

# Kombinasi Metode Naive Bayes dan K-Medoid dalam Memprediksi Penjurusan Siswa di Sekolah Menengah Atas

Devi Dwi Hariyanti<sup>a1</sup>, Gede Aditra Pradnyana<sup>a2</sup>, I Gede Mahendra Darmawiguna<sup>b3</sup>

<sup>a</sup>Jurusan Teknik Informatika, Universitas Pendidikan Ganesha  
Jalan Udayana no 11, Indonesia  
[1devidwihariyanti23@gmail.com](mailto:1devidwihariyanti23@gmail.com)

<sup>2</sup>gede.aditra@undiksha.ac.id (Corresponding author)

<sup>3</sup>mahendra.darmawiguna@undiksha.ac.id

## Abstract

*Majoring is a process of placement or distribution in the selection of teaching programs to students. The purpose of determining the majors is so that the lessons given to students are more focused. SMA Laboratorium Undiksha has problems in determining the majors and class division of students. The process of determining majors takes quite a long time, still using manual calculations in excel in determining students' majors. These problems can be overcome by developing a Student Class Majoring Prediction System Using a Combination of Naive Bayes and K-Medoid Methods. There are 22 criteria used in determining student class majors, namely gender, grades of report cards for semesters 3 to 5, math scores for semesters 3 to 5, science grades for semesters 3 to 5, social studies scores for semesters 3 to 5, Indonesian language grades for semester 3 to semester 5, grades in English for semester 3 to semester 5, student interests, and parents' interests. Based on the results of the calculation of the performance of the method, an accuracy value of 76% was obtained which indicates that the method used has a fairly good accuracy value in predicting student majors. The resulting precision of 84.33% indicates that the data category that is classified is in accordance with the actual category. The resulting recall of 70.67% indicates the success rate of the method in recognizing a category is good.*

**Keywords:** Prediction, Student Majoring, Naive Bayes, K-Medoid

## Abstrak

*Penjurusan merupakan suatu proses penempatan atau penyaluran dalam pemilihan program pengajaran kepada siswa. Tujuan dari penentuan penjurusan itu sendiri adalah agar kelak dikemudian hari pelajaran yang diberikan kepada siswa lebih terarah. SMA Laboratorium Undiksha memiliki permasalahan dalam penentuan penjurusan dan pembagian kelas siswa. Proses penentuan jurusan membutuhkan waktu yang cukup lama, masih menggunakan perhitungan manual di excel dalam menentukan penjurusan siswa. Permasalahan tersebut dapat diatasi dengan Pengembangan Sistem Prediksi Penjurusan Kelas Siswa Menggunakan Kombinasi Metode Naive Bayes Dan K-Medoid. Ada 22 kriteria yang digunakan dalam penentuan jurusan kelas siswa yaitu, jenis kelamin, nilai raport semester 3 sampai dengan semester 5, nilai matematika semester 3 sampai dengan semester 5, nilai ipa semester 3 sampai dengan semester 5, nilai ips semester 3 sampai dengan semester 5, nilai bahasa indonesia semester 3 sampai dengan semester 5, nilai bahasa inggris semester 3 sampai dengan semester 5, minat siswa, dan minat orang tua. Berdasarkan hasil perhitungan kinerja metode, didapatkan nilai akurasi sebesar 76% yang menunjukkan bahwa metode yang digunakan memiliki nilai akurasi yang cukup baik dalam memprediksi jurusan siswa. Precision yang dihasilkan sebesar 84,33% menunjukkan kategori data yang diklasifikasi telah sesuai dengan kategori yang sebenarnya. Recall yang dihasilkan sebesar 70,67% menunjukkan tingkat keberhasilan metode dalam mengenali suatu kategori sudah baik.*

**Kata Kunci:** Prediksi, Penjurusan Siswa, Naive Bayes, K-Medoid

## 1. Introduction

Penjurusan merupakan suatu proses penempatan atau penyaluran dalam pemilihan program pengajaran kepada siswa [1]. Penjurusan ini diadakan karena yang akan menentukan keberhasilan para siswa, baik pada waktu belajar di SMA maupun setelah perguruan tinggi maka diperlukan suatu bimbingan penjurusan. Williamson berpendapat bahwa di dalam penjurusan ini terdapat kaitan yang erat antara bimbingan penjurusan dengan bimbingan karir, yaitu merupakan suatu proses yang bebas, meluas dan berurutan [2]. Sekolah Menengah Atas (SMA) merupakan salah satu lembaga pendidikan yang mulai memperkenalkan jurusan dan membaginya dalam beberapa pilihan jurusan [1].

SMA Laboratorium Undiksha merupakan salah satu sekolah yang menerapkan proses penjurusan bagi siswanya untuk dibagi menjadi 3 jurusan yaitu, IPA, IPS dan Bahasa. Kurikulum yang digunakan SMA Laboratorium Undiksha saat ini ialah Kurikulum 2013, yang mengatur proses penjurusan dengan menggunakan nilai raport SMP, minat siswa dan minat orangtua. Penjurusan mulai dilakukan pada saat duduk di bangku kelas X (sepuluh). Bagi siswa baru, penentuan penjurusan merupakan hal yang harus mereka alami ketika memilih melanjutkan pendidikan di jenjang SMA. Inilah tahap yang sangat strategis karena memilih jurusan berarti menentukan masa depan. Namun pada kenyataannya, banyak siswa yang memilih jurusan bukan berdasarkan bakat yang dimilikinya. Terkadang, keputusan para siswa dipengaruhi oleh pendapat orangtua, teman atau figur-figur yang diidolakan.

