

Pengaruh Pendapatan Orang Tua terhadap Hasil Belajar Siswa Menggunakan Algoritma K-Means Clustering

M Arif Khairunnas¹, Asep Jamaludin², R Ibnu Adam³

^{1,2,3} Program Studi Informatika Universitas Singaperbangsa Karawang

Email : 1910631170104@student.unsika.ac.id

Abstrak

Pendidikan di Indonesia dianggap sebagai prinsip fundamental untuk memajukan bangsa, dengan peran penting keluarga dan tingkat pendapatan orang tua dalam membentuk hasil belajar siswa. Penelitian ini menggunakan data mining dengan algoritma k-means clustering untuk menganalisis pengaruh pendapatan orang tua terhadap hasil belajar siswa. Hasil belajar siswa menjadi indikator evaluasi pengetahuan formal, memperlihatkan pemahaman, motivasi, dan sikap siswa terhadap pembelajaran. Pentingnya meneliti hasil belajar siswa terletak pada penggunaannya sebagai indikator evaluasi dan identifikasi kelebihan serta kelemahan siswa. Algoritma k-means clustering, yang telah terbukti efektif dalam penelitian sebelumnya, digunakan untuk mempercepat proses pengelompokan data. Studi ini diharapkan dapat memberikan kontribusi pada pemahaman hubungan antara pendapatan orang tua dan hasil belajar siswa, serta menggambarkan efektivitas algoritma k-means clustering dalam konteks pendidikan.

Kata Kunci : *Pendidikan, Algoritma K-Means, Data Mining*

Abstract

Education in Indonesia is considered a fundamental principle for advancing the nation, with the important role of the family and parental income levels in shaping student learning outcomes. This research uses data mining with the k-means clustering algorithm to analyze the influence of parental income on student learning outcomes. Student learning outcomes become indicators of formal knowledge evaluation, showing students' understanding, motivation and attitudes towards learning. The importance of researching student learning outcomes lies in its use as an evaluation indicator and identification of student strengths and weaknesses. The k-means clustering algorithm, which has been proven effective in previous research, is used to speed up the data clustering process. This study is expected to contribute to the understanding of the relationship between parental income and student learning outcomes, as well as illustrate the effectiveness of the k-means clustering algorithm in an educational context.

Keywords: *Education, K-Means Algorithm, Data Mining*

PENDAHULUAN

Pendidikan memiliki peran krusial dalam memajukan suatu bangsa dan mengatasi ketertinggalan serta ketidaktahuan. Di Indonesia, pendidikan dianggap sebagai fondasi utama yang diamanahkan oleh Pasal 31 (1) UUD 1945, di mana pemerintah bertekad untuk memperkuat iman, ketakwaan, dan akhlak dalam pendidikan nasional guna membentuk karakter bangsa. Dalam perspektif psikologis, pendidikan diartikan sebagai suatu proses transformatif di mana individu mengalami perubahan perilaku melalui interaksi dengan lingkungan dan pemenuhan kebutuhan. Transformasi ini tercermin dalam berbagai aspek perilaku serta konsep belajar yang menjadi penanda perubahan tingkah laku individu dengan

tujuan mengubah pola tingkah laku secara umum melalui pengalaman dalam menghadapi lingkungan (Yannuansa. dkk., 2021).

Faktor internal dan eksternal memainkan peran penting dalam hasil belajar. Keluarga, sebagai faktor eksternal, memiliki dampak signifikan terhadap pencapaian siswa dalam belajar. Tingkat pendidikan orang tua kerap memberikan dorongan besar terhadap pendidikan anak-anak mereka, tercermin dalam pencapaian akademik siswa sebagai hasil dari proses pembelajaran. Hasil belajar merupakan cerminan dari upaya siswa selama terlibat dalam aktivitas belajar di sekolah yang tercermin dalam bentuk nilai (Yannuansa, dkk., 2021).

Evaluasi hasil belajar siswa memiliki peran penting sebagai indikator penilaian terhadap pengetahuan yang diperoleh dari pendidikan formal. Hal ini tercermin dalam nilai tes dan pencapaian dalam mata pelajaran yang dinilai oleh guru. Hasil belajar siswa membawa informasi tentang pemahaman siswa terhadap materi yang diajarkan, kemampuan, motivasi, bakat, minat, dan sikap siswa terhadap program pembelajaran. Selain itu, hasil belajar siswa juga memungkinkan evaluasi terhadap pencapaian sesuai dengan standar kompetensi dan kompetensi dasar yang ditetapkan, serta identifikasi kelebihan dan kelemahan siswa dalam proses pembelajaran (Sembiring, dkk., 2022).

Peran tingkat penghasilan orang tua juga turut berperan dalam memenuhi kebutuhan pendidikan siswa, memfasilitasi optimalisasi proses belajar-mengajar. Penghasilan merujuk pada penerimaan individu atau perusahaan sebagai hasil dari aktivitas tertentu, yang dapat berupa gaji, sewa, upah, bunga, atau laba dalam jangka waktu tertentu. Penelitian akan memanfaatkan teknik data mining, khususnya algoritma k-means, untuk mengamati pengaruh pendapatan orang tua terhadap hasil belajar siswa.

