Jurnal Sistem Komputer dan Informatika (JSON)
Volume 4, Nomor 2, Desember 2022
e-ISSN 2685-998X
DOI 10.30865/json.v4i2.5299

Penerapan Algoritma *K-Means* dan *Decision Tree* Dalam Analisis Prestasi Siswa Sekolah Menengah Kejuruan

Muhammad Bari Abdul Majid[1,*], Yusup Mad Cani[2], Ultach Enri[2]

[1]Fakultas Ilmu Komputer, Sistem Informasi, Universitas Singaperbangsa Karawang, Karawang, Indonesia
[2]Fakultas Ilmu Komputer, Informatika, Universitas Singaperbangsa Karawang, Karawang, Indonesia
Email: [1,*]1910631250021@student.unsika.ac.id, [2]1910631250054@student.unsika.ac.id, [3]ultach@staff.unsika.ac.id
Email Penulis Korespondensi: 1910631250021@student.unsika.ac.id
Submitted: **08/12/2022**; Accepted: **31/12/2022**; Published: **31/12/2022**

Abstrak–Penelitian ini bertujuan untuk mengetahui seberapa besar pengaruh dari penghasilan orang tua dan jarak rumah ke sekolah dalam prestasi siswa SMK Tarbiyatul Ulum. Karena jika mengetahui faktor yang mempengaruhi prestasi siswa, maka bisa diambil langkah atau tindakan untuk meningkatkan prestasi siswa. Metode yang dipakai dalam penelitian ini mecakup penggunaan klasterisasi dengan menggunakan K-Means lalu dilakukan pengklasifikasian dengan menggunakan metode Decision Tree dengan jumlah dataset yang dijadikan bahan penelitian ini sebanyak 157 data. Setelah diadakannya klasifikasi pemodelan menggunakan algoritma decision tree, maka didapati bahwa penghasilan orang tua tidak berpengaruh terhadap prestasi siswa, tetapi jarak rumah ke sekolah berpengaruh terhadap prestasi siswa. Lalu pada tahap evaluasi, algoritma decision tree tidak cocok digunakan dalam prediksi prestasi siswa, karena nilai akurasi dan nilai AUC dari algoritma ini adalah 68% dan 0.561, dimana nilai tersebut masuk ke dalam kategori failure.

Kata Kunci: Prestasi; Data Mining; K-Means; Decision Tree

Abstract–This study aims to determine how much influence parents' income and the distance from a student's home to school have on student achievement at SMK Tarbiyatul Ulum. Because if you know the factors that influence student achievement, steps or actions can be taken to improve student achievement. The method used in this study includes the use of clustering using K-Means and then classifying it using the Decision Tree method with a total of 157 datasets used as research material. After conducting the classification modeling using the decision tree algorithm, it was found that parents' income did not affect student achievement, but the distance from home to school did affect student achievement. Then at the evaluation stage, the decision tree algorithm is not suitable for use in predicting student achievement, because the accuracy and AUC values of this algorithm are 68% and 0.561, where these values fall into the failure category.

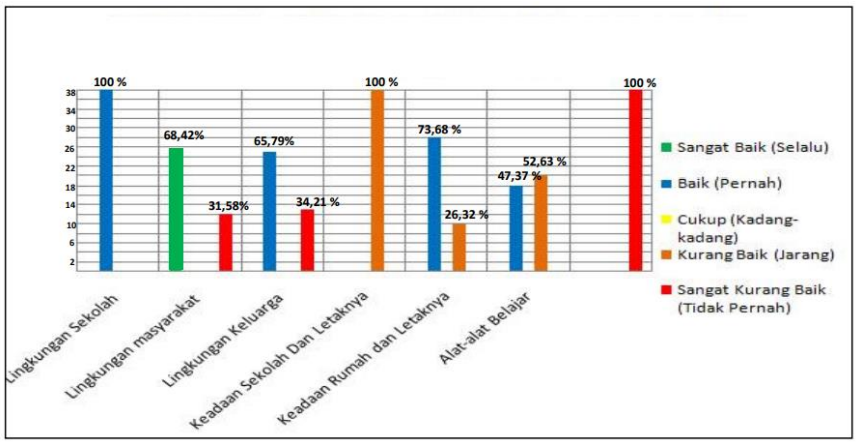
Keywords: Achivement; Data Mining; K-Means; Decision Tree