Selain penentuan jurusan, pembentukan atau pembagian kelas siswa juga merupakan hal yang sangat penting dalam rangka meningkatkan kualitas proses belajar mengajar yang dilakukan. [3] menyatakan bahwa hal tersebut dikarenakan perbedaan kemampuan yang dimiliki oleh siswa di setiap kelasnya yang dapat berdampak pada tidak efektifnya proses pembelajaran yang berlangsung. Permasalahan yang umum terjadi saat pembentukan kelas siswa adalah perbedaan kemampuan yang dimiliki oleh siswa di setiap kelasnya. Pengelompokan siswa dengan kemampuan yang sama merupakan hal yang sangat penting dalam rangka meningkatkan kualitas proses belajar mengajar. Dengan pengelompokan siswa di kelas yang sesuai, mereka akan dapat saling membantu dalam proses pembelajaran.

Tujuan dari penentuan penjurusan itu sendiri adalah agar kelak dikemudian hari, pelajaran yang diberikan kepada siswa lebih terarah, karena tidak jarang juga siswa-siswi yang asal-asalan dalam menentukan jurusan yang mereka ambil. Begitu pula pada pembentukan kelas siswa adalah untuk mengelompokkan siswa yang memiliki kemampuan yang sama agar memudahkan proses pembelajaran.

Seiring perkembangan teknologi hal tersebut dapat diatasi dengan teknik klasifikasi data dengan data mining. Data mining atau dalam bahasa Indonesia disebut penggalian data merupakan suatu proses pencarian korelasi, pola dan tren baru yang berguna dalam media penyimpanan data berukuran besar menggunakan teknologi pengenalan pola seperti teknik-teknik statistik dan matematis [4]. Klasifikasi adalah proses untuk menemukan model atau fungsi yang menjelaskan atau membedakan konsep atau kelas data, dengan tujuan untuk dapat memperkirakan kelas dari suatu objek yang labelnya tidak diketahui [4]. Dalam teknik klasifikasi dengan data mining terdapat beberapa metode yang dapat digunakan yaitu C4.5, Rainforest, Naïve Bayesian, neural network, genetic algorithm, fuzzy, case-based reasoning, dan K-Nearest Neighbor. Clustering adalah proses pengelompokkan kumpulan data menjadi beberapa kelompok sehingga objek di dalam satu kelompok memiliki banyak kesamaan dan memiliki banyak perbedaan dengan objek dikelompok lain [4]. Clustering memiliki beberapa metode yang dapat digunakan dalam mengelompokkan sebuah data diantaranya, K-Means, K-Medoids, Self-Organizing Map (SOM) dan Fuzzy C-Means.

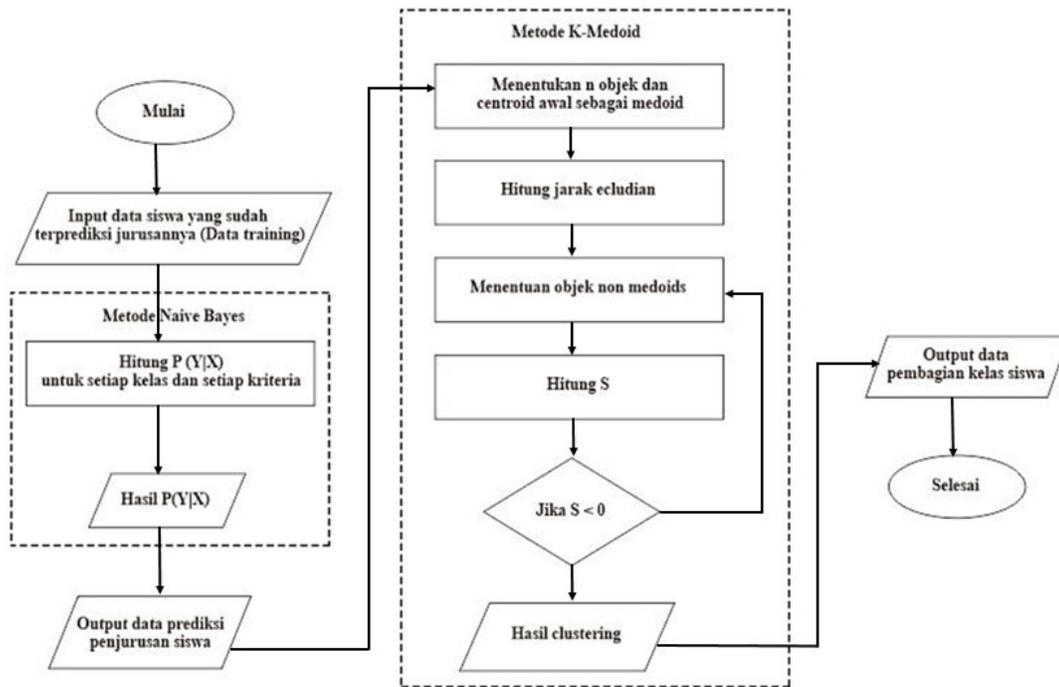
Pada penelitian ini, dirancang sebuah metode baru untuk prediksi penjurusan kelas siswa dengan mengkombinasikan algoritma Naive Bayes dan K-Medoid. Menurut penelitian yang dilakukan oleh [5] dengan judul "Optimalisasi K-Medoid Dalam Pengklasteran Mahasiswa Pelamar Beasiswa Dengan Cubic Clustering Criterion", K-Medoid lebih unggul dibandingkan k-means dalam mendeteksi outlier. K-medoids tidak menentukan nilai rata-rata dari objek dalam cluster sebagai titik acuan, tapi menggunakan medoid (median), yang merupakan objek yang paling