Data mining memanfaatkan statistik, matematika, kecerdasan buatan, dan pembelajaran mesin untuk mendapatkan informasi berharga dari berbagai basis data besar. Penguasaan materi pembelajaran menjadi faktor penting dalam menilai prestasi siswa, yang menjadi kebijakan kunci dalam menentukan pencapaian mereka. Salah satu teknik dalam data mining, clustering, adalah proses pengelompokan data secara otomatis tanpa informasi label kelas. Metode k-means, salah satu algoritma klasterisasi yang populer, digunakan karena keunggulannya dalam implementasi yang mudah.

Penelitian sebelumnya telah menerapkan Algoritma K-Means, seperti penelitian oleh Shefia Natalia Br Sembiring, Hendryan Winata, dan Sri Kusnasari (2020) yang mengamati pengelompokan prestasi siswa di SD Norambe dengan Algoritma K-means, serta penelitian oleh Aceng Supriyadi, Agung Triayudi, dan Ira Diana Sholihati (2021) yang membandingkan algoritma K-Means dengan K-Medoids pada pengelompokan armada kendaraan truk berdasarkan produktivitas. Dalam konteks ini, penelitian ini berjudul: "PENGARUH PENDAPATAN ORANG TUA TERHADAP HASIL BELAJAR SISWA MENGGUNAKAN ALGORITMA K-MEANS CLUSTERING." Penelitian ini akan menggunakan metode elbow untuk menentukan jumlah cluster yang optimal dan memanfaatkan rapidminer sebagai alat bantu untuk perhitungan yang efisien.

METODE

Untuk mencapai tujuan penelitian ini, digunakan metode atau tahapan yang mengacu pada metodologi Knowledge Discovery in Database (KDD) yang terdiri dari lima langkah, yaitu *data selection*, *data preprocessing*, *data transformation*, *data mining*, dan *evaluation*

HASIL DAN PEMBAHASAN

Data Selection

Penelitian ini mengenai proses data selection pada dataset dari SMAIT Ulil Albab. Tahapan data selection dilakukan untuk memilih data yang relevan dari dataset awal yang terdiri dari beberapa atribut seperti nama wali kelas, nama siswa, nilai kriteria ketuntasan minimal (KKM), nilai keterampilan dan pengetahuan setiap mata pelajaran, nilai akhlak, kepribadian, absensi siswa, akumulasi nilai, nilai rata-rata rapor kelas XI IPA dan IPS, nama orang tua, pekerjaan, penghasilan, dan alamat orang tua.

Tabel 1. Dataset Awal

Atribut	Tipe Data	Keterangan
Nama wali kelas	String	Nama wali kelas XI IPA dan IPS SMAIT Uilil Albab
Nama siswa	String	Menyebutkan semua nama siswa kelas XI IPA dan IPS di SMAIT Uilil Albab
Nilai kriteria ketuntasan minimal (KKM)	Numerik	Menunjukkan standart nilai capaian yang harus dicapai siswa pada setia mata pelajarannya
Nilai keterampilan	Numerik	Menunjukkan nilai keterampilan setiap mata pelajaran semua siswa kelas XI SMAIT Uilil Albab
Nilai Pengetahuan	Numerik	Menunjukkan nilai pengetahuan setiap mata pelajaran semua siswa kelas XI SMAIT Uilil Albab
Nilai Akhlaq	Boolean	Menunjukkan kategori akhlaq yang dimiliki pada setiap siswa kelas XI SMAIT Uilil Albab
Nilai Kepribadian	Boolean	Menunjukkan kategori kepribadian yang dimiliki pada setiap siswa kelas XI SMAIT Uilil Albab
Absensi (Kehadiran siswa)	Numerik	Menunjukkan jumlah sakit, alfa, dan izin setiap siswa kelas XI SMAIT Uilil Albab
Jumlah akumulasi nilai	Numerik	Menunjukkan hasil akumulasi semua nilai setiap siswa kelas XI SMAIT Uilil Albab
Nilai rata-rata rapor	Numerik	Menunjukkan hasil nilai rata-rata nilai setiap siswa kelas XI SMAIT Uilil Albab
Nama orang tua	String	Menunjukkan nama lengkap ayah dan ibu semua siswa kelas XI SMAIT Uilil Albab
Pekerjaan Orang tua	String	Menunjukkan pekerjaan ayah dan ibu semua siswa kelas XI SMAIT Uilil Albab
Penghasilan Orang tua	String	Menunjukkan penghasilan ayah dan ibu semua siswa kelas XI SMAIT Uilil Albab
Alamat	String	Menunjukkan alamat tempat tinggal orang tua siswa disertai kelurahan, kecamatan, kodepos, kabupaten dan provinsi tempat tinggal orang tua semua siswa kelas XI SMAIT Uilil Albab

Tabel 2. Dataset Selection Data

Atribut	Tipe Data	Keterangan
Nama siswa	String	Menyebutkan semua nama siswa kelas XI IPA dan IPS di SMAIT Uilil Albab
Nilai rata-rata rapor	Numerik	Menunjukkan hasil nilai rata-rata nilai setiap siswa kelas XI SMAIT Uilil Albab
Penghasilan Orang tua	String	Menunjukkan penghasilan ayah dan ibu semua siswa kelas XI SMAIT Uilil Albab

Tabel 1 ini memuat beragam atribut dengan tipe data dan penjelasan keterangan masing-masing atribut. Beberapa atribut yang disebutkan di antaranya adalah nama wali kelas, nama siswa, nilai KKM, nilai keterampilan, nilai pengetahuan, nilai akhlaq, nilai kepribadian, absensi siswa, akumulasi nilai, nilai rata-rata rapor, nama orang tua, pekerjaan orang tua, penghasilan orang tua, dan alamat orang tua. Dalam tahap seleksi data, peneliti memilih tiga atribut yang dianggap relevan untuk penelitian ini, yaitu nama siswa, nilai rata-rata rapor, dan penghasilan orang tua. Ketiga atribut ini dipilih untuk digunakan dalam proses preprocessing.

