

Klasifikasi Penjurusan Siswa Menengah Atas Menggunakan Metode K-Nearest Neighbor

Classification of majors for high school students using the K-Nearest Neighbor method

Abul Abbas Al marsi¹, Agung Nilogiri^{2*}, Reni Umilasari³

¹ Mahasiswa Program studi Teknik Informatika, Fakultas Teknik, Universitas Muhammadiyah Jember

¹Email : abula625@gmail.com

² Dosen Fakultas Teknik, Universitas Muhammadiyah Jember* Koresponden Author

Email : agungnilogiri@unmuhjember.ac.id

³ Dosen Fakultas Teknik, Universitas Muhammadiyah Jember

Email : reni.umilasari@unmuhjember.ac.id

Abstrak

Pelaksanaan kurikulum 2013 bertujuan untuk penyesuaian program pendidikan pada siswa, agar memiliki kemampuan dalam melakukan observasi, bertanya, bernalar, dan mengkomunikasikan apa yang diperoleh setelah menerima materi pembelajaran. SMA Negeri 2 Situbondo merupakan salah satu SMA Negeri di Kota Situbondo yang sudah menerapkan Kurikulum 2013 pada penentuan penjurusan. Tujuan penelitian ini untuk mengetahui tingkat *akurasi*, *presisi* dan *recall* menggunakan algoritma *Klasifikasi K-Nearest Neighbor* dalam penjurusan di SMA Negeri 2 Situbondo. Penulis menggunakan metode *K - Nearest Neighbor* (KNN) dengan *Euclidean Distance* untuk meng*Klasifikasikan* ke dalam dua jurusan yaitu IPA dan IPS. Digunakan 350 data yang di bagi menjadi dua yaitu data *testing* 20% dan data *training* 80%, Hasil dari penelitian ini didapatkan akurasi sebesar 76%, presisi 78% dan *recall* 96%. Akurasi tersebut didapatkan pada data *training* *Fold K = 5* skenario 3 dengan *KNN 9*. Kinerja algoritma KNN cukup baik dalam *Klasifikasi* penjurusan, disimpulkan hasil *akurasi* data final mengalami peningkatan 6% hal tersebut tidak menyebabkan terjadinya *overfitting* atau *underfitting*.

Kata Kunci : *Klasifikasi, K-Nearest Neighbor, Euclidean Distance, data mining, Penjurusan.*

Abstract

*This implementation aims to adjust the education program for students, so that they have the ability to observe, asking questions, reason, and communicate what is obtained after receiving learning materials. SMA Negeri 2 Situbondo is one of the public high schools in Situbondo City that has implemented the 2013 Curriculum in determining majors. The purpose of this study was to determine the level of accuracy, precision and recall using the K-Nearest Neighbor classification algorithm in majors at SMA Negeri 2 Situbondo. The author uses the K - Nearest Neighbor (KNN) method with Euclidean Distance to classify into two majors, namely science and social studies. 350 data is used which is divided into two, namely 20% testing data and 80% training data, The results of this study obtained 76% accuracy, 78% precision and 96% recall. This accuracy is obtained from the training data *Fold K = 5* scenario 3 with *KNN 9*. The performance of the KNN algorithm is quite good in the classification of majors, it is concluded that the accuracy of the final data has increased by 6%, this does not cause overfitting or underfitting.*

Keywords : *Classification, K-Nearest Neighbor, Euclidean Distance, data mining, Major.*

1. PENDAHULUAN

Pelaksanaan kurikulum 2013 telah banyak diterapkan di Sekolah Menengah Atas (SMA). Kurikulum 2013 bertujuan untuk penyesuaian program pendidikan pada satuan pembelajaran dengan keadaan dan kekhasan kemampuan yang terdapat di daerah pada siswa, supaya siswa unggul dalam segi observasi, tanya jawab, penalaran, serta dapat menyampaikan materi pelajaran yang sudah dipelajari. Pada kurikulum 2013 penjurusan siswa SMA dilakukan saat awal masuk sekolah yaitu pada kelas X. Hal ini sangatlah berbeda dengan kurikulum yang sebelumnya yaitu Kurikulum Tingkat Satuan Pembelajaran (KTSP), dalam kurikulum ini penjurusan siswa dilakukan pada tahun ke-2 yaitu kelas XI.

SMA Negeri 2 Situbondo ialah salah satu sekolah yang menerapkan Kurikulum 2013. Hal ini bertujuan agar siswa SMA Negeri 2 Situbondo dapat menyalurkan bakat, atensi serta keahlian di jurusan yang ada. Penjurusan yang dicoba SMA Negeri 2 Situbondo baik IPA maupun IPS hanya dilakukan dengan memperhitungkan angket peminatan. Sehingga hal tersebut masih kurang akurat serta menimbulkan banyak siswa yang merasa salah masuk jurusan. Oleh sebab itu dibutuhkan *Klasifikasi* penjurusan yang pas serta akurat, salah satunya menggunakan teknologi Data Mining.