terletak dipusat sebuah cluster. Dengan demikian, metode partisi masih dapat dilakukan berdasarkan prinsip meminimalkan jumlah dari ketidaksamaan antara setiap objek dan titik acuan yang sesuai (medoid). K-Medoid juga dapat digunakan untuk dataset dengan nilai domain kontinyu maupun diskrit sedangkan k-means hanya cocok untuk kasus domain kontinyu. Menurut penelitian [6] tentang "Penerapan Metode K-Medoid pada Pengelompokan Rumah Tangga Dalam Perlakuan Memilah Sampah Menurut Provinsi" menyatakan bahwa K-Medoid cukup efisien untuk datase yang kecil. Algoritma K-Medoid hadir untuk mengatasi kelemahan algoritma K-Means yang sensitif terhadap outlier. Naive Bayes Classifier merupakan salah satu metode yang digunakan dalam data mining yang didasarkan pada teori keputusan bayes [9]. Menurut penelitian yang dilakukan oleh [7] pada "Metode Klasifikasi dengan Algoritma Naive Bayes Untuk Rekomendasi Penjurusan SMA Terang Bangsa", kelebihan metode Naive Bayes adalah sederhana tetapi memiliki akurasi tinggi. [8] juga menyatakan bahwa Naive Bayes merupakan metode statistik sederhana dan memiliki akurasi yang baik dalam proses pengklasifikasian pada penelitiannya yang berjudul "Klasifikasi Jurusan Menggunakan Metode Naive Bayes Pada Sekolah Menengah atas Negeri (SMAN) 1 Fatuleu Tengah". Maka dari itu, dengan mengkombinasikan algoritma K-Medoid dan Naive Bayes ini dapat menghasilkan prediksi yang lebih spesifik, bukan hanya menentukan jurusan akan tetapi sampai pada kelas yang diperoleh. Dengan pengelompokan siswa di kelas yang sesuai, mereka akan dapat saling membantu dalam proses pembelajaran. Selain itu, membagi kelas siswa sesuai dengan kemampuannya dapat mempermudah tenaga pendidik dalam menentukan metode atau strategi pembelajaran yang sesuai.

Faktor utama dalam menentukan penjurusan kelas siswa pada SMA Laboratorium Undiksha adalah nilai rata-rata raport SMP, nilai raport IPA SMP, nilai raport Matematika SMP, nilai raport IPS SMP, nilai raport Bahasa Indonesia SMP, nilai raport Bahasa Inggris, peminatan siswa dan peminatan orangtua. Metode Naive Bayes digunakan untuk memprediksi jurusan siswa berdasarkan komponen atribut yang telah ditentukan. Sedangkan algoritma K-Medoid digunakan untuk pembentukan kelas siswa yang telah terprediksi jurusannya menggunakan metode Naive Bayes

## **2. Reseach Methods**

Tahapan-tahapan dari sistem prediksi penjurusan kelas siswa dengan kombinasi metode Naive Bayes dan K-Medoid yang dikembangkan dapat dilihat pada gambar 1. Tahapan awal adalah memproses data siswa yang sudah terprediksi jurusannya untuk dijadikan data training dalam proses prediksi dengan metode Naive Bayes. Hasil dari proses ini adalah prediksi jurusan masing-masing siswa yang akan diprediksi kelasnya menggunakan metode K-Medoid.



Gambar 1. Flowchart Sistem Prediksi Penjurusan Kelas Siswa Menggunakan Kombinasi Metode Naive Bayes dan K-Medoid

Tabel 1. Data Training

JK	RAPORT			MTK		
	3	4	5	3	4	5
L	82	82	82	78	78	78
	IPA			IPS		
	3	4	5	3	4	5
	80	77	83	82	75	76
	BIN			BING		
	3	4	5	3	4	5
	78	78	79	77	77	80
	MINAT					
	1	2	ORTU	HSL		
	IPA	BHS	IPA	IPA		

Tabel 2. Data Testing

JK	RAPORT			MTK		
	3	4	5	3	4	5
L	77,7	77,7	78,7	75	77	76
	IPA			IPS		
	3	4	5	3	4	5
	76	76	79	78	78	79
	BIN			BING		
	3	4	5	3	4	5
	77	77	80	78	76	77
	MINAT					
	1	2	ORTU			