Tabel 4.2 menunjukkan hasil dari seleksi data, yang terdiri dari tiga atribut yang dipilih: nama siswa, nilai rata-rata rapor kelas XI SMAIT Uilil Albab, dan penghasilan orang tua. Peneliti melakukan seleksi data untuk memfokuskan analisis pada atribut-atribut tertentu yang dianggap relevan dalam proses penelitian. Dengan mempersempit atribut yang digunakan, diharapkan dapat memudahkan dalam proses analisis dan pengolahan data yang lebih fokus.

Data Preprocessing

Tabel 3. Dataset Hasil Selection Data

No	Nama	Rata-Rata	Pendapatan Orang tua
1	ALZENA	88	Rp.6.000.000 - Rp.10.000.000
2	KHANSA ZAKIYAH	88	Rp. 5.800.000
3	M. IKHWAN AL FAJRI	90	Rp. 5.800.000
4	MUHAMMAD SYUKUR HIFZHAN	88	Rp. 3.000.000 - Rp. 6.000.000
5	NADIA FATMA ARIANTI	85	Lebih dari Rp. 5.000.000
6
19	SHAHIRA AULIYA LIM	86	Rp.2.000.000 - Rp.5.000.000
20	SILVI INDRI SARI	82	Rp. 3.000.000 - Rp. 6.000.000
21	SYALU AULIA HAKIM	84	Rp.3.000.000 - Rp.6.000.000
22	WILDA RAHMANYAH	86	Rp. 3.000.000 - Rp. 6.000.000

Tabel 4. Dataset Hasil Cleaning Data

No	Nama	Rata-Rata	Pendapatan Orang tua (Rp. -000)
1	ALZENA	88	6000
2	KHANSA ZAKIYAH	88	5800
3	M. IKHWAN AL FAJRI	90	5800
4	MUHAMMAD SYUKUR HIFZHAN	88	5000
5	NADIA FATMA ARIANTI	85	10000
6	10000
19	SHAHIRA AULIYA LIM	86	5800
20	SILVI INDRI SARI	82	3000
21	SYALU AULIA HAKIM	84	3000
22	WILDA RAHMANYAH	86	5000

Data preprocessing merupakan langkah penting sebelum memulai proses data mining karena memastikan kualitas dataset sebelum pemodelan. Tahapan utama dalam data preprocessing adalah Cleaning Data dan Feature Engineering. Cleaning Data melibatkan penghapusan missing value dan outliers dari dataset. Missing value terjadi ketika kolom atribut tidak memiliki nilai atau hanya berisi simbol seperti "-", sementara outliers dalam kategori data merupakan nilai yang ekstrim jika dibandingkan dengan data lainnya. Setelah proses selection data, dilakukan pembersihan data dengan menghilangkan simbol "-" yang menandakan "sampai" pada data pendapatan orang tua. Selanjutnya, angka "-000" dihapus agar nilai dapat lebih mudah diinterpretasikan oleh algoritma. Hasil dari proses cleaning data dapat dilihat pada Tabel 4.

Nilai	Income	Nama_ALZENA	Nama_CHALIEF AKBAR SETIAWAN	Nama_DAHILA PUTRI	Nama_DHAFA FAUZAN RISHANDA	Nama_EDWIN MAJIB HIDAYAT	Nama_KHANSA ZAKIYAH	Nama_M. IKHWAN AL FAJRI	Nama_MUHAIMIN SAPUTRA	...	Nama_NAUFAL TAQIUDIN YUSUF
0	88	6000	1	0	0	0	0	0	0	...	0
1	88	5800	0	0	0	0	1	0	0	...	0
2	90	5800	0	0	0	0	0	1	0	...	0
3	88	5000	0	0	0	0	0	0	0	...	0
4	88	5800	0	0	0	0	0	0	0	...	0
5	89	5000	0	0	0	0	0	0	0	...	0

Gambar 1. Hasil One-Hot Encoding

	Nilai	Income	Nama_ALZENA	Nama_CHALIEF AKBAR SETIAWAN	Nama_DAHILA PUTRI	Nama_DHAFA FAUZAN RISHANDA	Nama_EDWIN MAJIB HIDAYAT	Nama_KHANSA ZAKIYAH	Nama_M. IKHWAN AL FAJRI	Nama_MUHAIMIN SAPUTRA	...	Nama_NAUFAL TAQIUDIN YUSUF	Nama_N AFIF
count	22.000000	22.000000	22.000000	22.000000	22.000000	22.000000	22.000000	22.000000	22.000000	22.000000	...	22.000000	22.000000
mean	86.727273	3545.454545	0.045455	0.045455	0.045455	0.045455	0.045455	0.045455	0.045455	0.045455	...	0.045455	0.045455
std	2.548236	1959.768517	0.213201	0.213201	0.213201	0.213201	0.213201	0.213201	0.213201	0.213201	...	0.213201	0.213201
min	82.000000	1000.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.000000
25%	85.000000	2000.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.000000
50%	86.500000	3000.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.000000
75%	88.000000	5800.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.000000
max	93.000000	6000.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	...	1.000000	1.000000