Klasifikasi merupakan proses untuk mendapatkan model guna mendeskripsikan perbedaan suatu konsep maupun kelas data, bertujuan agar bisa mengira-ngira kelas dari yang labelnya tidak diketahui. karena itu riset ini menganalisis akurasi, presisi dan recall pelaksanaan metode *Klasifikasi K-Nearest Neighbor* (K-NN) digunakan untuk mengklasifikasikan siswa SMA dalam penjurusan Minat Program Studi, Riset di SMA Negeri 2 Situbondo.

2. LANDASAN TEORI

a. Kurikulum 2013

Kurikulum dalam dunia pendidikan merupakan sebuah sistem yang dilaksanakan oleh sebuah lembaga pendidikan guna mempermudah proses pelaksanaan

pembelajaran dan mengikatkan hasil belajar para peserta didik. Dengan diterapkannya K13 dalam suatu lembaga pendidikan diharapkan mampu membentuk siswa yang produktif, kreatif serta inovatif. K13 menitik beratkan ukuran pedagogik modern dalam pendidikan, yaitu dengan pendekatan ilmiah di pendidikan seluruh pelajaran, serta tahapan memperoleh dan menghimpun data dicoba beserta evaluasi otentik, Kemendikbud (2013: 210).

b. K-Nearest Neighbor (K-NN)

Algoritma K-NN masuk ke dalam kelompok *instance-based learning*. KNN diterapkan dengan menemukan kelompok k objek pada data *training* berdasarkan kedekatan objek data *testing* (Suyanto, 2018). Tujuannya agar dapat mengklasifikasikan objek baru lewat atribut serta contoh dari data *training*. Algoritma ini dihitung bersumber pada jarak minimum dari data baru ke informasi latih (Chandra, 2013). Setelah itu diperoleh nilai yang sering muncul selaku hasil prediksi dari data tersebut.

c. Euclidean Distance

Euclidean (query instance) melakukan perhitungan kuadrat jarak dari tiap-tiap objek terhadap information sample yang tersedia. Pada perhitungan jarak terdapat berbagai rumus seperti *Manhattan distance*, *Square Euclidean Distance*, dan *Euclidean Distance*. Secara umum digunakan dalam menjelaskan jarak antar dua objek x dan y, lalu dipakai rumus jarak *Euclidean Distance* dapat dicermati pada persamaan (1).

$$D(a, b) = \sqrt{\sum_{k=1}^d (X_k - Y_k)^2}$$
$$X^1 = [X_1, X_2, \dots, X_p] \& Y^1 = [Y_1, Y_2, \dots, Y_p]$$
$$D_{(x,y)}^2 = (X_1 - Y_1)^2 + (X_2 - Y_2)^2 + \dots + (X_p - Y_p)^2 \quad (1)$$

Matriks $D(a, b)$ merupakan jarak skala di ke dua vektor a dan b maka matriks ukuran d dimensi. Contoh, ada 2 buah objek akan diukur kedekatannya menggunakan *Euclidean Distance*. Berikut 2 buah objek tersebut:

$$a = [0, 2, 9, 3]$$
$$b = [7, 3, 1, 5]$$

Jadi, jarak ke-2 objek a dan b apabila dihitung adalah:

$$d_{Euclidean}(a, b) = \sqrt{(0 - 7)^2 + (2 - 3)^2 + (9 - 1)^2 + (3 - 5)^2} = 10,86$$

Jadi, kedekatan antara objek a dan b adalah 10,86.

d. K-Fold Cross Validation

K-Fold Cross Validation adalah model metode *cross validation* yang populer. Dimana dataset akan dibagi sebanyak K secara acak. Sebagai gambaran, *5-Fold Cross Validation* data eksperimen nya sebagai berikut:

Tabel 1. 5-Fold Cross Validation

Eksperimen Ke-	Data Latih	Data Uji
1	K2, K3, K4, K5	K1
2	K1, K3, K4, K5	K2
3	K1, K2, K4, K5	K3
4	K1, K2, K3, K5	K4
5	K1, K2, K3, K4	K5

Sumber: Wihardi, 2013.