IPA	BHS	IPA
-----	-----	-----

Berikut ini adalah contoh proses yang dilakukan dalam sistem prediksi penjurusan kelas siswa yang akan dikembangkan:

- a. Input data training yang telah terprediksi jurusannya. Diberikan data training sebanyak 84 data yang terletak pada tabel di bawah ini.
- b. Hitung jumlah kelas dan peluang setiap atribut. Jika jenis atribut data numerik, Maka hitung mean dan standar deviasi
- c. Hitung probabilitas masing-masing kriteria berdasarkan nilai data testing. Untuk data numerik hitung probabilitas menggunakan rumus densitas gauss

$$= \frac{1}{\sqrt{2\pi \times 3.14 \times 3.291}} e^{-\frac{(77,7-82,871)^2}{2 \times (3,291)^2}}$$

$$= 6,526$$

- d. Hitung probabilitas masing-masing kelas/konsentrasi dengan cara mengalikan hasil probabilitas per kriteria.

$$P(\text{IPA}) = \frac{15541941892}{15541941892+2,97+13149,41} = 0,99$$

$$P(\text{IPS}) = \frac{2,97}{15541941892+2,97+13149,41} = 1,91$$

$$P(\text{BAHASA}) = \frac{13149,41}{15541941892+2,97+13149,41} = 8,46$$

Sehingga hasil prediksi siswa 1 yaitu masuk ke kelas/konsentrasi BAHASA.

- e. Setelah hasil prediksi diketahui, dilanjutkan untuk penentuan kelas yang diperoleh oleh siswa. Pada contoh kali ini hanya menggunakan kelas/konsentrasi IPA. Maka dari itu, pilih nama siswa beserta kriterianya yang terprediksi di kelas/konsentrasi IPA kemudian akan dihitung menggunakan algoritma K-Medoid. Berikut adalah tabel hasil prediksi kelas/konsentrasi IPA:
- f. Tentukan (k) jumlah kluster dari n objek yaitu sebanyak 2 dengan keterangan C1 = IPA 1 dan C2 = IPA 2.
- g. Menentukan centroid awal sebagai medoids.
- h. Tempatkan objek-objek non medoids ke dalam *Cluster* yang paling dekat dengan medoids berdasarkan jarak euclidean. Berikut contoh perhitungan jarak pada data ke 1:

$$c(x_1, c_1) = \sqrt{(1-1)^2 + (87,2-87,2)^2 + (88,1-88,1)^2}$$

$$\sqrt{+(88,8-88,8)^2 + (87-87)^2 + (87-87)^2 +}$$

$$\sqrt{(92-92)^2 + (94-94)^2 + (84-84)^2 + (85-85)^2}$$

$$\sqrt{+(83-83)^2 + (91-91)^2 + (92-92)^2 + (85-85)^2}$$

$$\sqrt{+(85-85)^2 + (88-88)^2 + (85-85)^2 +}$$

$$\sqrt{(90-90)^2 + (90-90)^2 + (0100-0100)^2}$$

$$\sqrt{+(0001-0001)^2 + (0100-0100)^2} = 0$$

- i. Tentukan objek non medoids
- j. Tentukan objek-objek non medoids ke dalam *Cluster* yang paling dekat dengan non medoids berdasarkan jarak euclidean. Berikut adalah hasil perhitungan jarak ke setiap non medoids.

Tabel 3. Hasil perhitungan jarak ke setiap medoids

data ke i	d1	d2	Terdekat	Klaster yg diikuti
d(x1)=	22,34052	36,0536	22,34052	1
d(x2)=	29,79242	23,03483	23,03483	2
d(x3)=	0	32,84031	0	1
d(x4)=	32,84031	0	0	2
Nilai Cost	22,34052	23,03483		
T. Cost	45,37536			

Tabel 3 merupakan tabel hasil perhitungan jarak ke setiap medoids. Dari jarak 2 cluster tersebut dipilih jarak yang paling terdekat dan jarak tersebutlah yang menjadi cluster yang diikuti.

k. Hitung nilai S

$$S = \text{Total cost baru} - \text{total cost lama} = 45,37536 - 45,37536 = 0$$

l. Karena nilai  $S > 0$  maka proses pengClusteran berhenti. Sehingga diperoleh anggota tiap Cluster sebagai berikut

Tabel 4. Hasil pengclusteran

D1	1	➔	Siswa 1	1
D2	2		Siswa 2	2
D3	1		Siswa 3	1
D4	2		Siswa 4	2

Tabel 4 merupakan hasil cluster. Dari 4 data jurusan Ipa diperoleh 2 siswa di cluster 1 dan 2 siswa di cluster 2. Hal tersebut menunjukkan bahwa siswa di cluster 1 tergolong ke kelas Ipa 1 dan siswa di cluster 2 tergolong ke kelas Ipa 2.