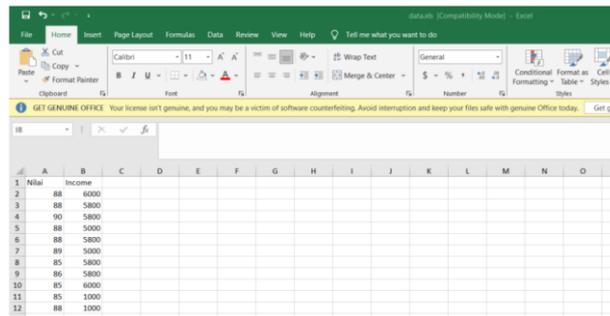
8 rows x 24 columns

Gambar 2. Hasil Feature Engineering

Metode One-Hot Encoding digunakan dalam proses pembersihan data untuk mengubah nilai atribut menjadi 0 atau 1, selain penghapusan data yang tidak diperlukan. Tujuan dari metode ini adalah untuk memastikan bahwa proses clustering dapat memberikan hasil yang lebih akurat dalam pemodelan. Penerapan metode ini dilakukan menggunakan pemrograman Python, dan hasilnya dapat dilihat pada Gambar 4.1. Setelah tahap One-Hot Encoding, langkah selanjutnya adalah melakukan Feature Engineering. Proses ini merupakan bagian dari preprocessing data yang bertujuan untuk mengembangkan serta memilih fitur yang akan digunakan dalam analisis data atau pembuatan model machine learning. Feature engineering menjadi tahap yang krusial karena kualitas fitur yang dihasilkan akan berdampak pada hasil akhir. Tahapan ini juga dilakukan menggunakan Bahasa pemrograman python, dengan hasilnya tergambar pada Gambar 4.2.

Data Transformation

Pada tahap data transformation dilakukan perubahan data menjadi data yang diperlukan dalam memining data. Pada tahapan ini tidak semua atribut hasil data cleaning yang akan diproses dikarenakan tidak seluruh atribut datanya relevan atau tidak berpengaruh pada hasil yang diperlukan. Tujuan dalam menghilangkan atribut data yang tidak relevan ini adalah untuk memudahkan proses clustering. Atribut yang dihilangkan adalah nama siswa. Perubahan dilakukan dengan bantuan Bahasa pemrograman python. Berikut hasilnya pada Gambar 4.3. Selain itu jenis data disimpan dalam format .csv hal tersebut dilakukan untuk mempermudah proses data mining.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	Nilai	Income													
2		88	6000												
3		88	5800												
4		90	5800												
5		88	5000												
6		88	5800												
7		89	5000												
8		85	5800												
9		86	5800												
10		85	6000												
11		85	1000												
12		88	1000												

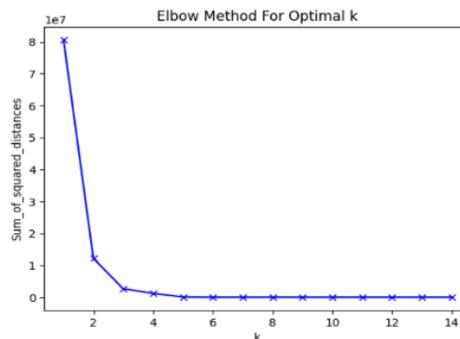
Gambar 3. Hasil Transformation

Data Mining

Pada tahap data mining, pemodelan dilakukan dengan algoritma K-Means Clustering. Metode ini dijalankan melalui dua percobaan menggunakan bahasa pemrograman python dan aplikasi rapid miner. Langkah-langkah yang terlibat dalam proses clustering data menggunakan algoritma K-Means meliputi:

1. Memilih K sebagai titik centroid.

Peneliti menggunakan metode Elbow untuk menentukan nilai K yang optimal. Melalui metode ini, penurunan nilai yang signifikan antara cluster pertama dan kedua menjadi kriteria dalam menentukan jumlah cluster terbaik, yang terlihat sebagai sudut pada grafik (Sulistiyawan, Hapsery & Arifanahum, 2021). Penerapan metode ini dilakukan dalam bahasa pemrograman python, dan hasilnya didasarkan pada dataset yang digunakan oleh peneliti, seperti yang terlihat pada Gambar 4.4.



Gambar 4. Grafik Hasil nilai K dengan Metode Elbow

Pada Gambar 4.4 terlihat adanya penurunan yang signifikan pada nilai k, diikuti dengan penurunan yang lebih lambat hingga mencapai nilai k yang stabil. Terjadi penurunan yang paling signifikan dari k=1 ke k=2, dan dari k=2 terjadi penurunan kembali yang cenderung jauh ke k=3 namun dari k=3 seterusnya jarak cenderung konstan. Oleh karena itu, dapat disimpulkan bahwa k=3 merupakan jumlah cluster yang optimal atau terbaik.

