

Implementasi Algoritma K-Nearest Neighbors Pada Penentuan Jurusan Siswa

M Daffa Alkhussayid*, Ferdiansyah

Fakultas ilmu komputer, Program Studi Teknik Informatika, Universitas Binadarma, Palembang, Indonesia

Email: ^{1,*}daffaalkhussayid29@gmail.com, ²ferdi@binadarma.ac.id,

Email Penulis Korespondensi: daffaalkhussayid29@gmail.com

Submitted: 28/08/2022; Accepted: 26/09/2022; Published: 30/09/2022

Abstrak—SMA Negeri 8 Palembang memiliki dua jurusan yakni IPA dan IPS. Penentuan jurusan dilakukan saat di kelas X untuk menentukan penjurusan siswa tersebut. Penelitian ini dilakukan dikarenakan guru mengalami kesulitan dalam menentukan peminatan jurusan siswa di SMA Negeri 8 Palembang yang memiliki peminatan penjurusan ada dua yakni IPA dan IPS yang ditujukan kepada siswa SMA kelas 10 dalam menentukan jurusan selanjutnya. Pada kasus penelitian, peneliti menggunakan klasifikasi dengan algoritma *k-nearest neighbors*, dan metode pengukuran jarak *euclidean distance* untuk memprediksi siswa-siswi dalam menentukan jurusan yang akan diambil oleh siswa. Sumber data penelitian ini berupa nilai raport di siswa kelas X SMA Negeri 8 Palembang, dan data penelitian ini ada 335 data siswa kelas X SMA Negeri 8 Palembang. Pengumpulan data yang diambil dari nilai raport mata pelajaran kelas X, ialah matematika, fisika, biologi, bahasa inggris, bahasa indonesia, sejarah, geografi, ekonomi, dan nilai psikotes. Dalam menentukan penjurusan siswa IPA dan IPS mendapatkan nilai rata-rata seluruh mata pelajaran dan nilai psikotes yang dihasilkan oleh siswa tersebut untuk masuk ke jurusan IPA dengan nilai rata-rata 80, nilai matematika 78, fisika 78, biologi 78, dan psikotes 80. Apabila siswa mendapat nilai rata-rata dibawah 80, maka akan di prediksi masuk IPS, untuk masuk ke dalam jurusan IPS dengan minimal nilai mata pelajaran geografi 70, ekonomi 70, sejarah 70, dan nilai psikotes 70. Hasil yang didapat dalam penelitian ini menggunakan metode *K-Nearest Neighbors* berdasarkan data training yang didapatkan dari data 335 data siswa, 101 diklasifikasikan kelas IPS sesuai prediksi, 10 data diprediksi IPS, tetapi data dinyatakan IPA, 2 data prediksi IPA dan, 222 data sesuai diprediksi IPA, dan untuk akurasi nya mendapatkan 96% dan hasil dari pengamatan menggunakan Website menggunakan K-NN menunjukkan data yang sama hasil yang di dapatkan melalui akurasi sebesar 96%.

Kata Kunci: SMA Negeri 8 Palembang; K-Nearest Neighbors; Prediksi Penjurusan; Machine Learning

Abstract—SMA Negeri 8 Palembang has two majors, namely science and social studies. Determination of majors is done when in class X to determine the majors of the student. This research was conducted because the teacher had difficulty in determining the majors at SMA Negeri 8 Palembang. In the research case, the researcher uses the classification with the *k-Nearest Neighbors* algorithm, and the Euclidean distance measurement method to predict the students in determining the majors that will be taken by students. The source of the data for this research is the report card scores for the X grade students of SMA Negeri 8 Palembang, and the data for this research are 335 data on the grade X students of SMA Negeri 8 Palembang. The data collection was taken from the grade X class report cards, namely mathematics, physics, biology, English, Indonesian, history, geography, economics, and psychological test scores. In determining the majors, students in science and social studies get the average score of all subjects and the psychological test scores produced by these students to enter the science department with an average score of 80, math score 78, physics 78, biology 78, and psychological test 80. If students get an average score below 80, it will be predicted to enter social studies, to enter the social studies department with a minimum score of 70 geography subjects, 70 economics, 70 history, and 70 psychological test scores. The results obtained in this study used the *K* method. - *Nearest Neighbors* based on training data obtained from 335 student data, 101 classified social studies class according to predictions, 10 data predicted social studies, but data declared natural science, 2 science prediction data and, 222 data according to natural science predictions, and the accuracy got 96% and the results of observations using the Website using *K-NN* show the same data results obtained through an accuracy of 96%.

Keywords: SMA Negeri 8 Palembang; K-Nearest Neighbors; Major Prediction; Machine Learning

1. PENDAHULUAN

Sekolah menengah atas (disingkat SMA) merupakan jenjang pendidik menengah pada pendidik formal di indonesia setelah lulus sekolah menengah pertama (disingkat SMP). Dimana sekolah menengah atas memiliki dua jurusan yaitu ipa dan ips. Penjurusan siswa yang dilakukan pada sekolah menengah atas sesuai minat yang dilakukan sejak peserta didik mendaftar SMA memiliki tujuan untuk memberikan kesempatan kepada siswa mengembangkan kompetensi sikap, kompetensi keterampilan siswa sesuai minat, bakat, dan kemampuan akademik dalam pelajaran keilmuan. Penentuan jurusan akan berdampak pada jenjang akademik berikutnya akan mempengaruhi bidang ilmu atau studi bagi siswa-siswi yang akan melanjutkan ke jenjang perguruan tinggi nantinya sehingga penjurusan yang tidak tepat bisa merugikan siswa pada masa depannya nanti[1].