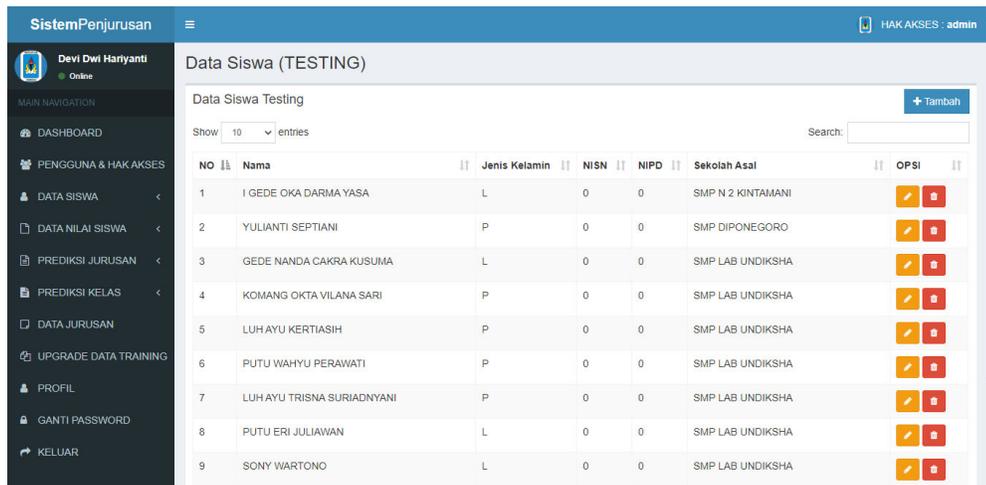
### 3. Result and Discussion

#### A. Pengembangan Sistem Prediksi Penjurusan Kelas Berbasis Web

Setelah melakukan perancangan, tahap selanjutnya adalah tahap implementasi atau pengembangan sistem. Pengembangan sistem mengacu pada hasil dari analisis kebutuhan dan proses perancangan. Sistem dikembangkan dengan bahasa pemrograman PHP dan basis data MySQL. Gambaran antar muka dari sistem yang dikembangkan dapat dilihat pada Gambar 4 sampai dengan Gambar 5.

Gambar 4 menunjukkan fitur fitur manajemen data siswa yang akan diprediksi jurusan dan kelasnya. Penambahan data siswa dapat dilakukan secara manual yaitu dengan menginputkan satu persatu. Selain data siswa, data nilai siswa juga diinputkan di sistem sesuai gambar no 5.

Proses penentuan jurusan dimulai dengan melakukan perhitungan rumus, dilanjutkan dengan proses prediksi jurusan hingga muncul hasil proses prediksi jurusan selesai seperti gambar no 6. Sedangkan pada proses pembagian kelas, dimulai dengan memilih fitur proses prediksi kelas hingga muncul hasil prediksi kelas siswa seperti gambar 7.



Gambar 4. Manajemen Data Siswa

**SistemPenjurusan** | HAK AKSES: admin

Devil Dwi Haryanti | Online

### Data Nilai Siswa (TESTING)

Data Nilai Siswa + Buat Data Nilai Siswa Baru

Show 10 entries Search:

NO	Nama	IPA 3	IPA 4	IPA 5	IPS 3	IPS 4	IPS 5	MTK 3	MTK 4	MTK 5	BIN 3	BIN 4	BIN 5	BING 3	BING 4
1	I GEDE OKA DARMA YASA	72.00	74.00	80.00	76.00	80.00	82.00	77.00	80.00	80.00	76.00	80.00	82.00	76.00	80.00
2	YULIANTI SEPTIANI	75.00	78.00	89.00	85.00	90.00	90.00	86.00	92.00	85.00	84.00	89.00	94.00	85.00	86.00
3	GEDE NANDA CAKRA KUSUMA	72.00	71.00	80.00	85.00	79.00	86.00	75.00	72.00	82.00	80.00	72.00	79.00	86.00	90.00
4	KOMANG OKTA VILANA SARI	70.00	75.00	80.00	81.00	80.00	88.00	79.00	72.00	82.00	78.00	79.00	79.00	80.00	86.00
5	LUH AYU KERTIASIH	70.00	76.00	79.00	82.00	88.00	85.00	78.00	72.00	82.00	79.00	81.00	80.00	80.00	86.00
6	PUTU WAHYU	73.00	71.00	79.00	84.00	84.00	84.00	72.00	72.00	82.00	76.00	72.00	79.00	80.00	79.00

Gambar 5. Manajemen Data Nilai Siswa

**SistemPenjurusan** | HAK AKSES: pengguna

Ahmad Neil Abror | Online

### Data Prediksi Jurusan Hasil Peminatan Siswa

Daftar Prediksi Jurusan

Show 10 entries Search:

NO	NISN	Nama	Hasil IPA	Hasil IPS	Hasil BAHASA	Prediksi Jurusan
1	0	GEDE NANDA CAKRA KUSUMA	102184	15810165.419079	148296118.97796	BAHASA
2	0	PUTU WAHYU PERAWATI	473754	9515.6505503835	103485331.36815	BAHASA
3	0	PUTU ERI JULIAWAN	740865	10889782531.837	3202041487997.1	BAHASA
4	0	LINDA SURYANI WIJAYA	211708	1047.4289550007	13025129.74937	BAHASA
5	0	I PUTU SATRIA WIBAWA	379539	77536156.199103	387025686.32503	BAHASA
6	0	PUTU IVAN JUNI ANTARA	165281	2669408.244286	3430962645.2382	BAHASA
7	0	KADEK DWIKI SUWITA KHARISMAWAN	110600	49.966313325196	701226.27466488	BAHASA
8	0	PUTU CAROLINA RATALIA PUTRI	747.31	134.97072639213	991.03220182484	BAHASA
9	0	JOANNE ELVARETTA SUNYOTO	102237	1076307300.0464	5006591463.9142	BAHASA
10	0	NI KOMANG GINA ASTUTI	355169	3311463.6967288	38316487.459078	BAHASA

