

PENERAPAN ALGORITMA K-MEANS UNTUK PENGELOMPOKAN PENILAIAN AKHIR SEMESTER DI SDN KADOKAN 01 SUKOHARJO

Karlina Kusuma Ningrum¹, Joni Maulindar², Anisatul Farida³

^{1,2,3}Teknik Informatika, Fakultas Ilmu Komputer, Universitas Duta Bangsa Surakarta

Email : 190103110@fikom.udb.ac.id¹, joni_maulindar@udb.ac.id², 3anisatul_farida@udb.ac.id³

ABSTRACT

Final Semester Assessment (PAS) is an annual assessment of educational standards. Final Semester Assessment is an achievement of competency in all subjects at school. Based on data from the results of the Final Semester Assessment (PAS) that researchers conducted at SDN Kadokan 01 Sukoharjo, where the data presented did't contain a specific category so that the informations to be conveyed to parents was unclear. Based on these problems, a study was conducted using the Final Semester Assessment (PAS) scores at SDN Kadokan 01 Sukoharjo using data mining techniques, namely the k-means clustering algorithm by grouping data into several clusters based on data similarity. Based on the result of the study, 3 clusters were indentified, namely cluster 0 with a low level of grouping in the Final Semestwr Assessment scores, there were 43 students, cluster 1 with a moderate level of grouping, there were 62 students, and cluster 2 with a high level of grouping, there were 63 students.

Keywords: Algorithm , Data Mining, Final Semester Assessment, Clustering

ABSTRAK

Penilaian Akhir Semester (PAS) merupakan penilaian tahunan terhadap standar pendidikan. Penilaian Akhir Semester merupakan kinerja kompetensi semua mata pelajaran di Sekolah. Berdasarkan data hasil survei Penilaian Akhir Semester (PAS) yang dilakukan peneliti di SDN Kadokan 01 Sukoharjo, dimanainformasi yang diberikan tidak mencantumkan kelas tertentu, sehingga informasi yang diberikan kepada orang tua siswa kurang jelas. Berdasarkan permasalahan tersebut maka dilakukan sebuah penelitian dengan menggunakan nilai Penilaian Akhir Semester (PAS) di SDN Kadokan 01 Sukoharjo dengan menggunakan teknik data mining yaitu algoritma *k-means clustering*, mengelompokkan data menjadi beberapa *cluster* berdasarkan kesamaan data. Berdasarkan hasil penelitian teridentifikasi 3 cluster yaitu *cluster 0* yang memiliki tingkat cluster rendah pada nilai penilaian akhir semester sebanyak 43 siswa, *cluster 1* dengan tingkat *cluster* sedang sebanyak 62 siswa, dan *cluster 2* dengan tingkat *cluster* tinggi sebanyak 63 siswa.

Kata Kunci: Algoritma K-Means, Data Mining, Penilaian Akhir Semester, Pengelompokan

Riwayat Artikel :

Tanggal diterima : 20-05-2023

Tanggal revisi : 21-05-2023

Tanggal terbit : 22-05-2023

DOI :

<https://doi.org/10.31949/infotech.v9i1.xxx>

INFOTECH journal by Informatika UNMA is licensed under CC BY-SA 4.0

Copyright © 20xx By Author



1. PENDAHULUAN

1.1. Latar Belakang

Pada akhir semester, sekolah atau madrasah melaksanakan kegiatan Penilaian Akhir Semester (PAS) untuk mengevaluasi hasil belajar siswa. Berdasarkan uraian diatas, diketahui bahwa peneliti memiliki landasan yang kuat. Dengan pemikiran ini, seseorang dapat mengkaji faktor-faktor yang menjadi ukuran peningkatan kualitas dan pemerataan pendidikan. Teknik analisis dapat digunakan untuk mendapatkan informasi dari data, untuk itu diperlukan data, dan untuk itu diperlukan pengolahan data, digunakannya teknik *data mining*. Sesuai Tacbir (2019), eksplorasi data mencakup pemanfaatan prosedur terukur, matematika, kesadaran buatan manusia, dan AI untuk memperoleh data dan informasi yang signifikan dari berbagai sumber informasi yang sangat banyak. Terdapat beragam teknik eksplorasi data, salah satunya adalah pengelompokan data (*clustering*). Salah satu metode pengelompokan data atau objek dalam data mining adalah “algoritma *clustering K-Means*”. Perhitungan ini digunakan untuk mengelompokkan informasi menjadi kelompok-kelompok berdasarkan kemiripan atau persamaan di antara mereka. Dengan menggunakan algoritma *K-Means*, setiap kelompok akan berisi data yang memiliki karakteristik serupa atau mirip. Algoritma *K-Means* telah dibuktikan berhasil oleh peneliti dalam pengelompokan data. Penelitian ini menggunakan data penilaian akhir semester SDN Kadokan 01 Sukoharjo untuk tiga mapel, yakni bahasa Indonesia, matematika, serta bahasa Inggris. Dalam penelitian, dilakukan pengelompokan data dengan Algoritma *K-Means*, yang menghasilkan tiga jenis cluster. Kluster 0 yakni kategori tingkat rendah, kluster 1 merupakan tingkat sedang, serta kluster 2 merupakan kategori tingkat tinggi.