Penempatan siswa sesuai dengan Menggunakan kapasitas kemampuannya atau sering disebut penjurusan peserta didik siswa baru di sekolah menengah ditentukan kemampuan akademik. Karena karakteristik suatu ilmu menuntut karakteristik yang sama dari mempelajarinya. Peserta didik siswa baru yang telah mempelajari suatu ilmu yang sesuai dengan ciri karakteristik kepribadiannya akan merasa senang. Pada waktu ini masih banyak peserta didik siswa baru sekolah menengah atas (SMA) yang masih ragu dalam memilih jurusan apa yang sesuai dengan kemampuan mereka masing-masing serta pihak sekolah masih melakukan proses penjurusan secara manual dan membutuhkan waktu yang sangat lama. Masalah ini sangat penting karena berpengaruh pada saat

peserta didik siswa baru memilih jurusan tersebut, oleh karena itu untuk mendapatkan keputusan yang tepat, SMA Negeri 8 Palembang untuk menentukan jurusan siswa sesuai dengan nilai, kecerdasan/iq, Matematika, Fisika, Biologi, kepribadian dan kemampuan yang dimiliki peserta didik siswa baru [2]

Apabila peserta didik siswa baru mengalami kesalahan dalam penjurusan kemungkinan yang akan terjadi rendahnya prestasi siswa atau dapat mengakibatkan terjadinya ketidakcocokan pada penjurusan. Permasalahan yang ada di SMA Negeri 8 Palembang ialah pada proses yang dilakukan dalam penentuan jurusan masih dilakukan dengan menggunakan *microsoft excel* dalam menghitung nilai-nilai siswa dalam menentukan jurusan. Cara ini bisa saja meleset dan tidak cocok dengan minat siswa dalam jurusan yang didapatkan menggunakan hasil *microsoft excel* secara manual, maka dari itu penulis berniat membangun sebuah sistem yang dapat membantu guru untuk menghitung nilai – nilai para siswa berdasarkan ketentuan sekolah. Pada kasus ini peneliti akan memakai klasifikasi dengan algoritma *k-nearest neighbors*, dan metode pengukuran jarak yang digunakan pada penulisan, penelitian ini menggunakan metode pengukuran jarak *euclidean distance*, untuk memprediksi siswa-siswi dalam menentukan jurusan yang akan diambil[3]. Diharapkan algoritma *k-nearest neighbors* dengan metode pengukuran jarak *euclidean distance* ini dapat menghasilkan tingkat akurasi yang baik dalam pengimplementasiannya pada data *mining* untuk menentukan jurusan siswa-siswi sma negeri 8 palembang[4].

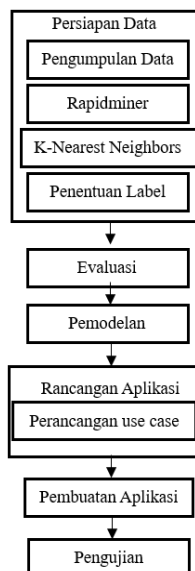
Terdapat penelitian yang relevan pada penelitian ini, ialah penelitian yang dilakukan oleh Habibi dan Santika (2020), dalam penelitiannya yang berjudul “Implementasi Algoritma *K-Nearest Neighbors* dalam Menentukan Jurusan Menggunakan Metode *Euclidean Distance* Berbasis Web Pada SMP Setia Gama”, permasalahan yang ada pada peneliti ini sebelumnya ialah pada siswa SMP Setia Gama bingung akan perihal jurusan yang harus mereka ambil dan guru yang menangani pemilihan jurusan masih manual melakukan penilaiannya. Perbedaan pada penelitian ini dengan peneliti sebelumnya ialah pada objek penelitian yang dilakukan peneliti sebelumnya siswa/i SMP Setia Gama. Persamaan pada peneliti ini dengan sebelumnya ialah menggunakan metode pengukuran jarak terdekat *Euclidean Distance* dan mengimplementasi pada algoritma *K-Nearest Neighbors* dengan Hasil yang didapatkan melalui klasifikasi prediksi berdasarkan nilai *accuracy*, *recall*, dan *precision* untuk mengetahui algoritma, maka didapatkan nilai *accuracy* paling tinggi sebesar 79,68% pada nilai $K = 7$, *recall* sebesar 85,00% dan *precision* sebesar 83,00% dengan nilai K yang sama.

Berdasarkan uraian diatas, maka peneliti tertarik mengambil permasalahan tersebut sebagai Penelitian ini penulis yang berjudul, “Implementasi Algoritma *K-Nearest Neighbors* pada Penentuan Jurusan Siswa SMA Negeri 8 Palembang”. Pada kasus penelitian, peneliti menggunakan klasifikasi dengan algoritma *k-nearest neighbors*, dan metode pengukuran jarak *euclidean distance* untuk memprediksi siswa-siswi dalam menentukan jurusan yang akan diambil oleh siswa. Diharapkan sistem penelitian yang dibangun menggunakan Algoritma *K-Nearest Neighbor* dengan metode pengukuran jarak *Euclidean Distance* ini dapat meningkatkan efisiensi dan efektifitas dalam membantu memprediksi penjurusan siswa/i SMA Negeri 8 Palembang.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Tahapan penelitian pada penentuan Jurusan SMA Negeri 8 Palembang menggunakan metode *K-Nearest Neighbors* terdiri dari 6 tahapan yaitu pengumpulan data, pemodelan, Evaluasi, Rancangan *Use Case*, pembuatan aplikasi, pengujian. Tahapan penelitian dapat dilihat pada Gambar 1.



Gambar 1. Tahapan Penelitian

Pada proses pengumpulan data yang dilakukan oleh peneliti yang didapatkan di sekolah SMA Negeri 8 Palembang, data yang didapatkan berupa data berbentuk *Excel*. Setelah mendapatkan data tersebut peneliti akan memasukkan data ke dalam Rapidminer untuk menggunakan metode *K-Nearest Neighbors* untuk di masukkan nilai K, setelah itu peneliti akan menentukan penentuan label, maka setelah menentukan penentuan label peneliti akan Evaluasi data untuk mendapatkan akurasi *Confusion Matrix*. Maka dari itu setelah mendatkan akurasi *confusion matrix*, peneliti akan membuat sebuah pemodelan dari hasil rapidminer ke *website*. Setelah mendapatkan model yang telah sesuai, untuk tahapan selanjutnya perancangan aplikasi dimana model akan diimplementasikan. Perancangan *Use Case* dimana proses jalannya sistem untuk *user*. Pada pembuatan aplikasi, peneliti menggunakan bahasa pemrograman *PHP* dan *MYSQL* dengan framework *Codeigniter*. Selanjutnya tahap pengujian dimana peneliti akan menguji data untuk mendapatkan hasil yang akurat.