Showing 1 to 10 of 209 entries Previous 1 2 3 4 5 ... 21 Next

Gambar 6. Hasil Prediksi Jurusan Siswa

**SistemPenjurusan** | HAK AKSES: admin

Devil Dwi Haryanti | Online

### Data Pembagian Kelas Manajemen Kelas Siswa

Daftar Siswa Kelas IPA I

NO	NISN	Nama
1	0	NI LUH MEILIANI
2	0	IDA BAGUS APSARA
3	0	KADEK MILA ARISANTINI
4	0	KADEK MEGAWATI
5	0	HADI PUTRA PERMANA
6	0	KADEK DENI MAHARDIKA
7	0	FAIZAL SALAM
8	0	PUTU ADI PUTRA WAHYUDI
9	0	I KADEK G. LANANG SURYA DININGRAT
10	0	SHERREN SANCHIA TANAYA
11	0	I KOMANG PUJI DARMAWAN
12	0	I WAYAN ARI ARDANA
13	0	RAHUL RIZKY

Gambar 7. Hasil Prediksi Kelas Siswa

**B. Pengujian Akurasi Metode**

Setelah sistem diimplementasikan dengan baik, langkah selanjutnya adalah melakukan pengujian sistem. Pengujian sistem dilakukan dengan metode black box testing dan white box testing. Metode black box testing digunakan untuk meyakinkan bahwa fitur-fitur atau fungsi-fungsi yang terdapat dalam sistem yang dibangun sudah berjalan baik. Sedangkan pengujian whitebox berupa pengujian terhadap algoritma algoritma yang digunakan. Setelah tahap pengujian selesai, selanjutnya dilakukan proses evaluasi sistem. Evaluasi sistem yang akan dilakukan berfokus pada evaluasi kinerja dari metode Naive Bayes dan K-Medoid dalam memprediksi jurusan dan kelas siswa. Evaluasi dilakukan dengan melakukan uji akurasi metode menggunakan confusion matrix pada metode naive bayes dan silhouette index pada metode k-medoid. Berikut hasil evaluasi kinerja metode menggunakan confusion matrix dan silhouette index.

Berdasarkan 30 data testing yang digunakan dalam mengevaluasi kinerja metode, didapatkan hasil confusion matrix pada tabel

*Tabel 6. Hasil Confusion Matrix MultiClass*

		PREDIKSI		
		IPS	IPA	BAHASA
AKTUAL	IPS	14	0	0
	IPA	2	4	1
	BAHASA	4	0	5

Dari hasil tabel 4.19 dapat dihitung nilai *sensitivity*, *precision*, *recall*, *specificity*, *f-measure* dan akurasi pada metode naive bayes. Berikut hasil perhitungan kinerja pada metode naive bayes.

$$Sensitivity\ IPA = \frac{TP}{TP+FN} \times 100\% = \frac{4}{4+2+1} \times 100\% = 57\%$$

$$Specificity\ IPA = \frac{TN}{TN+FP} \times 100\% = \frac{14+0+4+5}{14+0+4+5+0+0} \times 100\% = 100\%$$

$$Precision\ IPA = \frac{TP}{TP+FP} \times 100\% = \frac{4}{4+0+0} \times 100\% = 100\%$$

$$Recall\ IPA = \frac{TP}{TP+FN} \times 100\% = \frac{4}{4+2+1} \times 100\% = 57\%$$

$$Sensitivity\ IPS = \frac{TP}{TP+FN} \times 100\% = \frac{14}{14+0+0} \times 100\% = 100\%$$

$$Specificity\ IPS = \frac{TN}{TN+FP} \times 100 = \frac{4+1+0+5}{4+1+0+5+2+4} \times 100\% = 62\%$$

$$Precision\ IPS = \frac{TP}{TP+FP} \times 100\% = \frac{14}{14+2+4} \times 100\% = 70\%$$

$$Recall\ IPS = \frac{TP}{TP+FN} \times 100\% = \frac{14}{14+0+0} \times 100\% = 100\%$$

$$Sensitivity\ BAHASA = \frac{TP}{TP+FN} \times 100\% = \frac{5}{5+4+0} \times 100\% = 55\%$$

$$Specificity\ BAHASA = \frac{TN}{TN+FP} \times 100\% = \frac{14+0+2+4}{14+0+2+4+0+1} \times 100\% = 95\%$$

$$Precision\ BAHASA = \frac{TP}{TP+FP} \times 100\% = \frac{5}{5+0+1} \times 100\% = 83\%$$

$$Recall\ BAHASA = \frac{TP}{TP+FN} \times 100\% = \frac{5}{5+4+0} \times 100\% = 55\%$$

$$Sensitivity = \frac{Sensitivity\ IPA+Sensitivity\ IPS+Sensitivity\ BAHASA}{3} \times 100\% = \frac{57+100+55}{3} \times 100\% = 70,67\%$$