Keunggulan metode ini ialah semua dataset yang digunakan menjadi data *training* dan data *testing* agar meningkatkan akurasi Pemodelan *Klasifikasi*.

e. Confusion Matrix

Confusion Matrix ialah tata cara yang bisa dipakai memperkirakan kinerja prosedur *Klasifikasi*. Pengukuran kemampuan memanfaatkan *Confusion Matrix*, terselip 4 sebutan untuk gambaran hasil proses *Klasifikasi*. Keempat sebutan tersebut yakni *True Positive* (TP), *True Negative* (TN), *False Positive* (FP) *False Negative* (FN). Berikut uraian mengenai representasi hasil proses *Klasifikasi* (Santra, 2012):

Tabel 2. Confusion Matrix

	Nilai Sebenarnya	
	TRUE	FALSE
Nilai Prediksi	TRUE	FALSE
	FALSE	TRUE

Nilai Prediksi	TRUE	True Positive (TP)	False Positive (FP)
	FALSE	False Negative (FN)	True Negative (TN)

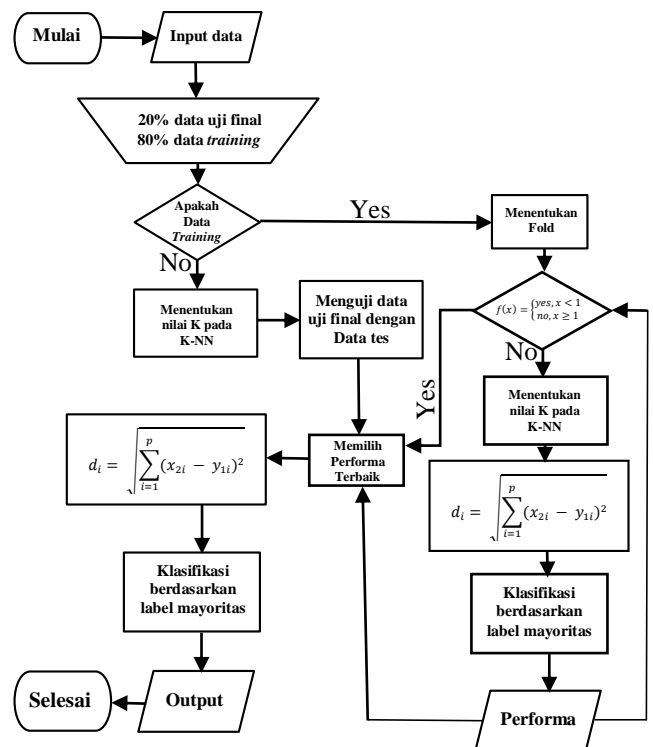
$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} \times 100\%$$

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

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

Sumber: Santra, 2012.

3. METODE PENELITIAN



Gambar 1. Flowchart Alur Klasifikasi K-Nearest Neighbor (KNN)

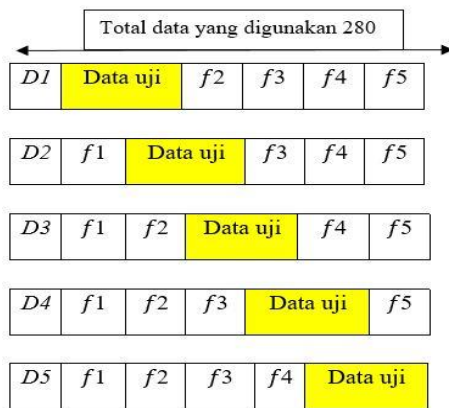
Sumber: Hasil gambar alur penelitian sendiri.

a. Implementasi

Dalam implementasi ini memiliki tujuan untuk pengolahan dan penerapan pada data yang telah terkumpul, memakai algoritma K-NN. Atribut yang dipakai adalah nilai rapor (Bahasa Indonesia, Bahasa Inggris, IPA, & IPS) selama 3 tahun. Data dibagi menjadi 2 yakni data *training* dan data *testing* dengan persentase 80% data *training* dan 20% data *testing* sehingga terdapat 280 data *training* dan 70 data *testing*. Dengan nilai tetangga terdekat pada *Klasifikasi* K-NN dengan nilai $k = 3, k = 5, k = 7, k = 9$

dan $k = 11$ sehingga akurasi dan presisi yang bermacam-macam pada setiap percobaan.

Dari 80% data *training* sebanyak 280 data diolah kembali dengan Metode *K-Fold Cross Validation* dengan membagi data sebanyak K secara acak dan K yang digunakan adalah $K=2$, $K=4$ dan $K=5$, (*Fold*): f_1, f_2, \dots, f_K , dan setiap *Fold* berisi 1 K bagian data. Kemudian dibangun K himpunan data: D_1, D_2, \dots, D_K , berisi $(K - 1)$ *Fold* data latih, 1 *Fold* untuk data uji. Misalkan, $K=5$, dibangun data sebanyak 5 himpunan. D_1 terdiri dari empat *Fold*: f_2, f_3, f_4, f_5 , untuk data latih, satu *Fold* f_1 data uji. D_2 terdiri *Fold*: f_1, f_3, f_4, f_5 , data latih, satu *Fold* f_2 data uji. Dan seterusnya himpunan data D_3, D_4, D_5 jadi setiap *Fold* bisa menjadi data uji.