2. Menentukan posisi awal centroid secara acak.

Penelitian ini menggunakan tiga cluster dan menentukan posisi awal centroid secara acak, dengan peneliti menguji coba 2, 3, 4, dan 5 sebagai posisi awal centroid secara berturut-turut. Hasil evaluasi menggunakan metode silhouette coefficient menunjukkan rata-rata nilai 0.77, 0.82, 0.93, dan 0.96 untuk masing-masing percobaan. Silhouette coefficient berperan dalam menilai kualitas dan kekuatan cluster serta mengukur tingkat kedekatan objek di dalamnya (Prima, 2021). Hasil eksperimen menggunakan pemrograman Python tergambar dalam Gambar 4.5.

```
range_n_clusters = [2, 3, 4, 5]
for n_cluster in range_n_clusters:
    kmeans = KMeans(n_clusters = n_cluster, max_iter =1000)
    kmeans.fit(dfkmeans)
    labels = kmeans.predict(dfkmeans)
    from sklearn.metrics import silhouette_score
    silhouette_avg = silhouette_score(dfkmeans, labels)
    print(silhouette_avg)

0.7706585044065096
0.8197928405530214
0.9320744376507828
0.9632369444433649
```

Gambar 5. Posisi awal centroid dan rata-rata nilai rata-rata silhouette

3. Data Process Training

Seperti yang dijelaskan pada point sebelumnya proses dilakukan secara berulang untuk memperoleh hasil clustering yang baik. Dengan total data = 22 orang, dan jumlah cluster = 3, ditemukan koordinat cluster pusat yang terbentuk terlihat pada Gambar 4.6. berikut:

```
[[ 86.         1285.71428571]
 [ 87.44444444 5666.66666667]
 [ 86.5        3000.        ]]
```

Gambar 6. Koordinat Cluster Pusat

Berdasarkan proses training yang dilakukan diperoleh label yaitu cluster yang digunakan meliputi cluster 1 yaitu 0 pendapatan orang tua rendah, cluster 2 yaitu 1 pendapatan orang tua tinggi dan cluster 3 yaitu 2 pendapatan orang tua sedang. Terlihat pada Gambar 4.7 berikut:

```
print(kmeans.labels_)
[1 1 1 1 1 1 1 1 0 0 0 0 0 0 2 2 2 2 2 2]
```

Gambar 7. Label Cluster

4. Hasil Clustering

Setelah dilakukan seluruh proses K-Means dengan bantuan python dan rapid miner, dimana hasil dari kedua proses tersebut memperoleh hasil yang sama yaitu nilai cluster 1 dengan label 0 berada pada rata-rata nilai siswa 86 dengan pendapatan orang tua rata-rata Rp. 1.333.000 sedangkan pada cluster 2 dengan label 1 nilai rata-rata siswa 88 dengan pendapatan orang tua Rp. 5.650.000 dan cluster 3 dengan label 2 nilai rata-rata siswa 87 dengan pendapatan orang tua Rp. 3.000.000. Berdasarkan hasil tersebut diperoleh kesimpulan bahwa nilai yang diperoleh siswa kelas XI SMAIT Ulil Albab dipengaruhi oleh pendapatan orang tua. Berikut hasil perhitungan yang dilakukan berdasarkan Bahasa pemrograman python dan rapid miner dengan bantuan Ms. Excel dalam proses kalkulasi keseluruhan data. Terlihat pada Gambar 4.8.

	Avarage				
	Nilai	Income	Cluster		
array([[88, 1000, 0],	86	1333	0	Rata-Rata Nilai Rendah = Income Rendah	
[82, 1000, 0],	88	5650	1	Rata-Rata Nilai Tinggi = Income Tinggi	
[85, 2000, 0],	87	3000	2	Rata-Rata Nilai Sedang = Income Sedang	
[82, 1000, 0],					
[87, 1000, 0],					
[93, 2000, 0],					
[85, 1000, 0],					
[85, 6000, 1],					
[88, 6000, 1],					
[85, 5000, 1],					
[89, 5000, 1],					
[88, 5000, 1],					
[88, 5000, 1],					
[90, 5000, 1],					
[88, 5000, 1],					
[86, 5000, 1],					
[84, 3000, 2],					
[89, 3000, 2],					
[88, 3000, 2],					
[86, 3000, 2],					
[86, 3000, 2],					

Gambar 4. 8 Hasil Clustering dengan K-Means

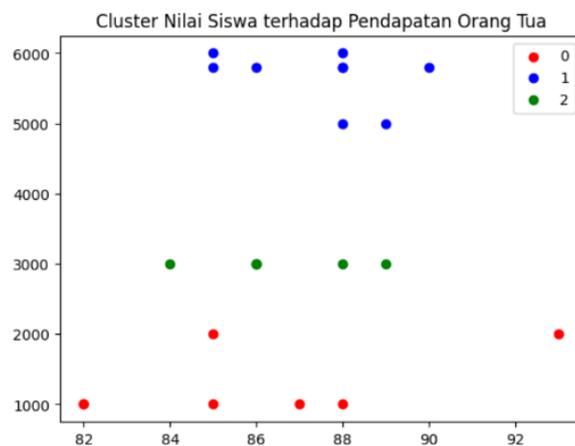
Cluster Model

Cluster 0: 6 items
 Cluster 1: 7 items
 Cluster 2: 9 items
 Total number of items: 22