$$Specificity = \frac{Specificity\ IPA + Specificity\ IPS}{3}$$

$$= \frac{\text{Specificity BAHASA}}{3} \times 100\%$$

$$= \frac{100+62+95}{3} \times 100\% = 85,67\%$$

$$\text{Precision} = \frac{\text{Precision IPA} + \text{Precision IPS} + \text{Precision BAHASA}}{3} \times 100\%$$

$$= \frac{100+70+83}{3} \times 100\% = 84,333\%$$

$$\text{Recall} = \frac{\text{Recall IPA} + \text{Recall IPS} + \text{Recall BAHASA}}{3} \times 100\%$$

$$= \frac{57+100+55}{3} \times 100\% = 70,67\%$$

$$F\text{-Measure} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \times 100\%$$

$$= 2 \times \frac{84,333 \times 70,67}{84,333 + 70,67} \times 100\% = 76,89\%$$

$$\text{Akurasi} = \frac{TP}{TP + E_{12} + E_{13} + E_{21} + E_{23} + E_{31} + E_{32} + TP}$$

$$= \frac{14}{14+0+0+2+4+1+4+0+5} \times 100\% = 76\%$$

Berdasarkan hasil perhitungan kinerja metode naive bayes diatas, didapatkan nilai akurasi sebesar 76% yang menunjukkan bahwa metode naive bayes memiliki nilai akurasi yang cukup baik dalam memprediksi jurusan siswa. *Precision* yang dihasilkan sebesar 84,33%. Hal tersebut menunjukkan kategori data yang diklasifikasi telah sesuai dengan kategori yang sebenarnya. *Recall* yang dihasilkan sebesar 70,67%. Hal tersebut menunjukkan tingkat keberhasilan sistem dalam mengenali suatu kategori sudah baik. *F-measure* merupakan gambaran pengaruh relatif antara *precision* dan *recall* atau disebut *harmonic mean*. *F-measure* yang dihasilkan oleh sistem sebesar 79,8% sedangkan pada perhitungan manual sebesar 76,89%. *Sensitivity* digunakan untuk mengukur *True Positives* yang diidentifikasi dengan benar. *Sensitivity* yang dihasilkan sebesar 70,67%. *Specificity* digunakan untuk mengukur *True Negatives* yang diidentifikasi dengan benar. *Specificity* yang dihasilkan sebesar 85,67%.

Evaluasi kinerja k-medoid dihasilkan melalui perhitungan silhouette index. SI digunakan untuk memvalidasi sebuah data, cluster tunggal, atau bahkan keseluruhan cluster. Metode ini banyak digunakan untuk memvalidasi cluster yang menggabungkan nilai kohesi dan separasi. Berikut hasil perhitungan silhouette index k-medoid pada jurusan IPA.

Untuk mencari nilai SI, tentukan nilai a dan b terlebih dahulu.

Perhitungan nilai a untuk data yang berada dalam cluster 1 sebagai berikut:

$$a_1^1 = \frac{1}{3-1} (d(x_1^1, x_2^1) + d(x_1^1, x_3^1) + d(x_1^1, x_4^1) + d(x_1^1, x_5^1))$$

$$a_1^1 = \frac{1}{2} (25,76836 + 31,583) = 28,67568$$

Perhitungan nilai b untuk data yang berada dalam cluster 1 sebagai berikut:

$$b_1^1 = \min \left\{ \frac{1}{m_n} \sum_{r \neq i}^{m_j} d(x_i^j, x_r^n) \right\}$$

$$b_1^1 = \min \left\{ \frac{1}{5} (36,5606 + 29,79242 + 36,74222 + 33,89233 + 146,5462) \right\}$$

$$b_1^1 = 56,70675$$

Perhitungan nilai Silhouette Index untuk data yang berada dalam cluster 1 sebagai berikut:

$$SI_1^1 = \frac{b_1^1 - a_1^1}{\max \{b_1^1, a_1^1\}} = \frac{56,70675 - 28,67568}{\max \{56,70675; 28,67568\}} = 0,494316$$

Berikut adalah hasil nilai SI untuk setiap data dalam cluster 1

Tabel 7. Hasil nilai SI untuk setiap data dalam cluster 1

	jarak	1	3
1	1	-	22,34052
	3	22,34052	-
	a	22,34052	22,34052
2	2	35,49244	29,79242
	4	36,0536	32,84031
	RATA2	35,77302	31,31636
	b	35,77302	31,31636
	SI	0,375492	0,286618
	SI1	0,331055	

Nilai SI untuk keseluruhan cluster 1:

$$SI_j = \frac{1}{m_j} \sum_{i=1}^{m_j} SI_i^j$$

$$SI_1 = \frac{1}{2} (0,37549 + 0,286618)$$

$$= 0,331055$$

Berikut adalah nilai SI untuk setiap data dalam cluster 2

Tabel 8. Hasil nilai SI untuk setiap data dalam cluster 2

	jarak	2	4
2	2	-	23,03483
	4	23,03483	-
	a	23,03483	23,03483
1	1	35,49244	36,0536
	3	29,79242	32,84031
	RATA2	32,64243	34,44695
	b	32,64243	34,44695
	SI	0,294328	0,331295
	SI2	0,312812	