1. PENDAHULUAN

Pendidikan merupakan proses pemantapan diri guna mempersiapkan diri untuk bisa bersaing sesuai zaman yang berjalan. Pendidikan memiliki peranan penting guna membentuk sumber daya manusia yang berkualitas karena perkembangan kemajuan suatu bangsa ditentukan oleh sumber daya manusia yang ada di dalamnya. Secara umum pendidikan merupakan faktor utama dalam semua bidang kehidupan. Oleh karena itu, pendidikan menjadi kebutuhan masyarakat yang paling dianggap penting. Hasil belajar merupakan hal yang sangat diperhatikan oleh orang lain terutama oleh diri sendiri dan orang tua. Untuk mencapai hasil pembelajaran yang maksimal itu sendiri terkadang memiliki kendala dalam proses belajar. Faktor yang mempengaruhi proses belajar terdiri dari dua faktor yaitu faktor yang terdapat dalam diri siswa (internal) dan faktor yang terdiri dari luar siswa (eksternal). Faktor internal merupakan faktor yang berasal dari diri orang tersebut dan bersifat biologis sedangkan, faktor eksternal merupakan faktor yang berasal dari luar seperti keadaan lingkungan, keluarga, lingkungan sekolah, dan lingkungan masyarakat yang dapat mempengaruhi diri seseorang untuk belajar [1].

Dalam dunia pendidikan prestasi merupakan suatu aspek yang sangat penting dimana keberhasilan dalam memahami atau menguasai suatu materi itu merupakan sebagai bekal untuk masa depannya. Prestasi merupakan sebuah hasil yang didapat dari proses belajar yang telah dilakukan [2]. Prestasi dalam belajar juga dapat menimbulkan kepercayaan tersendiri bagi seseorang yang memiliki prestasi guna dapat bersaing dimasa depan. Keberhasilan pendidikan dalam suatu sekolah dapat dilihat dari prestasi yang diraih oleh sekolah tersebut. Prestasi merupakan tolak ukur yang menjadi acuan untuk mengetahui keberhasilan belajar seseorang dimana seseorang yang memiliki prestasi tinggi dinyatakan sudah berhasil dalam proses belajar, namun jika ada seseorang yang memiliki prestasi rendah maka terdapat faktor-faktor penghambat untuk siswa tersebut berprestasi. Fungsi dari prestasi itu sendiri merupakan cara untuk mengetahui kemajuan pemahaman pengetahuan nya setelah melaksanakan proses kegiatan belajar. Nilai yang diberikan oleh pengajar atau guru merupakan sebuah bukti hasil belajar seorang peserta didik [3]. Dalam hal ini, prestasi akademik menjadi ukuran tingkat pencapaian tujuan pembelajaran dengan dilihat dari hasil nilai Raport, baik itu prestasi akademik maupun non-akademik yang telah diperolehnya. Seperti yang sudah dijelaskan sebelumnya, bahwa terdapat 2 faktor yang mempengaruhi prestasi seorang siswa, faktor internal memang faktor paling penting untuk mempengaruhi prestasi seorang siswa, tapi faktor eksternal tidak kalah pentingnya dalam mempengaruhi prestasi seorang siswa. Salah satu contoh faktor eksternal adalah jarak rumah seorang siswa ke sekolahnya, penghasilan orang tua, lingkungan siswa tersebut, dan lainnya. Tapi apakah sebegitu pengaruhnya faktor eksternal terhadap prestasi seorang siswa? Terutama dalam hal jarak rumah seorang siswa ke sekolahnya dan juga penghasilan orang tua.

Jurnal Sistem Komputer dan Informatika (JSON)
Volume 4, Nomor 2, Desember 2022
e-ISSN 2685-998X
DOI 10.30865/json.v4i2.5299



Gambar 1. Grafik Faktor Eksternal Yang Mempengaruhi Hasil Belajar Siswa **[4]**

Grafik pada Gambar 1. adalah hasil penelitian yang dilakukan oleh Elisa dan teman-temannya, untuk mencari tahu kenapa hasil belajar siswa sekolah menengah atas pada mata pelajaran ekonomi itu rendah. Dari grafik yang ditampilkan dalam faktor eksternal yang mempengaruhi hasil belajar siswa, lingkungan keluarga mempunyai peranan penting dalam belajar siswa [4]. Dalam hal itu, maka akan diteliti lebih lanjut apakah faktor eksternal memang berpengaruh penting terhadap prestasi siswa di sekolah atau tidak. Dengan data yang dimiliki, yaitu penghasilan orang tua, jarak antara rumah ke sekolah dan hasil raport siswa, akan dilakukan penelitian apakah memang betul faktor eksternal dengan data yang ada, berpengaruh terhadap prestasi seorang siswa. Karena prestasi belajar bagi siswa sangat penting karena prestasi belajar merupakan salah satu gambaran tingkat keberhasilan dari kegiatan selama mengikuti pelajaran [5]. Dan dengan diketahuinya faktor mana saja yang dapat berpengaruh dalam prestasi siswa, bisa dilakukan tindakan lebih lanjut untuk meningkatkan prestasi siswa. Terdapat satu cara untuk membuktikan hal tersebut, yaitu dengan menggunakan metode klasifikasi *data mining*. Dengan klasifikasi yang ada pada *data mining*, kita dapat mengetahui apakah faktor eksternal berpengaruh terhadap prestasi siswa. Hal itu, bisa dilakukan dengan menghubungkan antara faktor eksternal dengan hasil prestasi seorang siswa, apakah hubungan tersebut saling berpengaruh atau tidak.