Gambar 2. Alur *Cross Validation*

Sumber: Hasil gambar *Cross Validation*.

Pengujian data *training* dan penentuan jurusan, setelah mengumpulkan data akan dilakukan kelas “Jurusan IPA” dan “Jurusan IPS” dari data siswa di SMA Negeri 2 Situbondo, terdapat 5 variable untuk membuat data *training* yang terdiri dari:

- Nilai Matematika
- Nilai IPA
- Nilai IPS
- Nilai Bahasa Indonesia
- Nilai Bahasa Inggris

Untuk menentukan nilai mata pelajaran tersebut diambil dari rekap nilai rapor SMP selama tiga tahun pada saat seleksi masuk.

b. Klasifikasi Algoritma K-NN

- Terdapat data objek yang dilakukan pengukuran tingkat kedekatannya

menggunakan *Euclidean Distance* sebagai berikut:

Tabel 3 Sampel Data X

NO	No daftar	Nama	B.INDO	B.IGN	IPA	IPS	MTK	Minat
71	***	NOER	87	84	83	85	85	?

Sumber: Dataset Excel.

Tabel 4 Sampel Data Y

NO	No daftar	Nama	B.INDO	B.IGN	IPA	IPS	MTK	Minat
1	***	INDRI	87	87	80	81	85	IPA

Sumber: Dataset Excel.

- Maka, jarak objek X dan Y jika dihitung adalah:

$$d_{Euclidean}(x,y) = \sqrt{(87 - 87)^2 + (87 - 84)^2 + (80 - 83)^2 + (81 - 85)^2 + (85 - 85)^2}$$

$$d_{Euclidean}(x,y) = 5,830951895$$

- Langkah-langkah untuk memastikan *Euclidean* pada dokumen kedua serta ketiga sama saja serupa memutuskan *Euclidean* awal. Berikut ini ialah nilai *Euclidean* (D_2) dan (D_3):

Tabel 5 Rangkang *Euclidean Distance*

Dokumen Uji	Rangkang	<i>Euclidean Distance</i>	Jurusan
D1	66	5,83095	IPA
D2	10	4,24264	IPA
D3	1	2,82842	IPA
.....
D9	38	5,09901	IPS
D10	2	3,16227	IPA
.....
D16	63	5,74456	IPS
D17	3	3,31662	IPS
D18	14	4,35889	IPA
.....

Sumber: Data Hasil Perhitungan Excel.

- Pengujian Menggunakan Metode K-NN Setelah dilakukan pengukuran jarak *Euclidean*, selanjutnya mengatur jarak *Euclidean* dari nilai terkecil sampai nilai

yang terbesar dengan $k=3$, $k=5$, $k=7$, $k=9$ dan $k=11$. Sebagai data uji dilakukan pada Dokumen Uji (DU) dengan jarak *Euclidean* seperti yang ditunjukkan pada tabel berikut:

Tabel 6 *Euclidean Distance* $k = 3$

Rangking	<i>Euclidean Distance</i>	jurusan
1	2,82842	IPA
2	3,16227	IPA
3	3,31662	IPS

Sumber: Data Hasil Perhitungan Excel.

- Jadi jarak *Euclidean* terkecil berada pada Dokumen Uji (D3), maka data uji memiliki kelas yang sama dengan dokumen 3 (D3) yaitu kelas IPA.

Tabel 7 *Euclidean Distance* $k = 5$

Rangking	<i>Euclidean Distance</i>	jurusan
1	2,82842	IPA
2	3,16227	IPA
3	3,31662	IPS
4	3,60555	IPA
5	3,70555	IPA

Sumber: Data Hasil Perhitungan Excel.

- Jadi berdasarkan *Euclidean* minat mayoritas terkecil adalah IPA, maka data uji memiliki kelas yaitu kelas IPA.

Tabel 8 *Euclidean Distance* $k = 7$

Rangking	<i>Euclidean Distance</i>	jurusan
1	2,82842	IPA
2	3,16227	IPA
3	3,31662	IPS
4	3,60555	IPA
5	3,70555	IPA
6	3,74165	IPA
7	3,87298	IPS

Sumber: Data Hasil Perhitungan Excel.

- Jadi berdasarkan *Euclidean* minat mayoritas terkecil adalah IPA, maka data uji memiliki kelas yaitu kelas IPA.

Tabel 9 *Euclidean Distance* $k = 9$

Rangking	<i>Euclidean Distance</i>	jurusan
1	2,82842	IPA
2	3,16227	IPA

3	3,31662	IPS
4	3,60555	IPA
5	3,70555	IPA
6	3,74165	IPA
7	3,87298	IPS
8	4	IPS
9	4,12310	IPA

Sumber: Data Hasil Perhitungan Excel.