1.2. Tinjauan Pustaka

1. Penilaian Akhir Semester

PAS adalah sebuah metode evaluasi pembelajaran yang menggambarkan sejauh mana siswa memahami materi pembelajaran dalam satu semester. Tujuan dari PAS adalah untuk mengukur pencapaian kompetensi secara menyeluruh, meliputi pengetahuan, sikap, dan keterampilan yang telah diperoleh oleh siswa. Hasil penilaian akhir semester dapat digunakan oleh sekolah untuk menentukan persentase kelulusan siswa pada satuan pelajaran berdasarkan standar kelulusan. (Rintan, 2022)

2. Data Mining

Data mining adalah proses mengungkap dan mengidentifikasi informasi berharga dan relevan dari berbagai sumber data yang terhubung menggunakan teknik statistik, matematika, AI, dan pembelajaran mesin. (Novi, 2021)

3. Clustering

Clustering melibatkan pemisahan sekelompok objek data menjadi kelompok yang disebut cluster. Setiap objek dalam suatu cluster memiliki sifat-sifat

yang serupa dan berbeda dari cluster lainnya. (Karin, 2022)

4. Algoritma K-Means

“Algoritma *K-Means*” yakni metode untuk menganalisis kelompok, di mana N pengamatan dibagi menjadi kelompok-kelompok (*cluster*) berdasarkan kelompok rata-rata (*mean*) terdekat. Algoritma *K-Means* ini dapat diimplementasikan dengan mudah, memiliki kinerja yang relatif cepat, fleksibel dalam penyesuaian, dan umum digunakan. Algoritma *k-means* mengklasifikasikan data berdasarkan apakah mereka dekat dengan pusat massa. Data dikelompokkan dengan memaksimalkan data yang sama dalam sebuah cluster dan meminimalkan data yang sama di seluruh cluster. (Shefia, 2022)

5. Robust Scaler

Robust Scaler adalah penskala fitur menggunakan statistic *outlier* dan digunakan untuk menghapus median dan menskalakan data berdasarkan rentang interkuartil. Transformasi data dilakukan dengan menghitung statistik sesuai dari sampel dalam set pelatihan untuk melakukan pemusatan dan penskalaan secara independen dari setiap fitur. Setelah itu, rentang media dan interkuartil dicatat dan digunakan dalam data berikutnya.

6. Metode Elbow

Metode *elbow* membandingkan sudut yang terbentuk pada titik-titik tertentu berdasarkan jumlah klaster yang berbeda untuk menentukan jumlah klaster yang optimal. Metode *elbow* membandingkan sudut yang terbentuk pada titik-titik tertentu berdasarkan jumlah klaster yang berbeda untuk menentukan jumlah klaster yang optimal. Jika nilai jumlah klaster pertama dan jumlah klaster kedua membentuk sebuah siku pada diagram, atau jika penurunan jumlah klaster setelah itu tidak signifikan, maka jumlah klaster tersebut dapat dianggap tepat. Sebagai alternatif, untuk membandingkan jumlah klaster yang tepat, dilakukan perhitungan *Sum of Square Error (SSE)* untuk setiap klaster dengan mengkuadratkan selisih antara setiap titik dengan pusat klaster. Semakin banyak *cluster K*, semakin rendah nilai *SSE*-nya. Persamaan menghitung *SSE* adalah sebagai berikut:

$$SSE = \sum_{K=1}^K \sum_{Xi} |x_i - c_k|^2$$

Keterangan:

K = cluster ke- c

x_i = jarak data obyek ke- i

c_k = pusat cluster ke- i

(Dewa Ayu, 2019)

7. Silhouette Coefficient

Koefisien *Silhouette* digunakan untuk mengukur keunggulan dan keefektifan klasterisasi, yaitu sejauh mana objek ditempatkan dengan tepat dalam kluster yang sesuai. Metode ini menggabungkan proses ekstraksi dan pengkodean data. Koefisien *silhouette* untuk setiap titik data ke- i harus dihitung

sebelum koefisien siluet dapat dihitung. Sesuai dengan persamaan yang diberikan di bawah ini, nilai maksimum indeks Silhouette Global dalam rentang cluster 2 hingga n-1 digunakan untuk menghitung koefisien silhouette.

$$SC = maks_k SI(k)$$

Keterangan :

$SC = Silhouette\ Coefficient$

$SI = Silhouette\ Index\ Global$

$k = jumlah\ cluster$

(Dewa Ayu, 2019)

1.3 Metodologi Penelitian

1. Metode Pengumpulan Data

Dataset yang digunakan merupakan kumpulan data yang telah terkumpul berisi kumpulan hasil nilai Penilaian Akhir Semester yang dikelompokkan ke dalam bentuk format file *Microsoft Excel* (xls).