2.2 K- Nearest Neighbors

K-Nearest Neighbors merupakan sebuah metode untuk klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya terdekat[5]. Tujuan dari Algoritma K-NN untuk mengklasifikasi objek baru berdasarkan atribut, dimana hasil dari sampel uji yang baru diklasifikasikan berdasarkan mayoritas dari kategori *K-Nearest Neighbors*[6]. Algoritma K-NN menggunakan klasifikasi ketetanggaan sebagai nilai prediksi dari sampel uji yang baru. Jarak yang digunakan adalah jarak *Euclidian Distance* [7]. *Euclidean Distance* didefinisikan sebagai berikut;

$$d = (\sum_{i=1}^n (x_i - y_i)^2) \tag{1}$$

Penjelasan :

d = Jarak *Euclidean* yang akan dicari

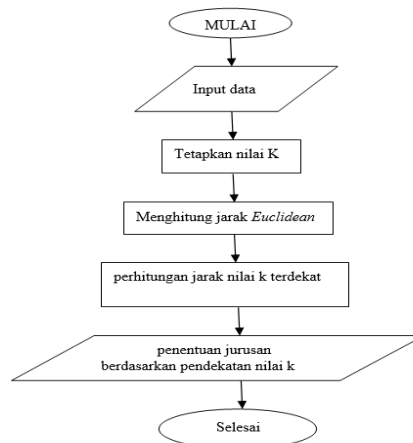
x_i = Nilai citra acuan yang telah disimpan di dalam *database*

y_i = Nilai citra uji

n = Jumlah data citra yang ada di dalam *database*

Berikut ini merupakan Flowchart Algoritma *K- Nearest Neighbors* dimana peneliti akan mengelola data[8] [9]

Tahapan peneliti menggunakan data dapat dilihat gambar 2.



Gambar 2. Flowchart Algoritma K-NN

Keterangan :

1. Menginput data

2. Masukkan nilai k

3. Menghitung jarak *Euclidean*

4. Setelah mendapatkan jarak *Euclidean* maka urutkan perhitungan nilai K jarak yang terdekat

5. Maka hasil yang menentukan jurusan dari pendekatan nilai K

2.3 Euclidian Distance

Metode *K-Nearest Neighbors* akan dilakukan perhitungan kuadrat jarak *Euclidean (Queryinstance)* masing-masing objek terdapat pada data sampel yang diberikan dalam perhitungan jarak dan ada beberapa rumus jarak contohnya seperti *Euclidean Distance*. Karena secara umum untuk mendefinisikan jarak antara objek x dan y, maka digunakan rumus jarak *Euclidean Distance* seperti dibawah ini [2].

$$d(x, y) = \sqrt{\sum_{i=1}^n (X_i - Y_i)^2} \tag{2}$$

Penjelasan :

X_i = Sampel data

Y_i = Data uji dan data testing

$d(x,y)$ = Jarak
 n = Dimensi data

2.4 Klasifikasi

Klasifikasi yaitu sebuah sistem yang melakukan proses menentukan target data input dengan benar, tetapi tidak mungkin bahwa kinerja suatu sistem tidak bisa seratus persen. Seluruh algoritma klasifikasi membuat model dengan tingkat akurasi tinggi, umumnya model ini digunakan untuk memprediksi data latih dengan benar. Pada proses penemuan model yang menggambarkan dan membedakan kelas data untuk tujuan memprediksi kelas dari objek penjurusan kelas tidak diketahui [10]. Proses klasifikasi didasari oleh 3 komponen;

1. Kelas, variabel yang berupa kategorikal yang mempresentasikan atau (label) yang terdapat pada objek contohnya IPA dan IPS.
2. *Predictor*, variabel yang dipresentasikan oleh karakteristik (atribut) data contohnya biologi, fisika, matematika, bahasa indonesia, bahasa inggris.
3. Data training satu set data yang berisi nilai yang digunakan untuk menentukan kelas yang cocok berdasarkan prediksi.

2.5 Evaluasi

Evaluasi dengan cara mengamati dan menganalisa hasil dari algoritma yang digunakan untuk memastikan bahwa pengujian benar. Pengujian dilakukan mengukur ke akuratan hasil, tujuan evaluasi untuk mengetahui mana yang terbaik dari hasil uji. Akurasi didefinisikan tingkat kedekatan antara prediksi dan nilai aktual, pengukuran akurasi terhadap model menggunakan *Confusion Matrix*[4]. Berikut ini tabel *Confusion Matrix* dapat dilihat di tabel 1.

Tabel 1. Confusion Matrix

Classification	Group	
	+	-
+	<i>True Positives</i>	<i>False Negatives</i>
-	<i>False Positives</i>	<i>True Negatives</i>

Perhitungan akurasi, presisi, recall *Confusion matrix*[11];

$$\text{Akurasi} = \frac{TP+TN}{TP+FP+FN+TN} \quad (3)$$

$$\text{Presisi} = \frac{TP}{TP+FP} \quad (4)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (5)$$

Keterangan:

- Akurasi Merupakan proporsi jumlah prediksi yang benar.
- Presisi adalah proporsi kasus hasil positif yang benar.
- TP(*True Positive*) adalah jumlah *record* positif diklasifikasikan sebagai positif oleh *classifier*.
- TN(*True Negative*) adalah jumlah *record* negatif yang diklasifikasikan sebagai negatif oleh *classifier*.
- FP(*False positive*) adalah jumlah *record* negatif yang diklasifikasikan sebagai positif oleh *classifier*.
- FN(*False Negative*) adalah jumlah *record* positif yang diklasifikasikan sebagai negatif oleh *classifier*.

2.6 Metodologi Penelitian

a. Pengumpulan Data

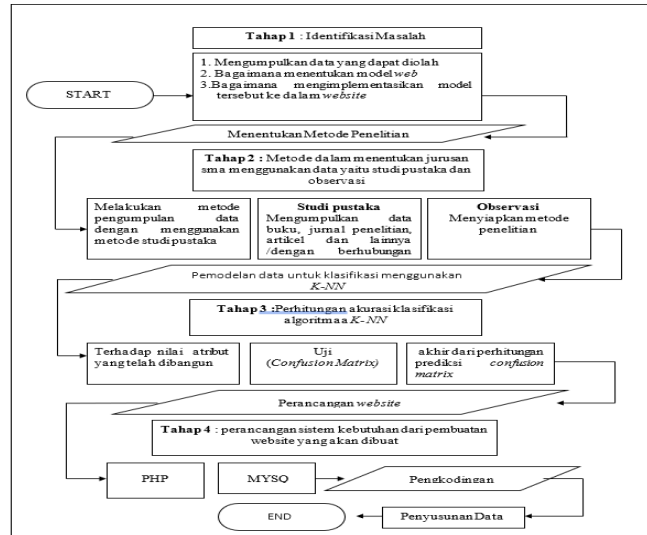
Data-data dengan cara membaca dan mempelajari buku-buku, jurnal, ataupun referensi lain yang berkaitan dengan penelitian yang dibahas. Studi pustaka pada penelitian ini yaitu mencari referensi dari *e-book* serta jurnal yang berkaitan dengan judul penelitian.

b. Dataset

Dataset merupakan pengumpulan data-data dengan cara melihat dari nilai mata pelajaran kelas x yaitu matematika, fisika, biologi, bahasa inggris, bahasa indonesia, sejarah, geografi, ekonomi, dan nilai psikotes. Maka untuk menentukan penjurusan IPA dan IPS melihat dari nilai rata-rata seluruh mata pelajaran dan nilai psikotes. Maka dari nilai raport kelas xi itu tidak dipakai, karena kelas xi telah masuk ke penjurusan, karena menentukan penjurusan itu dari kelas x. Dan untuk sumber dataset nya dari SMA Negeri 8 Palembang, diambil pada tanggal 20 Februari 2022 dengan persetujuan Sekolah SMA Negeri 8 Palembang. Dan untuk data nya primer yang telah diambil oleh peneliti secara langsung sehingga data yang didapatkan berupa data mentah.