- Jadi berdasarkan *Euclidean* minat mayoritas terkecil adalah IPA, maka data uji memiliki kelas yaitu kelas IPA.

Tabel 10 *Euclidean Distance* $k = 11$

Rangking	<i>Euclidean Distance</i>	jurusan
1	2,82842	IPA
2	3,16227	IPA
3	3,31662	IPS
4	3,60555	IPA
5	3,60555	IPA
6	3,74165	IPA
7	3,87298	IPS
8	4	IPS
9	4,12310	IPA
10	4,24264	IPA
11	4,24264	IPA

Sumber: Data Hasil Perhitungan Excel.

- Jadi berdasarkan *Euclidean* minat mayoritas terkecil adalah IPA, maka data uji memiliki kelas yaitu kelas IPA.

4. HASIL DAN PEMBAHASAN

a. Pengumpulan Data

Dilakukan pengumpulan data dengan mendatangi langsung instansi sekolah SMA Negeri 2 Situbondo, karena untuk mendapatkan data siswa baru hanya operator sekolah saja yang dapat mengakses situs penerimaan peserta didik baru (PPDB Jatim) dan melakukan wawancara dengan pihak bimbingan konseling (BK), untuk mengetahui alur penjurusan yang ada di sekolah tersebut. Digunakan data siswa baru yang mendaftar di SMA Negeri 2 Situbondo yang terdiri dari 350 data.

b. Data Set

Data yang dipakai dipenelitian ini ialah data nilai siswa tahun 2019/2020 berupa nilai rata-rata rapor SMP selama tiga tahun (nilai IPA, IPS, Bahasa Indonesia, Bahasa Inggris dan

Matematika) sebanyak 350 data. Atribut atau variable yang digunakan terdiri dari 7 atribut yang terdapat pada data SMA Negeri 2 Situbondo, atribut tersebut berupa nomor daftar,

nama Siswa, nilai IPA, nilai IPS, nilai Bahasa Inggris, nilai Bahasa Indonesia, dan nilai Matematika.

Tabel 11 Sampel Dataset

No Daftar	Nama	B.Ind	B.Ign	Ipa	Ips	Mtk	Minat
****	Anita Renias Nur Zakiyah	90	86	84	90	87	IPA
****	Siti Nur Lutfiyah	89	87	87	89	85	IPA
****	Neli Fadilatul Masrurroh	90	88	86	85	87	IPS
****	Abd. Mu'iz Samsul Arifin	80	88	90	89	88	IPA
****	Zainul Mu'in	89	89	85	85	87	IPS
****	Rovi Qotun Safira	90	89	87	81	87	IPA
****	Maechel Andrian Moudianto Yunior	90	86	83	86	88	IPA

Sumber: PPDB Online Guru Bimbingan Konseling.

c. Confusion Matrix

Proses penilaian dilakukan terhadap kumpulan dataset nilai siswa baru di SMA Negeri 2 Situbondo untuk memperoleh

performansi akurasi yang maksimal. Proses penilaian memakai *Confusion Matrix*. *Confusion Matrix* bermanfaat dalam menjabarkan mutu classifier untuk mengidentifikasi tuple di dalam kelas.

Tabel 12 Hasil K-Fold Cross Validation=5 Iterasi Ke-1 k=3

NO	No Daftar	NAMA	B.IND	B.ING	IPA	IPS	MTK	Minat	K-NN	Confusion Matrix
71	40103202900242	NOER ARCHIE WICAKSONO ARDIANSYAH	87	84	83	85	85	IPS	IPA	FP
72	40103202900310	MOCH. FAJAR MISWANTO	90	84	80	85	85	IPA	IPA	TP
73	40103202900228	CICI YULI ASTUTI	89	89	79	83	84	IPS	IPA	FP
74	40103202900107	AISYAH IMANIA	89	83	84	85	83	IPA	IPS	FN
75	40103202900028	ALIF IBRAHIM	88	83	80	90	83	IPA	IPA	TP
76	40103202900106	LUBIS NURIL UBADILLA	90	80	80	83	90	IPS	IPA	FP
77	40103202900207	IFTITAHIL FIRDAUSIYAH	90	76	85	84	88	IPS	IPS	TN
78	40103202900264	MOCHAMAD GIGIH LANANG PRAKOSO	84	83	79	89	88	IPS	IPA	FP
79	40103202900007	NIDYA MAULITA MAHMUDA	90	82	83	82	86	IPA	IPS	FN
80	40103202900353	JESIKA APRILIANA	86	83	86	83	85	IPA	IPA	TP
...

Sumber: Perhitungan Excel.