Suatu data yang ingin diolah, terutama dalam metode klasifikasi *data mining* memerlukan suatu label atau target agar data tersebut dapat diolah. Namun, terkadang data yang dimiliki dan akan diteliti tersebut belum memiliki label. Terdapat satu cara untuk mendapatkan label tersebut, yaitu dengan menggunakan metode klasterisasi *data mining*. Hasil dari klasterisasi tersebut bisa dijadikan label dalam data tersebut.

Banyak penelitian yang sudah dilakukan untuk melakukan analisis prestasi seorang siswa sekolah. Salah satunya adalah penelitian yang dilakukan oleh khairunnisa dan temannya, pada penelitian itu mereka melakukan analisis faktor penyebab turunnya prestasi belajar mahasiswa pada masa pandemi menggunakan klasifikasi C4.5. Hasilnya, faktor utama dalam menurunnya prestasi mahasiswa adalah pemahaman materi dan algoritma C4.5 dapat diterapkan dalam analisis faktor penyebab turunnya prestasi mahasiswa, karena hasil dari algoritma C4.5 memiliki nilai akurasi yang tinggi, yaitu 97.5% [6]. Dalam hal melakukan klasterisasi, sudah banyak juga penelitian yang dilakukan, contohnya penelitian yang dilakukan oleh Asrul Sani tentang penerapan metode *k-means clustering* pada sebuah perusahaan, dan dalam penelitian tersebut algoritma *k-means clustering* berhasil mengklasterisasi sebuah data transaksi pada perusahaan menjadi 4 kelompok cluster, C1 berjumlah 10 kode artikel, nilai untuk C2 berjumlah 4 kode artikel, nilai untuk C3 berjumlah 15 artikel dan untuk C4 berjumlah 8 artikel [7]. Kemudian penelitian tentang pengaruh Pendidikan orang tua, penghasilan orang tua, dan minat belajar mahasiswa terhadap prestasi mahasiswa jurusan Managemen Fakultas Ekonomi Stambuk 2014 Universitas HKBP NOMMENSEN, dalam penelitian tersebut menghasilkan kesimpulan bahwa tingkat Pendidikan, pendapatan orang tua dan minat mahasiswa berpengaruh positif terhadap prestasi mahasiswa jurusan Managemen Fakultas Ekonomi Stambuk 2014 Universitas HKBP NOMMENSEN. Hal itu bisa dibuktikan dengan hasil perhitungan koefisien determinasi (R2) diperoleh persentase sumbangan variabel independen yaitu tingkat pendidikan orangtua (X1), penghasilan orangtua (X2), dan minat belajar mahasiswa (X3) berpengaruh terhadap variabel dependen yaitu prestasi belajar mahasiswa (Y), yaitu sebesar 68,2% yang berarti tingkat pendidikan orangtua (X1), penghasilan orangtua (X2), dan minat belajar mahasiswa (X3) dapat menjelaskan prestasi belajar mahasiswa (Y) dan sisanya yaitu sebesar 31,8% dijelaskan oleh variabel lain yang tidak dikaji dalam penelitian ini [8].

Ada juga penelitian yang dilakukan Akon, Mashudi, dan juga Thomas. Mereka melakukan penelitian tentang Pengaruh Penghasilan dan Motivasi Orang Tua terhadap Hasil Belajar Siswa, dimana mereka mendapatkan kesimpulan bahwa tidak terdapat pengaruh signifikan antara tingkat penghasilan dan motivasi orang tua secara simultan terhadap hasil belajar siswa di SMP Negeri 1 Simpang Hulu. Hal ini dibuktikan dengan nilai Fhitung < Ftabel (2.070 < 4,977) [9]. Kemudian, penelitian yang dilakukan oleh Raja dan teman-temannya dalam melakukan Analisa pengaruh penggunaan gawai terhadap nilai akhir siswa SMA. Dalam penelitiannya menggunakan klasifikasi *data mining*, dilakukan pengklasifikasian dengan algoritma *decision tree*, didapatkan bahwa

Jurnal Sistem Komputer dan Informatika (JSON)
Volume 4, Nomor 2, Desember 2022
e-ISSN 2685-998X
DOI 10.30865/json.v4i2.5299