Nilai SI untuk keseluruhan cluster 2:

$$SI_j = \frac{1}{m_j} \sum_{i=1}^{m_j} SI_i^j$$

$$SI_2 = \frac{1}{2} (0,294328 + 0,331295)$$

$$= 0,312812$$

Nilai SI global:

$$SI = \frac{1}{k} \sum_{j=1}^k SI_j$$

$$SI = \frac{1}{2} (0,331055 + 0,312812) = 0,32193$$

Berdasarkan pengujian diatas diperoleh hasil nilai *silhouette indeks* IPA dengan jumlah cluster adalah 0,32193 yang berarti tergolong dalam *weak structure*. Nilai *silhouette indeks*

dengan jumlah 0,23574 dan nilai silhouette indeks IPS sebesar 0,16301 yang juga tergolong dalam weak structure.

#### 4. Conclusion

Pengujian akurasi peneliti melakukan dua uji akurasi yaitu pada algoritma Naive Bayes dan K-Medoid. Hasil akurasi algoritma Naive Bayes sebesar 76%. Sedangkan algoritma K-Medoid, hasil akurasi silhouette index yaitu Bahasa= 0,23575, IPA = 0,32193 dan IPS = 0,16301. Pengembangan Sistem Prediksi Penjurusan Kelas Siswa Menggunakan Kombinasi Algoritma Naive Bayes dan K-Medoid dikembangkan dengan menggunakan bahasa pemrograman PHP dan javascript dengan Framework CodeIgniter. Metode Naive Bayes digunakan untuk memprediksi jurusan siswa berdasarkan komponen atribut yang telah ditentukan. Sedangkan algoritma K-Medoid digunakan untuk pembentukan kelas siswa yang telah terprediksi jurusannya menggunakan metode Naive Bayes sebelumnya. Terdapat beberapa hal yang dapat dijadikan bahan pertimbangan untuk pengembangan berikutnya. Diharapkan pengembangan selanjutnya dapat mengkombinasikan dengan metode lain dari salah satu metode klusterisasi pada proses penentuan kelas siswa agar mendapatkan tingkat akurasi yang lebih tinggi. Diharapkan pengembang selanjutnya dapat mengembangkan sistem dengan menambah kriteria yang digunakan seperti tes akademik siswa yang dilakukan di sekolah. Penelitian selanjutnya dapat dilakukan dengan pengujian penelitian eksperimen atau user experience (UX) terhadap Sistem Prediksi Penjurusan Kelas Siswa Menggunakan Kombinasi Algoritma Naive Bayes dan K-Medoid (Studi Kasus: SMA Laboratorium Undiksha).

#### References

- [1] B. Novianti, "IMPLEMENTASI DATA MINING DENGAN ALGORITMA C4 . 5 UNTUK PENJURUSAN SISWA ( STUDI KASUS : SMA NEGERI 1 PONTIANAK )," vol. 04, no. 3, 2016.
- [2] V. N. Widowati, "Studi Kasus Tentang Proses Penjurusan beberapa SMA di Yogyakarta," p. 2, 2015.
- [3] G. A. Pradnyana and A. A. J. Permana, "Sistem Pembagian Kelas Kuliah Mahasiswa Dengan Metode K-Means Dan K-Nearest Neighbors Untuk Meningkatkan Kualitas Pembelajaran," *JUTI J. Ilm. Teknol. Inf.*, vol. 16, no. 1, p. 59, 2018.
- [4] J. Han, M. Kamber, and J. Pei, "Data Mining – Concepts & Techniques," 2011, pp. 1–744.
- [5] S. Defiyanti and M. Jajuli, "Optimalisasi K - Medoid Dalam Pengklasteran Mahasiswa Pelamar Beasiswa Dengan Cubic Clustering Criterion," vol. 03, no. 01, pp. 211–218, 2017.
- [6] D. A. Silitonga, A. P. Windarto, and D. Hartama, "Penerapan Metode K-Medoid pada Pengelompokan Rumah Tangga Dalam Perlakuan Memilah Sampah Menurut Provinsi," pp. 313–318, 2019.
- [7] A. Y. Kencana, S. Astuti, J. Teknik, I. Komputer, and U. D. Nuswantoro, "METODE KLASIFIKASI DENGAN ALGORITMA NAÏVE BAYES," vol. 15, no. 3, pp. 195–200, 2016.
- [8] A. H. Hailitik, B. S. Djahi, and Y. Y. Nabuasa, "KLASIFIKASI JURUSAN MENGGUNAKAN METODE NAÏVE BAYES PADA," vol. 5, no. 2, pp. 21–27, 2017.
- [9] I. M. A. Wirawan, *Metode Penalaran Dalam Kecerdasan Buatan*, 1st ed. Depok: Rajawali Pers, 2017.
- [10] G. A. Pradnyana, I. K. A. Suryantara, and I. G. M. Darmawiguna, "Impression Classification of Endek (Balinese Fabric) Image Using K-Nearest Neighbors Method", *KINETIK*, vol. 3, no. 3, pp. 213-220, Apr. 2018.