Tabel 13 Hasil Confusion Matrix k=3

Confusion Matrix		Nilai Sebenarnya	
		IPA	IPS
Nilai Prediksi	IPA	29	13
	IPS	6	8

Sumber: Akurasi Excel.

$$Akurasi = \frac{29+8}{29+13+6+8} \times 100\% = 66\%$$

$$Presisi = \frac{29}{29+13} \times 100\% = 69\%$$

$$Recall = \frac{29}{29+6} \times 100\% = 83\%$$

Tabel 14 Hasil Confusion Matrix k=5

Confusion Matrix		Nilai Sebenarnya	
		IPA	IPS
Nilai Prediksi	IPA	29	14
	IPS	6	7

Nilai Prediksi	IPA	25	14
	IPS	10	7

Sumber: Akurasi Excel.

$$Akurasi = \frac{25+7}{25+14+10+7} \times 100\% = 57\%$$

$$Presisi = \frac{25}{25+10} \times 100\% = 64\%$$

$$Recall = \frac{25}{25+7} \times 100\% = 71\%$$

Tabel 15 Hasil Confusion Matrix k=7

Confusion Matrix		Nilai Sebenarnya	
		IPA	IPS
Nilai Prediksi	IPA	29	14
	IPS	6	7

Sumber: Akurasi Excel.

$$Akurasi = \frac{29+7}{29+14+6+7} \times 100\% = 64\%$$

$$Presisi = \frac{29}{29+14} \times 100\% = 67\%$$

$$Recall = \frac{29}{29+7} \times 100\% = 83\%$$

Tabel 16 Hasil *Confusion Matrix* $k=9$

<i>Confusion Matrix</i>		Nilai Sebenarnya	
		IPA	IPS
Nilai Prediksi	IPA	28	14
	IPS	7	7

Sumber: Akurasi Excel.

$$Akurasi = \frac{28+7}{28+14+7+7} \times 100\% = 63\%$$

$$Presisi = \frac{28}{28+14} \times 100\% = 67\%$$

$$Recall = \frac{28}{28+7} \times 100\% = 80\%$$

Tabel 17 Hasil *Confusion Matrix* $k=11$

<i>Confusion Matrix</i>		Nilai Sebenarnya	
		IPA	IPS
Nilai Prediksi	IPA	29	14
	IPS	6	7

Sumber: Perhitungan Excel.

$$Akurasi = \frac{29+7}{29+14+6+7} \times 100\% = 64\%$$

$$Presisi = \frac{29}{29+14} \times 100\% = 67\%$$

$$Recall = \frac{29}{29+6} \times 100\% = 83\%$$

d. Hasil Fold Cross Validation 2, 3 dan 5 Dengan K-Nearest Neighbor (KNN)

Setelah melakukan perhitungan akurasi, presisi dan recall, maka akan di ambil akurasi terbaik untuk dijadikan data *training* data final, berikut adalah hasil keseluruhan perhitungan *Confusion Matrix*:

Tabel 18 Akurasi Setiap k pada Fold $K=2$

KNN		3NN		
K-Fold		Akurasi	Presisi	Recall
2Fold	Skenario 1	55%	68%	59%
	Skenario 2	58%	58%	78%
KNN		5NN		
K-Fold		Akurasi	Presisi	Recall
2Fold	Skenario 1	57%	67%	67%
	Skenario 2	54%	55%	80%
KNN		7NN		
K-Fold		Akurasi	Presisi	Recall
2Fold	Skenario 1	56%	66%	76%
	Skenario 2	55%	52%	83%
KNN		9NN		
K-Fold		Akurasi	Presisi	Recall
2Fold	Skenario 1	59%	68%	68%
	Skenario 2	54%	55%	87%
KNN		11NN		
K-Fold		Akurasi	Presisi	Recall

2Fold	Skenario 1	63%	72%	70%
	Skenario 2	53%	54%	91%

Sumber: Kumpulan Akurasi Excel.

Tabel 19 Akurasi Setiap pada Fold $K=4$

KNN		3NN		
K-Fold		Akurasi	Presisi	Recall
4Fold	Skenario 1	57%	63%	75%
	Skenario 2	53%	68%	55%
	Skenario 3	60%	64%	68%
	Skenario 4	56%	56%	67%
KNN		5NN		
K-Fold		Akurasi	Presisi	Recall
4Fold	Skenario 1	63%	66%	84%
	Skenario 2	53%	67%	60%
	Skenario 3	61%	64%	75%
	Skenario 4	53%	54%	61%
KNN		7NN		
K-Fold		Akurasi	Presisi	Recall
4Fold	Skenario 1	61%	65%	82%
	Skenario 2	56%	67%	66%
	Skenario 3	63%	63%	82%
	Skenario 4	56%	56%	67%
KNN		9NN		
K-Fold		Akurasi	Presisi	Recall
4Fold	Skenario 1	64%	67%	86%
	Skenario 2	54%	67%	62%
	Skenario 3	63%	63%	82%
	Skenario 4	56%	56%	64%
KNN		11NN		
K-Fold		Akurasi	Presisi	Recall
4Fold	Skenario 1	64%	67%	86%
	Skenario 2	57%	68%	68%
	Skenario 3	61%	63%	80%
	Skenario 4	61%	60%	72%

Sumber: Kumpulan Akurasi Excel.

