoritma klasifikasi yang lain agar mendapatkan performa akurasi yang lebih baik lagi.
2. Peneliti selanjutnya dapat mengembangkan penelitian ini dengan membuat sebuah aplikasi berdasarkan hasil akhir dari penelitian ini.
3. Peneliti selanjutnya dapat menggunakan data yang lebih banyak dan menggunakan atribut yang lebih spesifik

untuk diklasifikasikan agar lebih akurat dalam mengklasifikasikan prestasi siswa.

6. REFERENSI

Ilham, Ahmad. 2020. Hybird Metode Boostrap dan Teknil Imputasi Pada Metode C4-5 Untuk Prediksi Penyakit Ginjal Kronis. Statistika, vol. 8, no. 1. Jurusan Teknik Informatika, Universitas Muhammadiyah Semarang.

Demsar, J., Tomaz Curk, dkk. 2013. Orange : Data Mining Toolbox in Python. Journal of Machine Learning Research 14 (2013) 2349-2353. Faculty of Computer and Information Science, University of Ljubljana.

Djamarah dan Zain. 2010. 26 Pengertian Belajar Menurut Para Ahli Pendidikan dan Daftar Pustakanya. https://www.gurupendidikan.co.id/pengertian-belajar/. Diakses 11 Februari 2021.

Fayyad, U. 1996. Knowledge Discovery and Data Mining : Towards a Unifying Framework. KDD-96. Information and Computer Science, University of California.

Malik, K. dan Mochammad Faid. 2016. Prediksi Prestasi Siswa SMP Nurul Jadid Menggunakan Algoritma C4.5. Nusantara Journal of Computers and its Applications, Vol. 2, No. 4. Jurusan Teknik Informatika, Sekolah Tinggi Teknologi Nurul Jadid.

Myrtveit, I., Erik Strensrud dan H. Olsson. 2001. Analyzing Data Sets with Missing Data : An Empirical Evaluation of Imputation Methods and Likelihood-Based Methods. IEEE Transactions of Software Engineering. Vol. 27, No. 11. DOI: 10.1109/32.965340. BI Norwegian Business School, Norwegia.

Naik, A. dan Lilavati Samant. 2016. Correlation review of classification algorithm using data mining tool: WEKA, Rapidminer , Tanagra ,Orange and Knime. Procedia Computer Science 85 (2016) 662 – 668. India : International Conference on Computational Modeling and Security.

Noviriandini, A. dan Nurajijah. 2019. Analisis Kinerja Algoritma C4.5 dan Naïve Bayes Untuk Memprediksi Prestasi Siswa Sekolah Menengah Kejuruan. Jurnal Ilmu Pengetahuan dan Teknologi Komputer. Vol. 5, No.1, E-ISSN: 2527-4864. Jurusan Ilmu Komputer, STMIK Nusa Mandiri.

Pemerintah Indonesia. 2003. Undang-Undang Republik Indonesia No. 20 Tahun 2003 Tentang Sistem Pendidikan Nasional. Lembaga Negara RI Tahun 2003 No. 20. Sekretarian Negara. Jakarta.

Purwodarminto, 1979. Pengertian Prestasi Belajar Menurut Para Ahli. https://silabus.org/pengertian-prestasi-belajar/. Diakses 11 Februari 2021.

Santra, A.K. dan C. Josephine Christy. 2012. Genetic Algorithm and Confusion Matrix for Document Clustering. International Journal of Computer Science Issues. Vol. 9, Issue 1, No 2, ISSN : 1694-0814. Bharathiar University, India.

Sikki, Muhammad Ilyas. 2009. Pengenalan Wajah Menggunakan K-Nearest Neighbour dengan Praproses Transformasi Wavelet. Jurnal Paradigma Vol X. No 2.

Sobur. 2009. Pengertian Prestasi Akademik. http://hitamandbiru.blogspot.com/2012/06/pengertian-prestasi-akademik.html. Diakses 4 Februari 2021.

Solochin, Achmad. 2017. Mengukur Kinerja Algoritma Klasifikasi dengan Confusion Matrix. https://achmatim.net/2017/03/19/mengukur-kinerja-algoritma-klasifikasi-dengan-confusion-matrix/. Diakses 12 Februari 2021.

Suyanto, Sity Ayu Novarina. 2018. Klasifikasi Jenis Infeksi Berdasarkan Hasil Pemeriksaan Leukosit Menggunakan K-Nearest Neighbor (KNN). Teknologi Informasi, Universitas Sumatera Utara, Medan.

Turban, dkk. 2005. Data Mining.

https://medium.com/@16611061/data-mining-d48b2389b61. Diakses 11 Februari 2021.