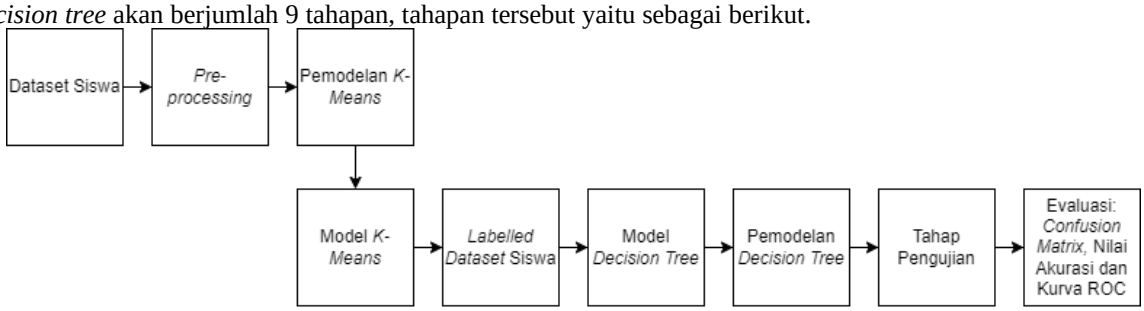
penggunaan gawai tidak berpengaruh terhadap nilai akhir siswa SMA dan performa algoritma *decision tree* tidak cocok diterapkan dalam penelitian ini [10].

Karena sudah banyak penelitian yang dilakukan dan sudah banyak cara yang dapat digunakan dalam analisis pengaruh prestasi siswa, maka *data mining* ini akan digunakan sebagai metode untuk membuktikan apakah jarak antara rumah ke sekolah dan juga penghasilan orang tua seorang siswa/i berpengaruh terhadap prestasi mereka. Pada dasarnya, penelitian untuk menganalisa prestasi siswa memang sudah pernah dilakukan, namun penyelesaian dan metode yang dilakukan berbeda-beda. Jadi, penelitian ini akan menggunakan algoritma *k-means* dan juga *decision tree* untuk menganalisa apakah faktor eksternal berpengaruh terhadap prestasi siswa di sekolah. Dan diharapkan hasil dari penelitian ini akan berguna bagi pihak sekolah untuk mengatasi permasalahan yang ada.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Tahapan penelitian yang akan dilakukan dalam analisis prestasi siswa menggunakan algoritma *k-means* dan *decision tree* akan berjumlah 9 tahapan, tahapan tersebut yaitu sebagai berikut.



Gambar 2. Tahapan Penelitian

Tahapan penelitian ini akan digunakan sebagai acuan untuk melakukan penelitian ini. Dimana penelitian akan dimulai dengan *dataset* siswa yang dilakukan *pre-processing.* Dalam tahap *pre-processing* akan diurai untuk dilihat dan dibersihkan agar *dataset* bisa digunakan dalam penelitian. *Dataset* sendiri didapatkan dari kuesioner yang dilakukan pada SMK Tarbiyatul Ulum dan yang mengisi adalah siswa sekolah tersebut. Setelah data dilakukan proses *pre-processing,* kemudian data akan dilakukan pemodelan menggunakan algoritma *k-means* dengan nilai k = 2 untuk dicari nilai rata-rata tersebut dan hasil dari *cluster* tersebut akan dijadikan *label* pada pemodelan selanjutnya. Setelah *dataset* sudah mendapatkan *label,* maka akan dilakukan pemodelan menggunakan algoritma *decision tree*, yang kemudian dari hasil pemodelan tersebut akan dilakukan analisis apakah penghasilan orang tua dan jarak rumah ke sekolah berpengaruh terhadap prestasi siswa di sekolah. Lalu tahap terakhir adalah dengan melakukan tahap pengujian dan evaluasi untuk mencari tahu apakah algoritma *decision tree* bisa digunakan dalam memprediksi prestasi siswa berdasarkan nilai akurasi dan juga kurva ROC.

2.2 Pemodelan Algoritma *K-Means*

Tahap pemodelan pada algoritma ini berfungsi untuk mencari nilai rata-rata dari data yang sudah ada menggunakan algoritma *k-means.* Algoritma *k-means* sendiri merupakan algoritma yang masuk ke dalam kategori *unsupervised learning,* kemudian sistem pengelompokkannya masuk ke dalam partisi [11]. Algoritma ini berfungsi sangat baik pada data numerik, yang dimana sangat cocok pada data yang ada pada penelitian ini. Karena nantinya yang akan menjadi pemodelan pada data ini adalah atribut Prestasi, RUTS1, RUAS1, RUTS2. Algoritma ini juga lebih baik dari algoritma *k-medoids* dalam melakukan *clustering* [12].Lalu dari hasil *k-means* ini akan digunakan sebagai label untuk tahapan selanjutnya. Secara umum, tahapan dalam *k-means* adalah sebagai berikut:

a) Menentukan k sebagai jumlah klaster yang akan dibentuk.
b) K *centroid* (titik pusat klaster) awal akan dibangkitkan secara random.