Tabel 20 Akurasi Setiap k pada Fold $K=5$

KNN		3NN		
K-Fold		Akurasi	Presisi	Recall
5Fold	Skenario 1	66%	69%	83%
	Skenario 2	55%	66%	68%
	Skenario 3	61%	67%	71%
	Skenario 4	66%	67%	81%
	Skenario 5	48%	50%	48%
KNN		5NN		
K-Fold		Akurasi	Presisi	Recall
5Fold	Skenario 1	59%	66%	71%
	Skenario 2	59%	68%	70%
	Skenario 3	66%	67%	85%
	Skenario 4	61%	62%	78%
	Skenario 5	50%	52%	59%
KNN		7NN		
K-Fold		Akurasi	Presisi	Recall
5Fold	Skenario 1	64%	67%	83%
	Skenario 2	62%	70%	76%
	Skenario 3	68%	69%	85%
	Skenario 4	50%	55%	66%
	Skenario 5	52%	53%	59%

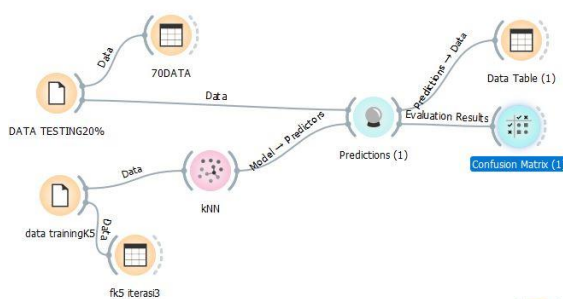
KNN		9NN		
K-Fold		Akurasi	Presisi	Recall
5Fold	Skenario 1	62%	67%	80%
	Skenario 2	61%	68%	72%
	Skenario 3	70%	70%	88%
	Skenario 4	48%	54%	69%
	Skenario 5	48%	50%	52%
KNN		11NN		
K-Fold		Akurasi	Presisi	Recall
5Fold	Skenario 1	66%	68%	86%
	Skenario 2	62%	69%	78%
	Skenario 3	64%	66%	85%
	Skenario 4	52%	56%	72%
	Skenario 5	52%	53%	59%

Sumber: Kumpulan Akurasi Excel.

Pada hasil algoritma *K-Nearest Neighbor* didapatkan akurasi tertinggi sebanyak 70% pada nilai *FoldK* = 5, skenario ke 3 di 9NN. Presisi tertinggi sebanyak 72% pada nilai *Fold K* = 2, skenario ke 1, 11NN. Untuk recall tertinggi sebanyak 91% pada nilai *Fold K* = 2, skenario ke 2 di 11NN. Skenario terbaik ditentukan berdasarkan akurasi tertinggi namun jika pada beberapa skenario akurasi memiliki nilai yang sama, maka skenario terbaik dipilih berdasarkan nilai presisi tertinggi.

e. Klasifikasi Data Final Dengan Metode K-NN

Setelah dilakukan pengujian dengan *Cross Validation* dan ditemukan 1 skenario terbaik. Selanjutnya, 1 skenario tersebut akan diuji dengan metode K-NN dimana data *testing* pada setiap skenario akan diganti dengan 70 data *testing* yang telah dipisah di awal. Berikut proses *Klasifikasi* dengan menggunakan metode K-NN menggunakan *software orange* ditunjukkan pada Gambar.



Gambar 3 Alur Orange

Sumber: Hasil gambar menurut alur Orange.

Berikut *Confusion Matrix* dari algoritma K-NN dengan *Fold K* = 5 pada skenario 3 dengan *k* = 9 yang dapat dilihat pada Gambar 4:

Gambar 4 Akurasi Orange

Sumber: Hasil gambar Orange.

Dapat di ketahui bahwa akurasi sebesar = $\frac{53+0}{53+15+2+0} \times 100\% = 76\%$, serta Presisi = $\frac{53}{53+15} \times 100\% = 78\%$ dan Recall = $\frac{53}{53+2} \times 100\% = 96\%$.

Tabel 21 Hasil perbandingan akurasi, presisi dan recall

KNN		9NN					
K Fold	Skenario	data testing Skenario 3			data final 70 data		
		Akurasi	Presisi	Recall	Akurasi	Presisi	Recall
5Fold	Skenario 3	70%	70%	88%	76%	78%	96%

Sumber: Akurasi Data Final.