2.7 Fase Penelitian

Fase penelitian ini merupakan dimana konsep alur yang akan dilakukan penelitian tersebut, hal ini dilakukan penelitian tersebut akan mudah di pahami. Adapun kerangka pemikiran dapat dilihat di gambar 3.



Gambar 3. Kerangka Pemikiran

Keterangan:

1. Identifikasi Masalah

Sebuah penelitian harus didasarkan pada masalah yang akan dipecahkan, pada tahap ini masalah diidentifikasi dan kemudian dicari solusi yang dapat menyelesaikannya. Inilah langkah-langkah yang harus dilakukan dalam memecahkan masalah tersebut:

- a. Menyediakan data yang bisa diolah untuk memprediksi.
- b. Menentukan model dan mendapatkan hasil dari hasil tersebut.
- c. Mengimplementasikan model tersebut ke dalam *website*.

2. Penentuan Metode penelitian

Sebuah penelitian memerlukan metode supaya penelitian tersebut cocok dan searah sehingga bisa menghasilkan suatu manfaat. Metode yang digunakan dalam dalam penelitian ini adalah studi pustaka dan observasi, dimana:

- a. Studi pustaka, mengumpulkan data-data dengan cara membaca buku, jurnal penelitian, artikel dan lainnya yang berhubungan dengan penerapan algoritma *k-nearest neighbors* pada penentuan jurusan SMA Negeri 8 Palembang.
- b. Observasi, pengumpulan data dengan cara mengamati langsung objek penelitian secara langsung untuk mengumpulkan data.

3. Perhitungan tingkat akurasi

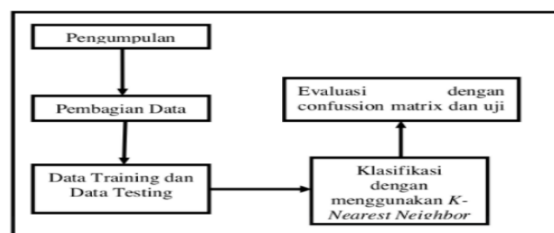
Pada tahapan ini model yang sudah dibangun ditentukan kemudian diuji evaluasi (*Confusion Matrix*) dengan menggunakan algoritma K-NN, yang nantinya untuk menentukan jurusan di SMA Negeri 8 Palembang.

4. Perancangan sistem *website*

Saat membuat sebuah *website*, penulis menggunakan pemrograman *PHP* dan *MYSQL* sebagai pengelolaan manajemen data di aplikasi *web* yang dibuat

2.8 Kerangka Pemikiran

Kerangka penelitian ini untuk menggambarkan secara umum, proses penelitian yang akan dilalui untuk menyelesaikan penelitian. Pada proses kerangka penelitian dapat dilihat pada gambar 4 [12]



Gambar 4. Kerangka Penelitian

Keterangan :

1. Awal penelitian dimulai dari tahapan pertama yaitu pengumpulan data dari SMA Negeri 8 Palembang.
2. Data nilai raport siswa tersebut kemudian di analisis untuk didapatkan data yang unik. Data yang telah diambil dari penelitian dalam bentuk *excel*. Proses seleksi data dengan cara mengelompokkan data yang berupa angka menjadi cukup dan tidak cukup.

- Untuk klasifikasi menggunakan *K-Nearest Neighbors* tersebut dengan cara membagi dua data yaitu *training* dan *testing*, *Training* yang akan digunakan data untuk latih, dan *testing* yang akan digunakan data sebagai uji data.
- Selanjutnya data tersebut akan di klasifikasi dengan melakukan akurasi *Confusion Matrix*.
- Melakukan pengecekan terhadap setiap nilai atribut dan model yang sudah dibangun. Lalu melakukan evaluasi menggunakan *Confusion matrix* dengan cara mengamati dan menganalisa hasil dari algoritma yang digunakan untuk memastikan bahwa hasil pengujian benar dan sesuai hasil pembahasan.

3. HASIL DAN PEMBAHASAN

3.1 Proses Penyelesaian Algoritma K-Nearest Neighbors

Pada tahap awal data tersebut akan dimasukkan kedalam rapidminer untuk melakukan penentuan label kelas, maka setelah melakukan penentuan label kelas data tersebut akan di kalsifikasikan K-Nearest Neighbors, selanjutnya data akan dibagi menjadi 2 yaitu *testing* dan *training*. Setelah data dibagi menjadi dua maka data tersebut akan dimasukkan kedalam metode *K-Nearest Neighbors* untuk memasukkan nilai K. Maka setelah memasukkan nilai K akan tampil hasil hitungan akurasi *Confusion Matrix*.

3.2 Pengumpulan Data

Penulis mendapatkan data dari sekolah SMA Negeri 8 Palembang, data nilai raport kelas X. Dalam penelitian tersebut, penulis mendapatkan data dalam bentuk excel dengan total 335 data. Pengumpulan data dapat dilihat pada tabel 2.