$$v = \sum_{k=1}^{n} \frac{Xi}{N} \quad (1)$$

i = 1,2,3 … n

Rumus Menentukan *Centroid* Awal

c) Jarak setiap objek ke masing-masing *centroid* dihitung dari masing-masing klaster. Lalu dalam jarak antara objek dan *centroid* dihitung menggunakan *Euclidian Distance.*

$$d(x,y) = |x - y| = \sqrt{\sum_{i=1}^{n}(xi - yi)^2} \quad (2)$$

Rumus Menghitung Jarak Antar *Centroid*

d) Alokasikan setiap objek pada *centroid* terdekat.
e) Lakukan iterasi, setelah itu dengan menggunakan persamaan tentukan posisi centroid baru.
Jika posisi *centroid* tidak sama, maka ulangi Langkah ke 3

Jurnal Sistem Komputer dan Informatika (JSON)
Volume 4, Nomor 2, Desember 2022
e-ISSN 2685-998X
DOI 10.30865/json.v4i2.5299

2.3 Pemodelan Algoritma *Decision Tree*

Tahap pemodelan ini berguna untuk mencari tahu bagaimana pola atau informasi yang dapat digunakan sebagai informasi yang berguna dari data yang disiapkan. Pada tahap pemodelan ini menggunakan algoritma *decision tree*, dimana algoritma ini akan berguna untuk pengambilan keputusan yang pemodelannya berbentuk pohon keputusan. Model dalam algoritma ini akan dibangun dengan tampilan seperti pohon dan dengan cabang yang terdiri dari data yang dibangun secara rekursif dan dengan kelas yang sama [13]. Dalam melakukan perumusan algoritma *decision tree*, formula yang dipakai adalah sebagai berikut [14]:

$$Entropy(S) = \sum_{i=1}^{n} - p_i \log 2\, p_i \quad (3)$$

Rumus menghitung *entropy*
Penjelasan:
S = Kumpulan Kasus
S = Jumlah Partisi S
S = Proporsi dari

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^{n} \frac{|S_i|}{|S|} Entropy(S_i) \quad (4)$$

Rumus menghitung *Gain*
Penjelasan:
S = Kumpulan Kasus
S = Atribut
S = Jumlah Partisi Atribut A
S_i = Jumlah Kasus Pada Partisi ke-i
S = Jumlah Kasus dalam S

2.4 Evaluasi

Pada tahap evaluasi ini berfungsi untuk melihat atau mengetahui kualitas dari pemodelan metode klasifikasi yang sudah dilakukan sebelumnya. Pada tahap ini sendiri akan diukur atau dilihat nilai akurasi, *confusion matrix*, *classification report*, nilai AUC dan kurva ROC dari prediksi yang sudah dilakukan algoritma *decision tree*. dari hasil nilai tersebut akan menjelaskan bahwa algoritma tersebut efektif untuk digunakan atau tidak [15]. Pada penelitian ini, nilai akurasi dan juga nilai AUC pemodelan algoritma *decision tree* yang sudah dilakukan sebelumnya, dilihat untuk mengetahui nilai tersebut akan masuk kedalam kelompok mana. Adapun menurut Gorunescu dalam [16], *performance* keakurasian AUC dapat diklasifikasikan menjadi lima kelompok yaitu:
a) 0.90 – 1.00 = *Excellent Classification*
b) 0.80 – 0.90 = *Good Classification*
c) 0.70 – 0.80 = *Fair Classification*
d) 0.60 – 0.70 = *Poor Classification*
e) 0.50 – 0.60 = *Failure*

3. HASIL DAN PEMBAHASAN

3.1 Pengumpulan Data

Data yang digunakan dalam penelitian ini diambil dari sekolah SMK Tarbiyatul Ulum. Data tersebut berisi beberapa atribut didalamnya, yaitu Jenis-Kelamin, Prestasi, Jarak Rumah-Sekolah, Penghasilan Orang Tua, RUTS1, RUAS1, RUTS2 dengan total data yang didapat adalah 158 *record*. Metode dalam pengambilan data tersebut adalah dengan menggunakan kuesioner dalam *google form*, dengan pengisi kuesioner tersebut adalah siswa SMK Tarbiyatul Ulum.Gunalan huruf kecil dan abjed untuk penomoran list.