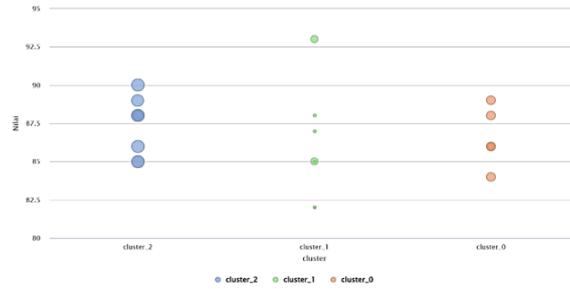
Gambar 4. 9 Cluster Model dengan Rapid Miner

Evaluasi

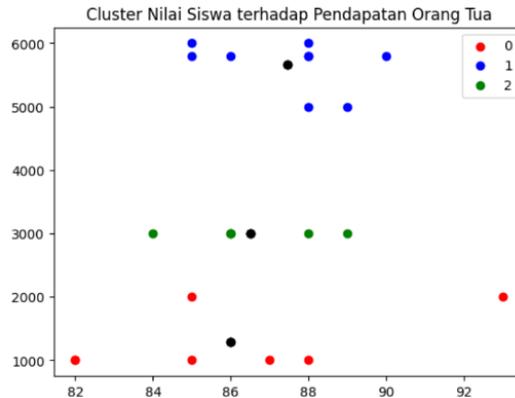
Setelah diperoleh hasil clustering selanjutnya dilakukan evaluasi model, terlihat model yang diperoleh pada Gambar 4.9. Berdasarkan model tersebut selanjutnya dilakukan evaluasi atau diuji untuk mengukur efektivitas kinerja hasil perhitungannya menggunakan uji validitas silhouette coefficient. Rata-rata model berdasarkan Silhouette coefficient terlihat pada Gambar 4.5 namun untuk melihat hasil dengan jelas dilakukan visualisasi model untuk memperjelas informasi yang diperoleh, visualisasi hasil clustering dengan K-Means sebagai berikut:



Gambar 10. Visualisasi Hasil Clustering di Python Programming

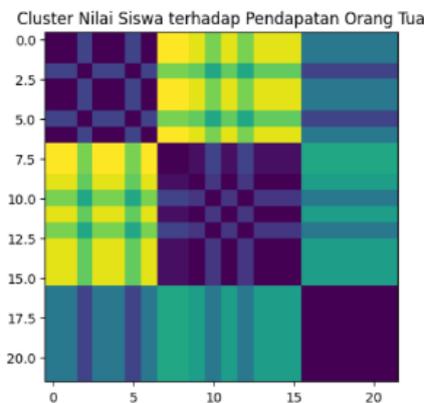


Gambar 11. Visualisasi Hasil Clustering di Rapid Miner

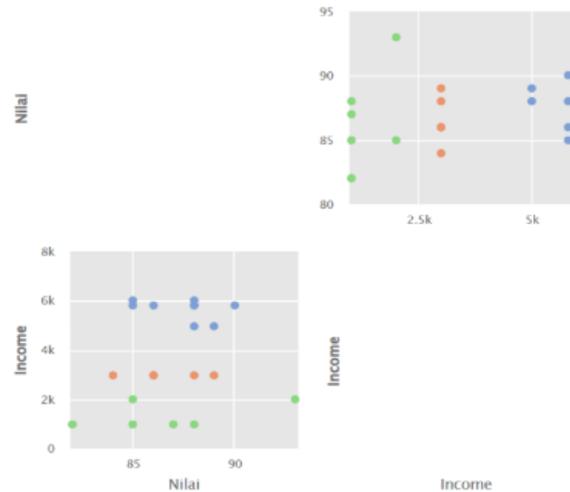


Gambar 12. Visualisasi Hasil Clustering dengan nilai Centroid di Python Programming

untuk memperjelas hasil evaluasi model, peneliti menerapkan metode similarity matrix. Similarity matrix adalah matriks yang elemen-elemennya mengukur seberapa mirip suatu pasangan satu sama lain. Semakin tinggi nilai ukuran maka semakin besar kemiripan antara keduanya. Semakin kecil nilai ukuran maka semakin besar ketidaksamaannya. Ukuran kesamaan yang digunakan bergantung pada tipe objek yang dievaluasi (Emifoniye, S. 2023). Hasil yang diperoleh terlihat pada Gambar 4.13 dan Gambar 4.14.



Gambar 13. Similarity Matrix di Python Programming



Gambar 14. Similarity Matrix di Rapid Miner

Pembahasan

Penelitian dengan algoritma clustering pada data penghasilan orang tua terhadap nilai hasil evaluasi siswa kelas XI SMAIT Ulil Albab dengan menggunakan algoritma K-Means bertujuan apakah terdapat pengaruh atau tidak antar keduanya. Berdasarkan hasil penelitian dengan bantuan python programming dan aplikasi rapid miner dalam proses clustering diperoleh hasil bahwa pendapatan orang tua siswa mempengaruhi nilai hasil belajar siswa. Pernyataan tersebut diperoleh melalui beberapa tahapan yaitu data selection, data preprocessing, data transformation, data mining, dan evaluasi model. Pada tahap data preprocessing diterapkan 2 metode yaitu data cleaning dengan metode OneHot Encoding dan metode Feature Engineering dalam proses pembersihan data. Sedangkan pada proses data mining diterapkan algoritma K-Means untuk menentukan cluster pada dataset dalam penarikan kesimpulan. Setelah proses memining data dilakukan evaluasi model dimana hasil clustering dari python programming dan aplikasi rapid miner ditampilkan dalam visualisasi untuk memperjelas informasi yang diperoleh serta penerapan metode Silhouette coefficient dan similarity matrix dalam penarikan kesimpulan hasil evaluasi model.