2. Metode Observasi

Pada penelitian ini observasi merupakan kumpulan nilai Penilaian Akhir Semester (PAS) siswa/siswi di SDN Kadokan 01 Sukoharjo.

3. Metode Analisis

Metode analisis diterapkan untuk mengidentifikasi masalah menggunakan metode *K-means*.

2. PEMBAHASAN

2.1 Implementasi Clustering K-Means

a. Memperoleh Data

Tabel 1 . Sample dataset

No	Nama Siswa	B Indonesia	Matematika	B Inggris	Jumlah	Rata-Rata
1	Abid Aqila Rajendra	77	77	86	240	80
2	Aceline Nindyana Putri	90	88	94	272	90,6667
3	Adyaraka Dharma Handaru	85	87	94	266	88,6667
4	Alfatih Rohman Prasetyono	76	76	83	235	78,3333
5	Anindita Putri Ardani	78	79	93	250	83,3333

Pada tabel 1 adalah contoh dataset yang menggunakan library Pandas untuk mengubah dataset yang berformat *comma-separated values* (csv) menjadi data frame.

Tabel 2. Struktur data

Column	Non-Null Count	Type
No	168 non-null	int64
Nama Siswa	168 non-null	object
B.Indonesia	168 non-null	int64
Matematika	168 non-null	int64
B.Ingggris	168 non-null	int64
Jumlah	168 non-null	int64
Rata-Rata	168 non-null	int64
Dtypes:float64 (1), int64 (5), object (1)		

Tabel 2 merupakan struktur konten data dengan menggunakan fungsi `dataset.info()`, sehingga semua informasi disertakan.

b. Data Cleaning

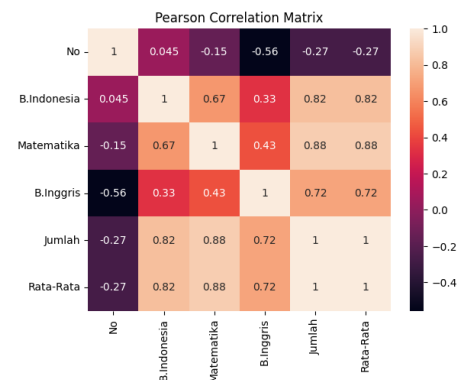
Menggunakan nilai rerata dan deviasi standar untuk mencari ambang batas threshold.

Tabel 3. Perhitungan threshold

Atribut	Nilai Plus (+)	Nilai Minus(-)
Threshold B Indonesia	88,20193358	1,48802256
Threshold Matematika	88,22756794	79,71290825
Threshold B Inggris	90,61897494	82,84531077

Table 3 menunjukkan perhitungan threshold untuk ketiga atribut tersebut.

c. Feature Selection



Gambar 1. Heatmap Person Correlation Matrix

Dari gambar 1 menunjukkan bahwa Bahasa Indonesia dan Bahasa Inggris cukup signifikan karena mendekati nilai 1. Bahasa Indonesia dan Matematika signifikan karena nilainya 0,67. Matematika dan Bahasa Inggris cukup signifikan.

d. Data Reduction

Data reduction digunakan untuk menghilangkan atribut yang tidak mempengaruhi proses pengelompokan nilai yaitu 'No' dan 'Nama Siswa'.

e. Data Transformation

Tujuan dari adanya transformasi data adalah untuk menyamakan distribusi data agar lebih mudah dalam mengklasifikasikan data. Langkah awal adalah memeriksa histogram atau boxplot dari setiap atribut guna menentukan transformasi yang tepat. Hal ini dapat membantu dalam

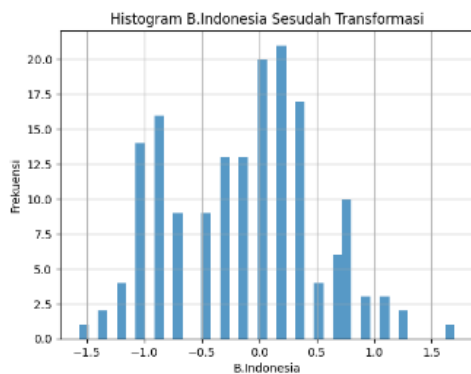
menampilkan data yang hilang atau outlier.