Tabel 1. *Dataset* Penelitian

Jenis-Kelamin	Prestasi	Jarak Rumah-Sekolah	Penghasilan Orang Tua	RUTS1	RUAS1	RUTS2
Lelaki	-	200-600	<1.000.000	80,25	80,75	82,75
Lelaki	-	200-600	1.501.000-2.000.000	85,00	80,00	80,25
Lelaki	-	600-1000	2.001.000-2.500.00	82,50	78,50	88,00
Lelaki	Provinsi	<200	1.000.000-1.500.000	80,25	80,25	80,25
Lelaki	-	200-600	1.000.000-1.500.000	82,50	73,25	88,00
Lelaki	-	200-600	2.001.000-2.500.00	80,25	85,75	80,00
Lelaki	-	600-1000	1.000.000-1.500.000	85,00	82,25	85,50
Lelaki	-	200-600	1.000.000-1.500.000	80,25	80,00	82,25
Lelaki	-	200-600	1.000.000-1.500.000	80,25	78,00	80,00

Jurnal Sistem Komputer dan Informatika (JSON)
Volume 4, Nomor 2, Desember 2022
e-ISSN 2685-998X
DOI 10.30865/json.v4i2.5299

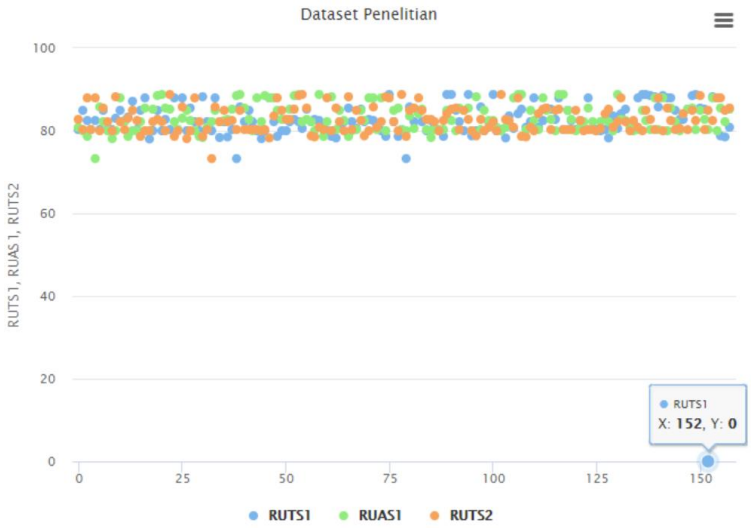
Jenis-Kelamin	Prestasi	Jarak Rumah-Sekolah	Penghasilan Orang Tua	RUTS1	RUAS1	RUTS2
Lelaki	-	200-600	1.501.000-2.000.000	83,00	80,00	88,25
Lelaki	-	200-600	1.501.000-2.000.000	85,00	88,00	82,25
Lelaki	-	600-1000	1.501.000-2.000.000	82,50	82,25	80,25
Lelaki	-	200-600	1.501.000-2.000.000	83,25	78,75	83,00

3.2 *Pre-processing Data*

Sebelum melakukan proses data, dilakukan beberapa tahapan terlebih dahulu seperti melakukan pemahaman data, pembersihan data dan transformasi data agar data yang dimiliki dapat dilakukan proses pada *data mining* [17]. Tahapan tersebut adalah sebagai berikut:

3.2.1 *Data Cleaning*

Pada tahapan ini, data akan diseleksi untuk dilihat apakah ada data yang perlu dihapus. Alasan data dihapus pun ada beberapa macam, yaitu jika data tidak lengkap, tidak akurat, tidak relevan dan lainnya. Tahap ini berdasar penelitian [18], bahwa *data cleaning* berguna untuk menghapus *noise data* yang dapat mengganggu proses penelitian. Salah satu cara untuk melihat data tersebut baik atau tidak adalah dengan menggunakan visualisasi data. Tujuan utama dari visualisasi data adalah untuk mengkomunikasikan informasi secara jelas dan efektif dengan cara grafis menurut Friedman dalam [19]. Tipe visualisasi data yang akan dipakai adalah *scratter/bubble*. Visualisasi data ini digunakan untuk melihat apakah ada data yang jauh berbeda dari kebanyakan data atau tidak. Karena, jika ada data yang berbeda jauh dari kebanyakan data lainnya, hal itu akan mempengaruhi proses pemodelan pada algoritma *k-means.* Visualisasi akan menggunakan *software* bantuan, yaitu *rapidminer*.



Gambar 3. Visualisasi *Dataset* Penelitian

Jika dilihat dari Gambar 2. bisa disimpulkan bahwa ada satu data yang tidak sesuai atau tidak seperti kebanyakan data pada umumnya. Hal ini akan mempengaruhi proses pemodelan pada algoritma *k-means clustering*. Maka dari itu, data tersebut akan dihapus dan tidak akan diolah pada tahap selanjutnya. Dengan itu, total data yang ada sekarang adalah 157 *record.*

3.2.1 *Transformation Data*

Tahap ini berdasar pada penelitian [20] agar data diubah menjadi format yang bisa digunakan dalam penelitian. Hal itu karena, data jarak rumah ke sekolah dan juga penghasilan orang tua masih berupa data dengan perkiraan saja, maka data tersebut akan diubah menjadi penomoran sesuai dengan kelompok yang ada sebelumnya. Lalu, karena tipe data 'string' juga tidak bisa digunakan dalam pemodelan *decision tree*, maka akan diubah data Jenis-Kelamin dan Prestasi menjadi angka, agar pemodelan bisa dilakukan. Hasil dari tranformasi data tersebut adalah sebagai berikut:

Tabel 2. *Dataset* Sesudah Transformasi Data

Jenis-Kelamin	Prestasi	Jarak Rumah-Sekolah	Penghasilan Orang Tua	RUTS1	RUAS1	RUTS2
1	0	2	1	80,25	80,75	82,75
1	0	2	2	85,00	80,00	80,25
1	0	3	3	82,50	78,50	88,00

Jurnal Sistem Komputer dan Informatika (JSON)
Volume 4, Nomor 2, Desember 2022
e-ISSN 2685-998X
DOI 10.30865/json.v4i2.5299

Jenis-Kelamin	Prestasi	Jarak Rumah-Sekolah	Penghasilan Orang Tua	RUTS1	RUAS1	RUTS2
1	1	1	2	80,25	80,25	80,25
1	0	2	2	82,50	73,25	88,00
1	0	2	3	80,25	85,75	80,00
1	0	3	2	85,00	82,25	85,50
1	0	2	2	80,25	80,00	82,25
1	0	2	2	80,25	78,00	80,00
1	0	2	2	83,00	80,00	88,25
1	0	2	2	85,00	88,00	82,25
1	0	3	2	82,50	82,25	80,25
1	0	2	2	83,25	78,75	83,00

Penjelasan Tabel 2. diatas:
Jenis Kelamin
1 Lelaki
2 Perempuan
Prestasi
0 Tidak Ada
1 Provinsi dan Lokal
Jarak Rumah-Sekolah
1 Dekat
2 Lumayan
3 Jauh
Penghasilan Orang Tua
1 Sedikit
2 Lumayan
3 Banyak
4 Sangat Banyak

3.3 Implementasi *K-Means* Menggunakan Bahasa pemrograman Python

Setelah dilakukan pengecekan dan juga *preprocessing data*, dan sudah dihasilkan data yang bisa digunakan dan layak untuk diteliti. Maka penelitian dimulai dengan melakukan pemodelan terhadap dataset siswa, proses dalam penelitian ini menggunakan bahasa pemrograman python. Langkah pertama yang dilakukan adalah dengan mencari label dari data yang ada, yaitu dengan cara melakukan pemodelan *k-means*.

```
df = pd.read_excel("Dataset Penelitian.xlsx")
feature_cols = ['Prestasi','RUTS1','RUAS1','RUTS2']
cluster = df[feature_cols]
cluster.head(10)
```

	Prestasi	RUTS1	RUAS1	RUTS2
0	0	80.25	80.75	82.75
1	0	85.00	80.00	80.25
2	0	82.50	78.50	88.00
3	1	80.25	80.25	80.25
4	0	82.50	73.25	88.00
5	0	80.25	85.75	80.00
6	0	85.00	82.25	85.50
7	0	80.25	80.00	82.25
8	0	80.25	78.00	80.00
9	0	83.00	80.00	88.25

Gambar 4. *Dataset* Untuk Pemodelan *K-Means*

Tetapi data yang digunakan untuk *clustering* hanyalah data Prestasi, RUTS1, RUAS1 dan RUTS2. Jumlah k dibagi menjadi 2, yang nantinya cluster tersebut akan digunakan sebagai label dari dataset siswa. Nilai k hanya dibagi 2 dikarenakan proses *clustering* ini digunakan untuk mencari nilai rata-rata dari dataset yang ada saja. Maka hal tersebut menghasilkan model seperti ini:

Jurnal Sistem Komputer dan Informatika (JSON)
Volume 4, Nomor 2, Desember 2022
e-ISSN 2685-998X
DOI 10.30865/json.v4i2.5299


```
from sklearn.cluster import KMeans
km = KMeans(n_clusters=2)
km.fit(cluster)
y_pred = km.predict(cluster)
y_pred
```

```
array([1, 1, 1, 1, 1, 1, 0, 1, 1, 1, 0, 1, 1, 0, 1, 0, 0, 1, 1, 0, 1, 1,
       1, 0, 1, 0, 1, 0, 1, 1, 0, 1, 1, 0, 1, 1, 1, 1, 1, 0, 1, 0, 1, 1,
       1, 1, 0, 0, 1, 1, 1, 1, 0, 0, 1, 0, 1, 1, 0, 1, 1, 1, 1, 1, 1, 0,
       1, 1, 0, 1, 1, 0, 0, 1, 1, 1, 0, 1, 1, 1, 1, 1, 0, 1, 0, 1, 1, 1, 1, 1,
       1, 0, 0, 0, 1, 1, 0, 1, 1, 0, 1, 1, 0, 1, 1, 1, 1, 1, 1, 0, 0, 1, 1,
       0, 1, 0, 0, 0, 1, 0, 1, 0, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 0,
       0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0,
       1, 1, 1], dtype=int32)
```