Setelah mengambil akurasi terbaik dari hasil perhitungan *K-Fold* maka di dapat skenario terbaik ada pada 5-Fold skenario ke-3 dengan nilai tetangga terdekat adalah 9NN, maka data *testing* skenario ke-3 diganti dengan data final sebanyak 70 data dan didapatkan hasil akurasi, presisi dan recall mengalami peningkatan seperti yang telah ditunjukkan oleh Tabel.

5. KESIMPULAN DAN SARAN

a. Kesimpulan

Hasil akurasi yang paling tinggi dalam *Klasifikasi* penjurusan menggunakan algoritma K-NN sebanyak 70%, akurasi tersebut didapatkan pada *FoldK*=5 skenario ke-3. Skenario ini menghasilkan beberapa nilai presisi sebanyak 70% dan nilai recall sebanyak 88% dengan nilai *k* tetangga terdekat adalah 9-NN. Maka dapat disimpulkan skenario ini adalah skenario terbaik dari algoritma K-NN. Setelah mengetahui skenario terbaik maka data *testing* digantikan dengan data final sebanyak 70 data dan mendapatkan hasil akurasi sebanyak 76%, presisi sebanyak 78% dan recall sebanyak 96%.

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. Seperti *Modified K-Nearest Neighbor*, *Rough Set*, dan *Support Vector Machine* (SVM).
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 penjurusan siswa.

6. REFERENSI

- Cahayani, N. (2014). Pelaksanaan Bimbingan dan Konseling dalam Pelayanan Peminatan Akademik Kurikulum 2013 di SMAN I Sooko Mojokerto. *Jurnal BK UNESA*, 4(3).
- Deekshatulu, B. L., & Chandra, P. (2013). Classification of heart disease using K-Nearest Neighbor and genetic algorithm. *Procedia technology*, 10, 85-94.
- Hakim, L. (2017). Analisis perbedaan antara kurikulum KTSP dan kurikulum 2013. *JURNAL ILMIAH DIDAKTIKA: Media Ilmiah Pendidikan dan Pengajaran*, 17(2), 280-292.
- Han, J., & Kamber, M. (2006). Classification and prediction. *Data mining: Concepts and techniques*, 347-350.
- Ishak, A., Siregar, K., Ginting, R., & Afif, M. (2020). Orange Software Usage in Data Mining Classification Method on The Dataset Lenses. In *IOP Conference Series: Materials Science and Engineering* (Vol. 1003, No. 1, p. 012113). IOP Publishing.
- Nuh, M. (2013). Kurikulum 2013. Kebudayaan, Kementrian Pendidikan. "Materi Pelatihan Guru Implementasi Kurikulum 2013."
- Kristanto, O. (2014). Penerapan Algoritma Klasifikasi Data Mining ID3 Untuk Menentukan Penjurusan Siswa SMAN 6 Semarang. Universitas Dian Nuswantoro, Semarang.
- Mulyasa, H.E, (2016). Pengembangan Dan Implementasi Kurikulum 2013. Bandung: Remaja Rosdakarya. http://repo.unikadelasalle.ac.id/index.php?p=show_detail&id=11179&keywords. Diakses 11 Februari 2021
- Prasetyo, E., (2012). Data Mining, Konsep dan Aplikasi menggunakan Matlab, Yogyakarta: Andi Offset.
- Santosa, B. (2007). Data Mining Teknik Pemanfaatan Data Untuk Keperluan Bisnis. Yogyakarta: Graha Ilmu, 978(979), 756.
- Santra, A. K., & Christy, C. J. (2012). Genetic algorithm and Confusion Matrix for document clustering. *International Journal of Computer Science Issues* (IJCSI), 9(1), 322.
- Sikki, M. I. (2009). Pengenalan wajah menggunakan K-Nearest Neighbor dengan praproses transformasi wavelet. *Paradigma*, 10(2), 159-172.
- Soucy, P., & Mineau, G. W. (2001, November). A simple KNN algorithm for text categorization. In *Proceedings 2001 IEEE international conference on data mining*. 0-7695-1119-8/01 2001 IEEE. Department of Computer Science UniversitC Laval, QuCbec, Canada. <https://www.semanticscholar.org/paper/A-simple-KNN-algorithm-for-text-categorization-Soucy-Mineau/3cd7c6ae098c9634d160527c5418c824cf3a6da2>. Diakses 23 Februari 2021.
- Sunjana, (2010, juni). Aplikasi Mining Data Mahasiswa Dengan Metode Klasifikasi Decision Tree. ISSN: 1907-5022. Universitas Widyatama, Yogyakarta. <https://journal.uui.ac.id/Snati/article/download/1857/1633>. Diakses 23 Februari 2021.
- Suyanto, Novarina, S. A. 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.
- Wihardi. (2013). “K-Fold Cross Validation” [Online], <http://blog.yayaw.web.id/riset/k-Folds-cross-validation>. Diakses 22 Februari 2021.