1. Bahasa Indonesia



Gambar 2. Bahasa Indonesia Sebelum Transformasi

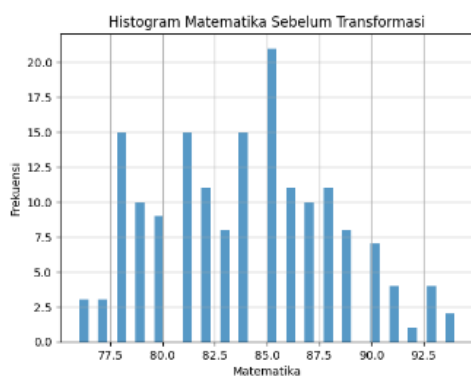
Gambar 2 menunjukkan proses konversi Bahasa Indonesia menggunakan Logaritma Transformasi.



Gambar 3. Bahasa Indonesia Sesudah Transformasi

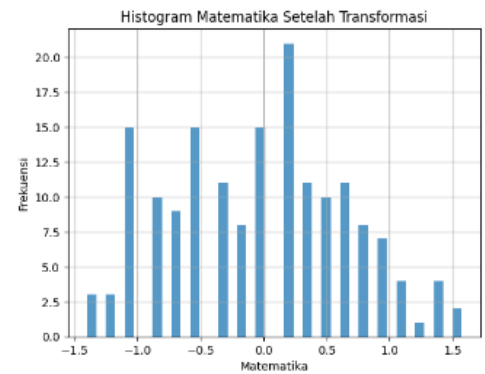
Gambar 3 menunjukkan histogram bahasa indonesia setelah transformasi dengan menggunakan Logaritma Transformasi dan Robust Scaler.

2. Matematika



Gambar 4. Matematika Sebelum Transformasi

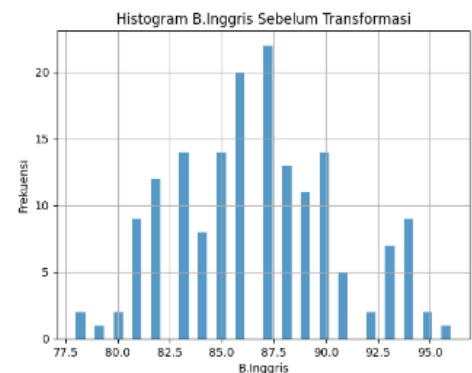
Gambar 4 menunjukkan proses transformasi matematika menggunakan Logaritma Transformasi.



Gambar 5. Matematika Setelah Transformasi

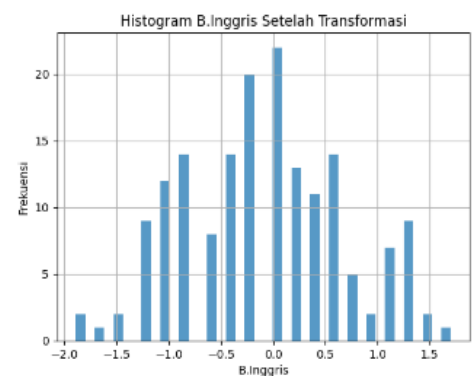
Gambar 5 menunjukkan histogram matematika setelah dilakukan ditransformasi dengan menggunakan Logaritma Transformasi dan Robust Scaler.

3. Bahasa Inggris



Gambar 6. Bahasa Inggris Sebelum Transformasi

Gambar 6 menunjukkan proses konversi bahasa inggris menggunakan logaritma transformasi.



Gambar 7. Bahasa Inggris Sesudah Transformasi

Gambar 7 menunjukkan histogram setelah dilakukan ditransformasi menggunakan Logaritma Transformasi dan Robust Scaler.

f. Clustering K-Means

Saat menggabungkan algoritma k-means, peneliti memilih 3 kluster.

Tabel 4. Titik pusat awal

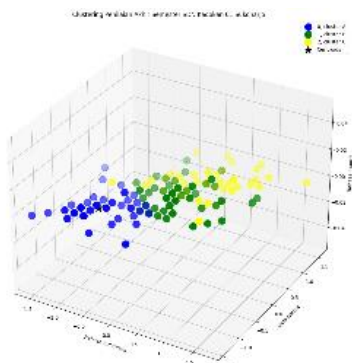
Titik Pusat (<i>Centroid</i>)		
-0,77019976	-0,747064	-0,54606761
0,15512561	0,15199978	-0,27676683
0,3197022	0,666857185	0,82186091

Pada tabel 4 merupakan titik fokus pada penelitian ini. Setelah menentukan titik pusat secara acak, dihitung jarak *Euclidean Distance* dan hasil akhir pada iterasi 15.