Pada penelitian yang dilakukan (Yannuansa, dkk., 2021) pemanfaatan algoritma K-Means Clustering dalam mengolah pengaruh pendapatan orang tua terhadap hasil belajar pada mata pelajaran produktif diperoleh hasil bahwa kelompok siswa yang termasuk mendapatkan hasil kurang dimana orang tua siswa mendapatkan penghasilan pas-pasan, dan siswa yang mendapatkan hasil baik orang tua siswa adalah golongan dari orang mampu karena untuk menunjang nilai prdofuktif diperlukan alat-alat praktikum yang memadai dan siswa dapat menguasai terlebih kamera dan alat video recorder, sedangkan siswa yang masuk kedalam kelompok sedang orang tua berpenghasilan standar yaitu golongan menengah. Berdasarkan penelitian terdahulu disimpulkan bahwa pendapatan orang tua mempengaruhi hasil belajar siswa.

Kemampuan K-Means dalam melakukan proses clustering menurut (Wanto. dkk., 2020) sangat baik karena tujuan yang ingin dicapai adalah untuk meminimalkan total varians intracluster atau fungsi kesalahan kuadrat. Dengan menerapkan teori tersebut menurut (Yannuansa, dkk., 2021) mampu menghasilkan algoritma yang sederhana namun hasil yang powerfull. Selain itu K-Means adalah algoritma yang paling umum digunakan dalam proses clustering karena proses yang sederhana dan penerapannyapun mudah dilakukan, K-Means dapat diterapkan dengan bantuan aplikasi data mining seperti rapid miner dan bahkan dapat diselesaikan dengan bantuan library pada python programming sehingga K-Means adalah metode yang layak untuk digunakan dalam proses data mining (Diwa Oktario Dacwanda & Yeissica Nataliani, 2021).

SIMPULAN

Berdasarkan hasil penelitian yang dilakukan maka diperoleh kesimpulan sebagai berikut:

1. Berdasarkan hasil penelitian yang dilakukan dengan menggunakan algoritma K-Means Clustering pada aplikasi rapid miner dan python programming pendapatan orang tua mempengaruhi prestasi belajar siswa kelas XI di SMAIT Ulil Albab. Dalam proses penelitian diperoleh nilai cluster 1 dengan label 0 berada pada rata-rata nilai siswa 86 dengan pendapatan orang tua rata-rata Rp. 1.333.000 sedangkan pada cluster 2 dengan label 1 nilai rata-rata siswa 88 dengan pendapatan orang tua Rp. 5.650.000 dan cluster 3 dengan label 2 nilai rata-rata siswa 87 dengan pendapatan orang tua Rp. 3.000.000. Hal tersebut membuktikan bahwa peningkatan prestasi siswa dapat meningkat jika kemampuan finansial dalam pemenuhan kebutuhan pembelajaran dapat terpenuhi.
2. K-Means Clustering adalah metode yang digunakan dalam penelitian ini, penerapan metode ini dilakukan dengan beberapa tahapan yaitu data selection, data preprocessing, data transformation, data mining, dan evaluasi model. Pada tahap data preprocessing diterapkan 2 metode yaitu data cleaning dengan metode One-Hot Encoding dan metode Feature Engineering dalam proses pembersihan data. Sedangkan pada proses data mining diterapkan algoritma K-Means untuk menentukan cluster pada dataset dalam penarikan kesimpulan. Setelah proses memining data dilakukan evaluasi model dimana hasil clustering dari python programming dan aplikasi rapid miner ditampilkan dalam visualisasi untuk memperjelas informasi yang diperoleh serta penerapan metode Silhouette coefficient dan similarity matrix dalam penarikan kesimpulan hasil evaluasi model.