Tabel 2. Data Nilai Raport

Matematika	Biologi	Fisika	Bahasa Indonesia	Bahasa Inggris	Geografi	Sejarah	Ekonomi	psikotes	Rata-rata	Kelas
67	82	68	82	80	78	78	70	80	74	IPS
67	82	73	78	75	77	79	72	80	76	IPS
67	80	69	77	77	77	78	78	80	74	IPS
67	78	66	77	65	77	82	70	80	73	IPS
67	79	67	78	77	81	78	76	85	75	IPS

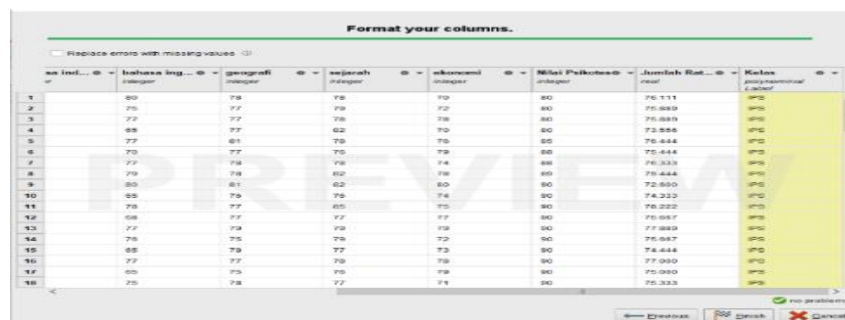
3.2.1 Kriteria Data Nilai

pengumpulan data-data dengan cara melihat dari nilai mata pelajaran kelas X yaitu matematika, fisika, biologi, bahasa inggris, bahasa indonesia, sejarah, geografi, ekonomi, dan nilai psikotes. Maka untuk menentukan penjurusan IPA dan IPS melihat dari nilai rata- rata seluruh mata pelajaran dan nilai psikotes. Maka dari nilai raport kelas XI itu tidak dipakai, karena kelas XI telah masuk ke penjurusan, karena menentukan penjurusan itu dari kelas X. Dalam Menentukan Jurusan SMA Negeri 8 Palembang, untuk masuk ke dalam jurusan IPA dengan nilai rata-rata 80 dan untuk nilai matematika 78, fisika 78, biologi 78 dan psikotes 80. Namun jika nilai rata-rata dibawah 80 maka akan di prediksi masuk IPS. Dan untuk masuk ke dalam jurusan IPS dengan minimal geografi 70, ekonomi 70, Sejarah 70 dan nilai psikotes 70.

3.3 Pengujian Data Menggunakan Rapidminer

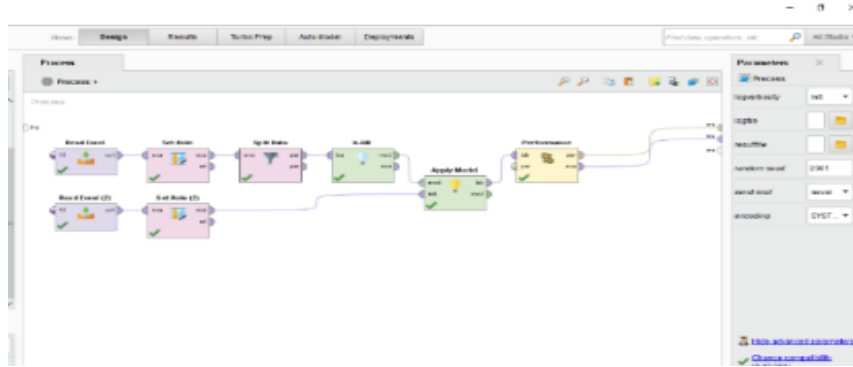
Dengan menggunakan metode *K-Nearest Neighbors*, memakai klasifikasi yang akan menghasilkan nilai akurasi yang telah kita tentukan dengan nilai k pada saat pengolahan klasifikasi data.[13]

- Pada tahap awal data akan di masukkan ke dalam rapidminer untuk melakukan penentuan label kelas, Penentuan label kelas dapat dilihat pada gambar 5.



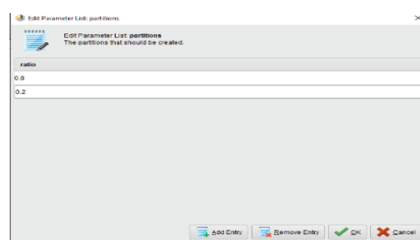
Gambar 5. Penentuan Label

- Setelah data melakukan penentuan label kelas, maka data tersebut akan di klasifikasi model *K-Nearest Neighbors*, klasifikasi model *K-Nearest Neighbors* dapat dilihat pada gambar 6.



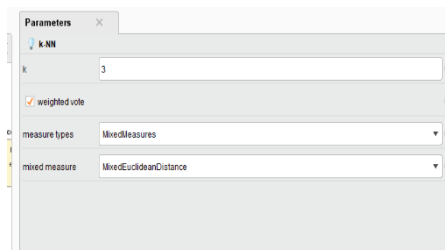
Gambar 6. Klasifikasi Model *K-Nearest Neighbors*

- Selanjutnya data tersebut akan di bagi dua data yaitu *testing* dan *training*, Pembagian data dapat dilihat pada gambar 7.



Gambar 7. Pembagian Data Training dan Testing

- Ketika data tersebut telah dibagi menjadi dua data, maka data tersebut akan di masukan ke dalam metode *K-Nearest Neighbors* untuk memasukkan nilai *K* untuk mendapatkan hasil hitungan Akurasi, menentukan nilai *k* dapat dilihat pada gambar 8.



Gambar 8. Menentukan Nilai *K* untuk Mengklasifikasi Data

- Tahap akhir, sesudah memasukkan nilai *K* maka akan tampil hasil hitungan akurasi *Confusion Matrix* Rapidminer, Hasil perhitungan akurasi *Confusion Matrix* dapat dilihat pada gambar 9. [14]

accuracy: 96.42%

	true IPS	true IPA	class precision
pred. IPS	101	10	90.99%
pred. IPA	2	222	99.11%
class recall	98.99%	95.69%	

Gambar 9. Hasil Hitungan Akurasi Rapidminer

3.4 Confusion Matrix

Dengan Hasil metode *K-Nearest Neighbors* berdasarkan data training, dari data 335 data siswa, 101 diklasifikasikan kelas IPS sesuai prediksi, 10 data diprediksi IPS tetapi ternyata IPA, 2 data prediksi IPA dan, 222 data sesuai diprediksi IPA, dan untuk akurasi nya mendapatkan 96%, dari hasil perhitungsn metode K-NN yang telah mendapatkan hasil akurasi sebesar 96% dapat dilihat tabel 3.[15].

Tabel 3. Confusion Matrix

Classification	Klasifikasi	
	IPS	IPA
IPS	101	10
IPA	2	222

$$\text{Accuracy} = \frac{101 + 222}{101 + 222 + 10 + 2} \times 100 = 96\%$$

accuracy: 96.42%			
	true IPS	true IPA	class precision
pred IPS	101	10	90.99%
pred IPA	2	222	99.11%
class recall	98.06%	95.69%	

Gambar 10. Hasil Confusion Matrix

3.5 Kondisi Situasional

Dalam kondisi situasi siswa memilih jurusan IPA, tetapi nilai mata pelajaran (Matematika kecil, Biologi kecil, Fisika kecil), tetapi nilai siswa (Geografi besar, Sejarah besar, Ekonomi besar) Maka siswa tersebut akan diprediksi kejurusan IPS. Dan untuk siswa memilih jurusan IPS tetapi nilai pelajaran (Matematika besar, Biologi besar, Fisika besar), dan siswa tersebut akan diprediksi masuk kejurusan IPA. Siswa dalam kondisi situasi dapat dilihat pada gambar 11.