Tabel 5. Clusterisasi tiap nomor

	<i>Cluster 0</i>	<i>Culster 1</i>	<i>Cluster 2</i>
No	2,3,6,7,8,1 0,11,16,17 ,18,19,20, 22,23,24,2 5,31,32,35 ,41,45,48, 49,54,56,5 9,60,64,66 ,68,69,72, 73,74,77,7 8,82,101,1 20,138	1,4,5,12,13, 14,15,21,26, 28,30,37,55, 75,876,80,8 4,85,86,89,9 0,91,92,93,9 4,96,97,98,9 9,100,103,1 04,105,106, 107,109,110 ,111,112,11 3,114,115,1 16,117,118, 119,121,127 ,129,130,13 2,133,139,1 40,141,142, 146,147,151 ,152,154,16 2	27,29,33,3 4,36,38,39 ,40,42,43, 44,46,47,5 0,51,52,53 ,57,58,61, 62,63,65,6 7,70,71,79 ,81,83,95, 102,108,1 22,123,12 4,125,126, 128,131,1 34,135,13 7,143,144, 145,148,1 49,150,15 3,155,157, 158,159,1 60,161,16 3,164,165, 166,167,1 68

Pada tabel 5 menunjukkan hasil clustering untuk setiap nomor. Setelah melakukan pemrosesan data, hasil dari clustering K-Means akan dapat diketahui dan disajikan dalam bentuk subplot 3D untuk divisualisasikan.



Gambar 8. Visualisasi Clustering K-Means

Visualisasi Gambar 8 menampilkan “hasil clustering k-means” dari data penilaian akhir semester di SDN Kadokan 01 Sukoharjo.

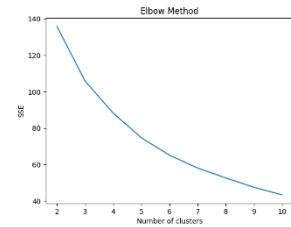
2.2 Pengujian

Dalam penelitian tersebut, para peneliti

melakukan eksperimen dengan menerapkan tiga model yang mencakup pendekatan Elbow dan evaluasi koefisien Silhouette guna mengidentifikasi jumlah cluster yang memberikan hasil optimal.

1. Elbow Method

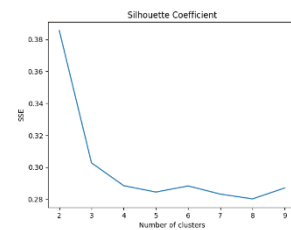
Elbow method digunakan untuk menentukan jumlah *cluster* terbaik dalam suatu kumpulan data.



Gambar 9. Hasil Uji Elbow Method

Pada pengujian *cluster* terbaik dengan menggunakan *Elbow Method* seperti terlihat pada gambar 9, cluster optimal untuk kumpulan data penilaian akhir semester adalah 3.

2. Silhouette Coefficient



Gambar 10. Hasil Uji Silhouette Coefficient

Pada pengujian cluster terbaik pada gambar 10 menggunakan *Silhouette* dengan hasil yakni jumlah *cluster* yang optimal yangmana kumpulan data penilaian akhir semester adalah 3.

Tabel 6. Evaluasi hasil pengujian

Jumlah Cluster	Elbow Method	Silhouette Coefficient
2	135,76530479 927092	0,38553423 538358367
3	105,77963182 73822	0,30266371 852353413
4	88,162942239 02338	0,28838230 915716556
5	74,638108812 47481	0,28439443 00438978
6	65,084361228 50367	0,28822459 740831446
7	58,019297444 68485	0,28311686 9006686
8	52,584728499 39092	0,28011776 13440898

9	47,463502685 665745	0,28696619 470382534
10	43,287224732 7019	0,29026290 20222415
11	41,178728503 5277	0,28324512 616290204
12	37,106704694 03689	0,30360615 741782904
13	34,955891253 871116	0,28964007 103191947
14	33,254643577 682245	0,28100295 75415456
15	30,378533355 552946	0,29576769 993395924
16	29,129366671 692935	0,28231617 84380226

Pada tabel 6 menunjukkan hasil pengujian masing-masing metode dengan perulangan 2 sampai 16. Ketika menggunakan Metode Elbow dan Koefisien Silhouette, diperoleh hasil cluster optimal sebanyak 3.