DAFTAR PUSTAKA

- Agustin, F. E., Fitria, A., & Hanifah, A. (2009). *Implementasi Algoritma KMeans Untuk Menentukan Kelompok Pengayaan Materi Mata Pelajaran Ujian Nasional (Studi Kasus: SMP Negeri 101 Jakarta)*. Jakarta: Universitas Islam Negeri Syarif Hidayatullah.
- Antonio, D., & Alatubir, V. (2017). *THE IMPLEMENTATION OF K-MEANS ALGORITHM FOR CLUSTERING THE SENIOR HIGH SCHOOLS IN THE PROVINCE OF YOGYAKARTA SPECIAL REGION BASED ON THE ABSORPTION CAPACITY SCORE OF BAHASA INDONESIA NATIONAL EXAMINATION*.
- Darmi, Y., Setiawan, A., Bali, J., Kampung Bali, K., Teluk Segara, K., & Bengkulu, K. (2016). PENERAPAN METODE CLUSTERING K-MEANS DALAM PENGELOMPOKAN PENJUALAN PRODUK. Dalam *Jurnal Media Infotama* (Vol. 12, Nomor 2).
- Fahmi, R. (2021). Clustering dan Aplikasinya dalam Data Mining. *Jurnal Informatika*, 3(1), 1-10.
- Fitrianingsih, B. G. S., & Rosyid, R. (2016). *PENGARUH PENDAPATAN ORANG TUA TERHADAP TINGKAT PENDIDIKAN ANAK DESA SUNGAI ASAM KABUPATEN KUBU RAYA*.
- Harahap, B. (t.t.). *Penerapan Algoritma K-Means Untuk Menentukan Bahan Bangunan Laris (Studi Kasus Pada UD. Toko Bangunan YD Indarung)*.
- HARYAJI, B. (2018). *PENERAPAN ALGORITMA K-MEANS UNTUK MEMETAKAN GARIS KEMISKINAN MENURUT PROVINSI DI INDONESIA*.
- Ikhwan, A., & Aslami, N. (2020). IMPLEMENTASI DATA MINING UNTUK MANAJEMEN BANTUAN SOSIAL MENGGUNAKAN ALGORITMA K-MEANS. *Jurnal Teknologi Informasi*, 4(2).
- Indrawan, B. (2018). *PENERAPAN ALGORITMA K-MEANS CLUSTERING UNTUK MENENTUKAN STRATEGI PROMOSI UNIVERSITAS ISLAM NEGERI SUNAN KALIJAGA YOGYAKARTA Skripsi*.
- Jannah, N., & Yulianto, T. (2016). *Mengelompokkan Siswa Berprestasi Akademik dengan Menggunakan Metode K Means Kelas VII MT*. 2(2).
- Kadek, I., Arta, J., Indrawan, G., Gede, & Dantes, R. (2019). DATA MINING REKOMENDASI CALON MAHASISWA BERPRESTASI DI STMIK DENPASAR MENGGUNAKAN

- METODE TECHNIQUE FOR OTHERS REFERENCE BY SIMILARITY TO IDEAL SOLUTION. *Jurnal Ilmu Komputer Indonesia (JIKI)*, 4(1).
- Kustanti, E., Ba'diah, U., Sinawati, & Fadlan, M. (2018). Klasifikasi Siswa berdasarkan Nilai pada Bidang Ekstrakurikuler Menggunakan Metode KNearest Neighbor. *Journal of Big Data Analytic and Artificial Intelligence*.
- Menhard. (2017). PENGARUH PENDAPATAN ORANG TUA DAN STATUS SOSIAL TERHADAP MINAT MELANJUTKAN PENDIDIKAN KE PERGURUAN TINGGI (STUDI KASUS STIE MAHAPUTRA RIAU). *Jurnal Wira Ekonomi Mikroskil*.
- Missa, D., Achmadi, S., & Mahmudi, A. (2021a). PENERAPAN METODE CLUSTERING DENGAN ALGORITMA K-MEANS PADA PENGELOMPOKAN DATA PENGHASILAN ORANG TUA SISWA. Dalam *Jurnal Mahasiswa Teknik Informatika* (Vol. 5, Nomor 1).
- Natalia Br Sembiring, S., Winata, H., Kusnasari, S., Informasi, S., & Triguna Dharma, S. (2022). *Pengelompokan Prestasi Siswa Menggunakan Algoritma K-Means*.
- Nishom, M. (2019). Pengelompokan Data Menggunakan Metode K-Means Clustering. *Jurnal Teknologi Informasi dan Komputer*, 2(1), 42-49.
- Oktario Dacwanda, D., & Nataliani, Y. (2021). Implementasi k-Means Clustering untuk Analisis Nilai Akademik Siswa Berdasarkan Nilai Pengetahuan dan Keterampilan. *AITI: Jurnal Teknologi Informasi*, 18(Agustus), 125–138.
- Pratiwi, L. (2013). (2013). *IMPLEMENTASI ALGORITMA K-MEANS DALAM DATA MINING UNTUK PEMINATAN JURUSAN BAGI SISWA KELAS X SMA NEGERI 29 JAKARTA PROGRAM STUDI TEKNIK INFORMATIKA FAKULTAS SAINS DAN TEKNOLOGI UNIVERSITAS ISLAM NEGERI SYARIF HIDAYATULLAH JAKARTA*.
- Rahman, F., Muhammad, D., & Firdaus, I. (2016). PENERAPAN DATA MINING METODE NAÏVE BAYES UNTUK PREDIKSI HASIL BELAJAR SISWA SEKOLAH MENENGAH PERTAMA (SMP). Dalam *AI Ulum Sains dan Teknologi* (Vol. 1, Nomor 2). <http://archive.ics.uci.edu>
- Sumantri, B., & Utami, H. N. (2018). Tahapan Proses Knowledge Discovery in Database (KDD) dalam Data Mining. *Jurnal Sistem dan Informatika*, 8(1), 48-56.
- Supriyadi, A., Triayudi, A., Sholihati, I. D., & Sawo Manila, J.(2021). *PERBANDINGAN ALGORITMA K-MEANS DENGAN K-MEDOIDS PADA PENGELOMPOKAN ARMADA KENDARAAN TRUK BERDASARKAN PRODUKTIVITAS*.
- Waluyo, J. (2019). *Penerapan Metode K-means Clustering Dalam Penilaian Kedisiplinan Siswa Untuk Layanan Bimbingan Dan Konseling Di SMP Negeri 3 Temanggung*.
- Yannuansa, N., Safa'udin, M., & Aziz, M. I. (2021). *Pemanfaatan Algoritma K-Means Clustering dalam Mengolah Pengaruh Pendapatan Orang Tua Terhadap Hasil Belajar Pada Mata Pelajaran Produktif*.