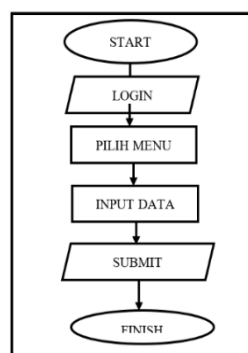
Kelas	predict...	confide...	confide...	No	Matema...	Biologi	fisika	bahasa ...	bahasa ...	geografi	sejarah	ekonomi	Nilai Psi...	Jumlah
IPA	IPS	0.500	0.500	33	70	67	77	80	75	60	79	74	91	91
IPA	IPS	0.500	0.500	39	70	83	77	80	80	81	78	79	91	80
IPA	IPS	0.500	0.500	44	70	84	77	77	74	82	80	72	91	81
IPA	IPS	0.650	0.342	47	67	66	76	78	77	61	78	77	91	90
IPS	IPA	0.328	0.674	73	75	71	71	80	85	79	79	85	92	77
IPA	IPS	0.500	0.500	78	76	77	76	80	71	77	79	71	94	81
IPA	IPS	0.751	0.249	80	61	74	72	80	78	77	78	76	94	90
IPA	IPS	0.500	0.500	82	79	74	77	78	80	78	87	73	94	81
IPA	IPS	0.500	0.500	84	83	75	77	80	80	77	79	90	94	82
IPA	IPS	0.642	0.358	103	71	79	79	82	79	61	78	66	94	91
IPA	IPS	0.608	0.392	205	72	71	81	81	78	80	77	70	99	81
IPS	IPA	0	1	248	65	78	73	79	77	65	81	70	102	80

Gambar 11. Nilai Siswa dalam Kondisi Situasi

Setelah melakukan membuat sebuah model dari rapidminer itu kemudian peneliti membuat aplikasi itu untuk mengimplementasikan model tersebut.

3.6 Rancangan Use Case

User (pengguna) menjelaskan mengenai proses jalannya sistem untuk user, tahap awal user akan login terlebih dahulu. Setelah login user akan di bawa ke tampilan Dashboard, lalu user akan memilih input data dan memasukkan data, maka setelah data dimasukkan, user akan dipilih otomatis jurusan dengan menggunakan metode K- Nearest Neighbors jarak dekat[16] Use Case dimana user akan mengenai proses jalannya suatu sistem yang dapat dilihat pada Gambar 12.



Gambar 12. Use Case

3.7 Pembuatan Aplikasi

Dalam pembuatan aplikasi, peneliti menggunakan bahasa pemrograman PHP dengan framework CodeIgniter, lalu menggunakan mysql untuk menyimpan data[17]. Digunakan sebagai interface untuk user melakukan penentuan jurusan dalam hal ini hamper serupa dengan system pakar berbasis web tetapi menggunakan algoritma machine learning

3.8 Tampilan K-Nearest Neighbors

Proses pengumpulan data yang dilakukan pengambilan data siswa SMA Negeri 8 Palembang dengan data sebanyak 350 dalam bentuk Excel, selain itu data tersebut akan diklasifikasi KNN (K-Nearest Neighbors) yang dilihat pada gambar 13.

```
if ($jumlah_benar <= 0) {
    // print array
    // dd($row) |> dd($data)
    $data['mtk'] = $row[1];
    $data['fisika'] = $row[3];
    $data['biologi'] = $row[2];
    $data['bi'] = $row[4];
    $data['inggris'] = $row[5];
    $data['geografi'] = $row[6];
    $data['sejarah'] = $row[7];
    $data['ekonomi'] = $row[8];
    $data['psikotes'] = $row[9];
    $data['rata_rata'] = $row[10];
    // dd($data)
    $data['kelas'] = $this->klasifikasi($data,3);
    // $jumlah_benar++;
    // if ($data['kelas'] == $data['klasifikasi']){
    //     $jumlah_benar++;
    // }
    $this->data_uji->insert($data);
}
}
}
// $akurasi = $jumlah_benar * 100 / $jumlah_soal;
session()->setFlashdata('message', 'Data Berhasil di Import dan di klasifikasi');
return redirect()->to('/admin/index_pengujian');
```

Gambar 13. Klasifikasi K-Nearest Neighbors

Setelah data di klasifikasi maka data tersebut akan di tentukan jarak *Euclidian Distance*, Codingan jarak *Euclidean Distance* dapat dilihat pada gambar 14.

```
}) {
    session()->setFlashdata('error', $this->validator->listErrors());
    // dd($this->validator->listErrors());
    return redirect()->back()->withInput();
} else {
    $data = [
        'mtk' => $this->request->getVar('mtk'),
        'fisika' => $this->request->getVar('fisika'),
        'biologi' => $this->request->getVar('biologi'),
        'bi' => $this->request->getVar('bi'),
        'inggris' => $this->request->getVar('inggris'),
        'sejarah' => $this->request->getVar('sejarah'),
        'ekonomi' => $this->request->getVar('ekonomi'),
        'geografi' => $this->request->getVar('geografi'),
        'psikotes' => $this->request->getVar('psikotes'),
        'rata_rata' => $this->request->getVar('rata_rata'),
    ];
    $mtk = $this->request->getVar('mtk');
    $fisika = $this->request->getVar('fisika');
    $biologi = $this->request->getVar('biologi');
    $psikologi = $this->request->getVar('psikotes');
    $nilai_rata_rata = ($mtk + $fisika + $biologi + $psikologi)/4;
    if($nilai_rata_rata > 80){
        $tambahan = " Menurut Rata Rata Anda dapat masuk ke : IPA " ;
    }else{
        $tambahan = " Menurut Rata Rata Anda dapat masuk ke : IPS " ;
    }
    $data['klasifikasi'] = $this->klasifikasi($data,5). " | ". $tambahan;
    // dd($data);
    $this->data_uji->insert($data);
    set_notifikasi_swal('success', 'Success', 'Hasil Klasifikasi Jurusan : '.$data['klasifikasi']);
    return redirect()->to(base_url('admin/index_pengujian'));
}
}
```