3. ALGORITMA ATAU PROGRAM

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

dataset = pd.read_csv('dataset.csv')
dataset.head()
dataset.info()
plt.figure(figsize=(20, 10))
plt.plot(dataset['Nama Siswa'], dataset['B.Indonesia'], label='B.Indonesia')
plt.plot(dataset['Nama Siswa'], dataset['Matematika'], label='Matematika')
plt.plot(dataset['Nama Siswa'], dataset['B.Inggris'], label='B.Inggris')
plt.xticks(rotation=90)
plt.legend()
plt.show()
sns.catplot(x='Nama Siswa', y='B.Indonesia', kind='bar', data=dataset, height=5, aspect=3)
plt.xticks(rotation=90)
plt.show()
sns.catplot(x='Nama Siswa', y='Matematika', kind='bar', data=dataset, height=5, aspect=3)
plt.xticks(rotation=90)
plt.show()
sns.catplot(x='Nama Siswa', y='B.Inggris', kind='bar', data=dataset, height=5, aspect=3)
plt.xticks(rotation=90)
plt.show()
```

```
dataset.isnull().sum()
dataset.describe()
sns.heatmap(dataset.corr(), annot = True,
xticklabels= True, yticklabels= True)
plt.title('Pearson Correlation Matrix')
plt.show(True)
dataset = dataset.drop(['No', 'Nama Siswa'],
axis = 1)
print ('The shape of the DataFrame is :',
dataset.shape, '\n')
dataset.head()
dataset[['B.Indonesia', 'Matematika', 'B.Inggris']].boxplot(figsize=(11,4))
plt.show()
datacluster = dataset.iloc[:, 0:3]
datacluster.head()
dataArray = np.array(datacluster)
plt.hist(dataArray[:,0], label='True Position', bins=40, alpha=0.75)
plt.xlabel('B.Indonesia')
plt.ylabel('Frekuensi')
plt.title('Histogram B.Indonesia Sebelum Transformasi')
plt.grid()
plt.show()
from sklearn import preprocessing
from sklearn.preprocessing import RobustScaler
dataArray_log = np.log(dataArray)
data_scale_bindonesia = RobustScaler().fit_transform(dataArray_log)
plt.hist(data_scale_bindonesia[:,0], label='True Position', bins=40, alpha=0.75)
plt.xlabel('B.Indonesia')
plt.ylabel('Frekuensi')
plt.title('Histogram B.Indonesia Sesudah Transformasi')
plt.grid()
plt.show()
x = pd.DataFrame([dataset['B.Indonesia'], data_scale_bindonesia[:,1]])
x = x.transpose()
x.columns = ['Normal plot', 'After Log Transformation and RobustScaler']
plt.figure(figsize=(10,5))
plt.title('B.Indonesia Log Transformation and RobustScaler')
x.boxplot()
plt.hist(dataArray[:,1], label='True Position', bins=40, alpha=0.75)
plt.xlabel('Matematika')
plt.ylabel('Frekuensi')
plt.title('Histogram Matematika Sebelum Transformasi')
plt.grid()
plt.show()
```

```

from sklearn import preprocessing
from sklearn.preprocessing import RobustScaler

dataArray_log = np.log(dataArray)
data_scale_matematika = RobustScaler().fit_transform(dataArray_log)

plt.hist(data_scale_matematika[:,1], label='True Position', bins=40, alpha=0.75)

plt.xlabel('Matematika')
plt.ylabel('Frekuensi')
plt.title('Histogram Matematika Setelah Transformasi')
plt.grid()
plt.show()

x = pd.DataFrame([dataset['Matematika'], data_scale_matematika[:,1]])
x = x.transpose()
x.columns = ['Normal plot', 'After Log Transformation and RobustScaler']
plt.figure(figsize=(10,5))
plt.title('Matematika Log Transformation and RobustScaler')
x.boxplot()
plt.hist(dataArray[:,2], label='True Position', bins=40, alpha=0.75)

plt.xlabel('B.Ingggris')
plt.ylabel('Frekuensi')
plt.title('Histogram B.Ingggris Sebelum Transformasi')
plt.grid()
plt.show()

from sklearn import preprocessing
from sklearn.preprocessing import RobustScaler

dataArray_log = np.log(dataArray)
data_scale_binggris = RobustScaler().fit_transform(dataArray_log)

plt.hist(data_scale_binggris[:,2], label='True Position', bins=40, alpha=0.75)

plt.xlabel('B.Ingggris')
plt.ylabel('Frekuensi')
plt.title('Histogram B.Ingggris Setelah Transformasi')
plt.grid()
plt.show()

x = pd.DataFrame([dataset['B.Ingggris'], data_scale_binggris[:,2]])
x = x.transpose()
x.columns = ['Normal plot', 'After Log Transformation and RobustScaler']
plt.figure(figsize=(10,5))
plt.title('B.Ingggris Log Transformation and RobustScaler')
x.boxplot()

datascale_df = pd.DataFrame([data_scale_bindonesia[:,0], data_scale_matematika[:,1], data_scale_binggris[:,2]])

datascale = np.array(datascale_df.transpose())

from sklearn.cluster import KMeans
dataKmeans = KMeans(n_clusters = 3, init = 'k-means++', n_init=1, random_state=42)
dataKmeans.fit(datascale)
label = dataKmeans.predict(datascale)
print(dataKmeans.cluster_centers_)
print(dataKmeans.labels_)
dataset['cluster'] = dataKmeans.labels_
dataset.head()

centers = dataKmeans.cluster_centers_
fig = plt.figure(figsize=(15, 13))
ax = fig.add_subplot(projection='3d')
ax.scatter(datascale[label == 0,0], datascale[label == 0,1], s = 200, c = 'blue', label = '0, cluster A')
ax.scatter(datascale[label == 1,0], datascale[label == 1,1], s = 200, c = 'green', label = '1, cluster B')
ax.scatter(datascale[label == 2,0], datascale[label == 2,1], s = 200, c = 'yellow', label = '2, cluster C')
plt.scatter(centers[:,0], centers[:,1], marker = '*', c='black', s= 250, label = 'Centroids')
plt.legend()
ax.set_xlabel('Bahasa Indonesia')
ax.set_ylabel('Matematika')
ax.set_zlabel('Bahasa Inggris')
ax.set_title("Clustering Penilaian Akhir Semester SDN Kadokan 01 Sukoharjo")
plt.show()

for i in range(2,17):
kmeans = KMeans(n_clusters = i, init = 'k-means++', max_iter = 300, n_init =10, random_state= 42)
kmeans.fit(datascale)
sse.append(kmeans.inertia_)
print(i, sse)
plt.plot(range(2, 17), sse)
plt.title('Elbow Method')
plt.xlabel('Number of clusters')
plt.ylabel('SSE')
plt.show()

from sklearn.cluster import KMeans
sse = []
for i in range(2,4):
kmeans = KMeans(n_clusters = i, init = 'k-means++', max_iter = 300, n_init =10, random_state= 42)
kmeans.fit(datascale)
sse.append(kmeans.inertia_)
print(i, sse)

```