Gambar 14. Jarak *Eucliden Distance*

Dan untuk rumus menentukan jurusan tersebut nilai matematika, biologi, fisika dan psikologi ditambah maka akan menghasilkan nilai rata-rata. Codingan yang menghaikn nilai rata-rata pada penjurusan siswa dapat dilihat pada gambar 15.

```
public function klasifikasi($params,$k){
    $data_trainings = $this->data_training->get_data_training();
    foreach ($data_trainings as $row) {
        // dd($row)
        // $row = $row->toArray();
        $Euclidian_distance = 0;
        $Euclidian_distance += pow($row->fisika - $params['fisika'],2);
        $Euclidian_distance += pow($row->biologi - $params['biologi'],2);
        $Euclidian_distance += pow($row->bi - $params['bi'],2);
        $Euclidian_distance += pow($row->inggris - $params['inggris'],2);
        $Euclidian_distance += pow($row->sejarah - $params['sejarah'],2);
        $Euclidian_distance += pow($row->ekonomi - $params['ekonomi'],2);
        $Euclidian_distance += pow($row->psikotes - $params['psikotes'],2);
        $Euclidian_distance += pow($row->rata_rata - $params['rata_rata'],2);
        $Euclidian_distance = sqrt($Euclidian_distance);
    }
    $row->jarak = $Euclidian_distance;
}
//proses mengurutkan
usort($data_trainings,array($this->'custom_sort'));
//hitung sebanyak k
$jumlah_ipa = 0;
$jumlah_ips = 0;
for ($j = 0; $j < $k; $j++){
    if ($data_trainings[$j]->kelas == 'IPA') $jumlah_ipa++;
    if ($data_trainings[$j]->kelas == 'IPS') $jumlah_ips++;
}
//penentuan kelas
if ($jumlah_ipa > $jumlah_ips) return "IPA";
else return "IPS";
die(json_encode($data_trainings));
}
```

Gambar 15. Rumus menghasilkan nilai rata-rata

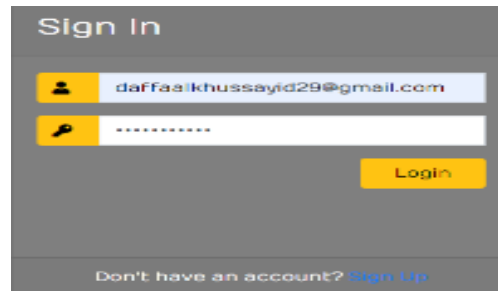
Apabila nilai rata-rata siswa 80 maka diprediksi IPA dan bisa untuk memasuki jurusan IPS Karena nilai rata-rata diatas 75, sedangkan nilai rata-rata siswa di bawah 80, maka akan diprediksi IPS untuk nilai rata-rata. Pada gambar 16 menjelaskan dimana siswa akan ditentukan jurusan IPA dan IPS melalui dari nilai rata-rata dengan menggunakan metode K-Nearest Neighbors.

```
if($nilai_rata_rata > 80){
    $tambahan = " Menurut Rata Rata Anda dapat masuk ke : IPA " ;
} else {
    $tambahan = " Menurut Rata Rata Anda dapat masuk ke : IPS " ;
}
$data['klasifikasi'] = $this->klasifikasi($data,5). " | ". $tambahan;
```

Gambar 16. Pelabelan Jurusan

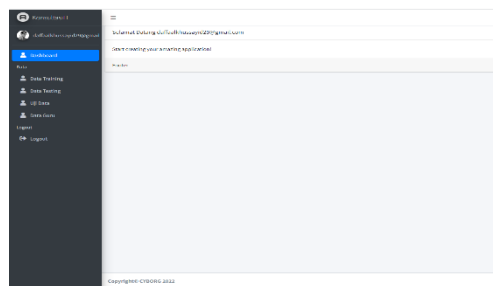
3.9 Tampilan Website

- a) Pada tahap awal *User* akan disuruh memasuki login pada di *Website* tersebut. Sebelum memasuki ke tampilan *Dashboard*, tampilan login dapat dilihat pada gambar 17.



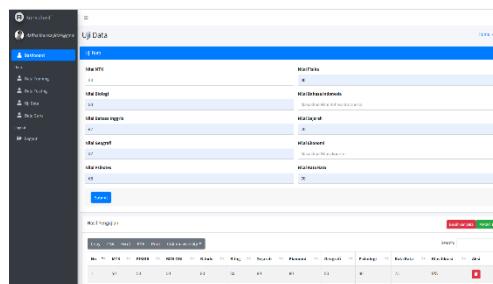
Gambar 17. Tampilan *Login*

- b) Setelah login user akan dibawa ke tampilan halaman *Dashboard* dimana ada banyak pilihan. Tampilan *Dashboard* dapat dilihat pada gambar 18.



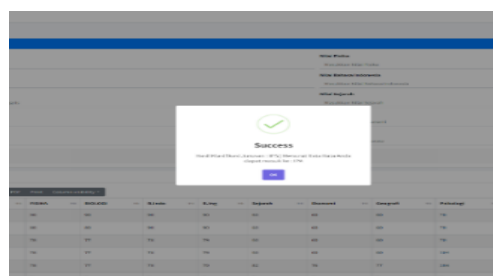
Gambar 18. Tampilan *Dashboard*

- c) Lalu itu *User* dapat menuju ke halaman uji data yang telah disediakan untuk memasukkan data-data nilai. Gambar 19 menjelaskan dimana *User* akan memasukkan data nilai untuk memprediksi masuk penjurusan IPA Dan IPS.



Gambar 19. Tampilan *Uji Data*

- d) Setelah *User* mengisi data-data nilai tersebut maka akan ditentukan jurusan dengan menggunakan metode *K-Nearest Neighbors*. Gambar 20 menjelaskan dimana user telah mengisi data nilai tersebut maka *user* akan di prediksi masuk ke penjurusan IPA Dan IPS dengan menggunakan metode *K-Nearest Neighbors*.



Gambar 20. Hasil Pemilihan Jurusan

- e) Dari hasil pengamatan menggunakan *Website*, dengan menggunakan *K-NN* telah mendapatkan akurasi sebesar 96% , Hasil akurasi menggunakan metode *K-Nearest Neighbors* dapat dilihat pada gambar 21.