```

plt.plot(range(2,4), sse)
plt.title('Elbow Method')
plt.xlabel('Number of clusters')
plt.ylabel('SSE')
plt.show()
from sklearn.metrics import silhouette_score
si = []
for i in range(2,17):
    kmeans = KMeans(n_clusters = i, init = 'k-
means++', max_iter = 300, n_init =10,
random_state= 42)
    kmeans = kmeans.fit(datascale)
    labels = kmeans.labels_
    si.append(silhouette_score(datascale,
labels, metric = 'euclidean'))
print(si)
silhouette_score(datascale, kmeans.labels_)
plt.plot(range(2,17), si)
plt.title('Silhouette Coefficient')
plt.xlabel('Number of clusters')
plt.ylabel('SSE')
plt.show()
si = []
for i in range(2,4):
    kmeans = KMeans(n_clusters = i, init = 'k-
means++', max_iter = 300, n_init =10,
random_state= 42)
    kmeans = kmeans.fit(datascale)
    labels = kmeans.labels_
    si.append(silhouette_score(datascale,
labels, metric = 'euclidean'))
print(si)
plt.plot(range(2,4), si)
plt.title('Silhouette Coefficient')
plt.xlabel('Number of clusters')
plt.ylabel('SSE')
plt.show()

```

4. KESIMPULAN

Dari hasil penelitian dalam “penerapan algoritma *k-means*” melalui beberapa langkah, diantaranya data *cleaning*, data transformasi, *data mining* (Algoritma *K-Means*), dan kemudian metode evaluasi. Ada 3 *cluster* yaitu C0, C1, dan C2. Nilai C0 rendah, C1 sedang, dan C2 tinggi. Data yang digunakan adalah 168 siswa yang tergabung dalam pengelompokan nilai penilaian akhir semester kategori rendah ada 43 siswa, kategori sedang ada 62 siswa, dan kategori tinggi ada 63 siswa.

PUSTAKA

“sklearn.preprocessing.RobustScaler.” <https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.RobustScaler.html>

A. Aditya, I. Jovian, and B. N. Sari, “Implementasi K-Means Clustering Ujian Nasional Sekolah Menengah Pertama di Indonesia Tahun 2018 / 2019,” vol. 4, pp. 51–58, 2020, doi: 10.30865/mib.v4i1.1784.

- A. Cluster, D. Transaksi, P. Minimarket, and S. Pandemi, “Analisa Cluster Data Transaksi Penjualan Minimarket Selama Pandemi,” vol. 3, no. 28, pp. 153–160, 2022.
- A. Sulistiyawati and E. Supriyanto, “Implementasi Algoritma K-means Clustering dalam Penentuan Siswa Kelas Unggulan,” *J. Tekno Kompak*, vol. 15, no. 2, p. 25, 2021, doi: 10.33365/jtk.v15i2.1162.
- D. A. I. C. Dewi and D. A. K. Pramita, “Analisis Perbandingan Metode Elbow dan Silhouette pada Algoritma Clustering K-Medoids dalam Pengelompokan Produksi Kerajinan Bali,” *Matrix J. Manaj. Teknol. dan Inform.*, vol. 9, no. 3, pp. 102–109, 2019, doi: 10.31940/matrix.v9i3.1662.
- D. N. P. Sari and Y. L. Sukestiyarno, “Analisis Cluster dengan Metode K-Means pada Persebaran Kasus Covid-19 Berdasarkan Provinsi di Indonesia,” *Prism. Pros. Semin. Nas. Mat.*, vol. 4, pp. 602–610, 2021, [Online]. Available: <https://journal.unnes.ac.id/sju/index.php/prisma/>
- I. N. M. Adiputra, “CLUSTERING PENYAKIT DBD PADA RUMAH SAKIT DHARMA KERTI MENGGUNAKAN ALGORITMA K-MEANS,” vol. 2, no. 2, pp. 99–105, 2021.
- I. Rintan, Ashadi, and S. Endang, “ANALISIS BUTIR SOAL PENILAIAN AKHIR SEMESTER (PAS) MATA PELAJARAN KIMIA KELAS XI TAHUN AJARAN 2019/2020 SMA NEGERI 1 SRAGEN,” vol. 11, no. 1, pp. 33–37, 2022, [Online]. Available: <http://repository.upi.edu/id/eprint/70527>
- J. Hutagalung, Y. H. Syahputra, Z. P. Tanjung, K. Medan, and S. Informasi, “Pemetaan Siswa Kelas Unggulan Menggunakan Algoritma K-Means Clustering,” vol. 9, no. 1, pp. 606–620, 2022.
- K. Annisa, B. S. Ginting, and M. A. Syari, “Penerapan Data Mining Pengelompokan Data Pengguna Air Bersih Berdasarkan Keluhannya Menggunakan Metode Clustering Pada PDAM Langkat,” *J. Ilmu Komput. dan Inform. Vol.*, vol. 6, no. 1, pp. 113–129, 2022.
- K. Novi, “Data Mining Pengelompokan Pasien Rawat Inap Peserta BPJS Menggunakan Metode Clustering (Studi Kasus : RSUD.Bangkalan),” *J. Inf. Technol.*, vol. 1, no. 2, pp. 47–53, 2021, doi: 10.32938/jitu.v1i2.1470.
- K. Vi, D. I. Sdn, and A. Tahun, “PENILAIAN AKHIR SEMESTER MATA PELAJARAN MATEMATIKA,” vol. 12, no. 2, pp. 146–171, 2020.
- Nasution, A. P. Windarto, and M. Fauzan, “Penerapan Algoritma K-Means Dalam Pengelompokan Data Penduduk Miskin Menurut Provinsi,” vol. 2, no. 2, pp. 76–83, 2020.
- N. Dwitri, J. A. Tampubolon, S. Prayoga, and P. P. P. A. N. W. F. I. R. H. Zer, “PENERAPAN ALGORITMA K-MEANS DALAM MENENTUKAN TINGKAT PENYEBARAN

- PANDEMI COVID-19 DI INDONESIA,” vol. 4, no. 1, pp. 128–132, 2020.
- S. Handoko, E. Tri, and E. Handayani, “IMPLEMENTASI DATA MINING UNTUK MENENTUKAN TINGKAT PENJUALAN PAKET DATA TELKOMSEL MENGGUNAKAN METODE K-MEANS CLUSTERING,” vol. 25, no. 1.
- S. N. Br Sembiring, H. Winata, and S. Kusnasari, “Pengelompokan Prestasi Siswa Menggunakan Algoritma K-Means,” *J. Sist. Inf. Triguna Dharma (JURSI TGD)*, vol. 1, no. 1, pp. 31–40, 2022, doi: 10.53513/jursi.v1i1.4784.
- Z. I. Alfianti, U. Bina, S. Informatika, K. Kabupaten, J. Barat, and D. Mining, “ALGORITMA K-MEANS,” pp. 111–122, 2020.