Berhasil import excel dan mendapatkan akurasi :
96.987951807229 %

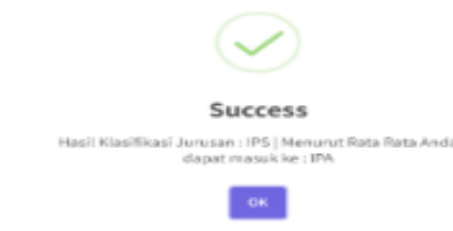


Gambar 21. Hasil Akurasi Menggunakan *Website*

- f) Dari hasil kondisi situasi nilai siswa matematika 81, fisika 80, biologi 79 dan nilai psikotes 85 maka siswa tersebut akan masuk ke jurusan IPA. Dari hasil klasifikasi nya siswa bisa masuk ke jurusan IPS. Sedangkan Untuk hasil kondisi situasi nilai siswa matematika 70, fisika 78, biologi 60 dan nilai psikotes 71 maka siswa tersebut akan masuk ke jurusan IPS. Dari hasil klasifikasi nya siswa tidak bisa masuk ke jurusan IPS. Hasil Menentukan jurusan siswa dalam nilai kondisi dapat dilihat pada gambar 22 dan gambar 23.



Gambar 22. Hasil Kondisi Situasi



Gambar 23. Hasil Kondisi Situasi

4. KESIMPULAN

Dari hasil penelitian terkait implementasi metode *K-Nearest Neighbors* pada penentuan jurusan SMA Negeri 8 Palembang dapat disimpulkan *metode K-Nearest Neighbors* berdasarkan data training yang didapatkan dari data 335 data siswa, 101 diklasifikasikan kelas IPS sesuai prediksi, 10 data diprediksi IPS, tetapi data dinyatakan IPA, 2 data prediksi IPA dan, 222 data sesuai diprediksi IPA, dan untuk akurasi nya mendapatkan 96% dari hasil rapidminer, sedangkan hasil dari pengamatan menggunakan *Website* menggunakan K-NN menunjukkan data yang sama memperoleh hasil melalui akurasi sebesar 96%, maka penelitian ini dapat digunakan dalam menentukan jurusan siswa/i SMA dengan akurat memilih IPA dan IPS sesuai dengan nilai rata-rata yang didapatkan oleh siswa/i tersebut. Untuk penelitian kedepan disarankan mencoba algoritma machine learning yang lain dengan tuning ataupun menambahkan hyperparameter agar hasil lebih akurat, dikarenakan peneliti hanya mencoba berdasarkan rujukan referensi sebelumnya dan model machine learning yang sudah ada.

REFERENCES

- [1] J. G. S. Km, "PENENTUAN JURUSAN SEKOLAH MENENGAH ATAS MENGGUNAKAN METODE K-NEAREST NEIGHBOR CLASSIFIER PADA SMAN 2 SINGINGI," vol. 3, no. 1, p. 8, 2020.
- [2] A. Sulistiyo, "Penentuan Jurusan Sekolah Menengah Atas Menggunakan Metode K-Nearest Neighbor Classifier Pada SMAN 16 Semarang," p. 5.
- [3] M. R. F. Risky Adit Pratama, *Analisis Big Data Sentimen Konsumen UMKM Sektor Kuliner Menggunakan Multi-Label K-Nearest Neighbor*. 2021.
- [4] S. T. R. Mustakim², *Perbandingan Teknik Pembagian Data untuk Klasifikasi Sarana Akses Air pada Algoritma K-Nearest Neighbor dan Naïve Bayes Classifier*. 2020.
- [5] P. Semmau, "PROSIDING SEMMAU 2017," p. 15, 2017.
- [6] M. Zunaidi and A. H. Nasyuha, "Analisis K-Nearest Neighbour Dalam Mendiagnosa Grandemultipara," vol. 3, p. 5, 2022.
- [7] C. Rahayuningtyas and D. S. Winarso, "Implementasi Algoritma k-Nearest Neighbor untuk Penjurusan Siswa SMA," *CAHAYATECH*, vol. 6, no. 2, p. 50, May 2019, doi: 10.47047/ct.v6i2.10.
- [8] A. M. Habibi, R. R. Santika, J. R. Ciledug, P. Utara, and K. Lama, "Implementasi Algoritma K-Nearest Neighbor dalam Menentukan Jurusan Menggunakan Metode Euclidean Distance Berbasis Web Pada SMP Setia Gama," vol. 3, no. 4, p. 8, 2020.
- [9] Y. Kustiyahningsih and N. Syafa'ah, "SISTEM PENDUKUNG KEPUTUSAN UNTUK MENENTUKAN JURUSAN PADA SISWA SMA MENGGUNAKAN METODE KNN DAN SMART," p. 10.

- [10] K. U. Syaliman bin Lukman and A. Labellapansa, “Analisa Nilai Lamda Model Jarak Minkowsky Untuk Penentuan Jurusan SMA (Studi Kasus di SMA Negeri 2 Tualang),” *J. Tek. Inform. Dan Sist. Inf.*, vol. 1, no. 2, Aug. 2015, doi: 10.28932/jutisi.v1i2.377.
- [11] R. K. Dinata, F. Fajriana, Z. Zulfa, and N. Hasdyna, “Klasifikasi Sekolah Menengah Pertama/Sederajat Wilayah Bireuen Menggunakan Algoritma K-Nearest Neighbors Berbasis Web,” *CESS J. Comput. Eng. Syst. Sci.*, vol. 5, no. 1, p. 33, Jan. 2020, doi: 10.24114/cess.v5i1.14962.
- [12] A. Y. Kencana and S. Astuti, “METODE KLASIFIKASI DENGAN ALGORITMA NAÏVE BAYES UNTUK REKOMENDASI PENJURUSAN SMA TERANG BANGSA,” vol. 15, no. 3, p. 6.
- [13] F. C. Anggoro, “PENERAPAN METODE K-NEAREST NEIGHBOUR UNTUK MENGANALISIS INVESTASI BUDIDAYA LOBSTER AIR TAWAR BERBASIS WEB,” vol. 3, no. 1, p. 6, 2019.
- [14] A. R. Kadafi, “Perbandingan Algoritma Klasifikasi Untuk Penjurusan Siswa SMA,” *J. ELTIKOM*, vol. 2, no. 2, pp. 67–77, Dec. 2018, doi: 10.31961/eltikom.v2i2.86.
- [15] A. Rahmawati and B. Hananto, “PREDIKSI PENENTUAN PROGRAM STUDI MENGGUNAKAN ALGORITMA K-NN PADA PELAJAR SMAN 6 KOTA DEPOK JURUSAN IPA,” p. 6, 2018.
- [16] U. Ramadhani and S. A. Putri, “Pemodelan Modified K-Nearest Neighbor Dalam Klasifikasi Jurusan Siswa Di SMAN 6 Pekanbaru,” p. 7, 2021.
- [17] I. Darmajaya, “Proseding Seminar Bisnis & Teknologi ISSN : 2407-6171,” p. 14, 2014.