Gambar 5. Model Algoritma *K-Means*

```
from sklearn.cluster import KMeans
km = KMeans(n_clusters=2)
km.fit(cluster)
y_pred = km.predict(cluster)
km.cluster_centers_
```

```
array([[8.33333333e-02, 8.62671795e+01, 8.43998333e+01, 8.32726282e+01],
       [3.09278351e-02, 8.06300793e+01, 8.24009358e+01, 8.23194607e+01]])
```

Gambar 6. *Centroid K-Means*

Dari pemodelan diatas bisa disimpulkan bahwa cluster_0 merupakan cluster dengan nilai rata-rata yang lebih tinggi daripada cluster_1. Maka hasil dari *k-means clustering* ini akan dijadikan acuan untuk pelabelan data. Hasil dari *k-means clustering* ini akan diganti nama cluster menjadi penamaan label yang sesuai dengan penelitian kali ini, yaitu cluster_0 akan diganti nama menjadi Tinggi dan cluster_1 akan diganti nama menjadi Rata-Rata.

3.4 Implementasi Algoritma Decision Tree Menggunakan Bahasa pemrograman Python

Setelah dilakukan *clustering* data dengan *k-means* dan hasil model dari *clustering* tersebut digunakan sebagai label dari dataset siswa. Dataset siswa yang sudah dilabelkan lalu dilakukan pemodelan dengan menggunakan salah satu algoritma dalam metode klasifikasi yaitu algoritma *decision tree.* Karena penelitian ini tentang analisis apakah jarak rumah ke sekolah dan penghasilan orang tua berpengaruh terhadap prestasi siswa, maka data yang dipakai untuk penelitian hanyalah jenis-kelamin, jarak rumah dan penghasilan orang tua.

```
df.head(7)
```

	Jenis-Kelamin	Kelas	Prestasi	Jarak Rumah-Sekolah	Penghasilan Orang Tua	RUTS1	RUAS1	RUTS2	label
0	1	X TBSM	0	2	1	80.25	80.75	82.75	1
1	1	X TBSM	0	2	2	85.00	80.00	80.25	1
2	1	X TBSM	0	3	3	82.50	78.50	88.00	1
3	1	X TBSM	1	1	2	80.25	80.25	80.25	1
4	1	X TBSM	0	2	2	82.50	73.25	88.00	1
5	1	X TBSM	0	2	3	80.25	85.75	80.00	1
6	1	X TBSM	0	3	2	85.00	82.25	85.50	0

```
feature_cols = ['Jenis-Kelamin','Jarak Rumah-Sekolah','Penghasilan Orang Tua']
tree = df[feature_cols]
X = tree
Y = df['label']
X.head(7)
```

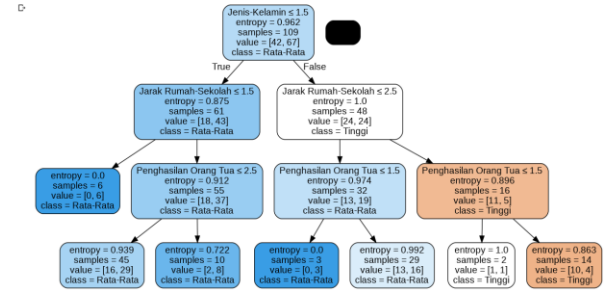
	Jenis-Kelamin	Jarak Rumah-Sekolah	Penghasilan Orang Tua
0	1	2	1
1	1	2	2
2	1	3	3
3	1	1	2
4	1	2	2
5	1	2	3
6	1	3	2

Gambar 7. *Dataset* Untuk Proses *Decision Tree*

Pemodelan dilakukan dengan membagi dataset siswa yang sudah dilabeli menjadi data *training* dan juga data *testing* dengan *ratio* 7:3 dan juga random_state adalah 1, lalu pada model *decision tree* memakai criteria entropy dan juga max_depth 3 (tiga) maka model yang dihasilkan menjadi seperti ini:

```
X_train, X_test, y_train, y_test = train_test_split (X,Y, test_size = 0.3, random_state=1)
clf = DecisionTreeClassifier(criterion='entropy',max_depth=3)
clf.fit(X_train, y_train)
from sklearn.tree import export_graphviz
import pydotplus
from IPython.display import Image

dot_data = StringIO()
export_graphviz(clf, out_file=dot_data,
                filled=True, rounded=True,
                special_characters=True,feature_names = feature_cols,class_names=['Tinggi','Rata-Rata'])
graph = pydotplus.graph_from_dot_data(dot_data.getvalue())
graph.write_png('Siswa.png')
Image(graph.create_png())
```



Gambar 8. Model Algoritma *Decision Tree*

Jurnal Sistem Komputer dan Informatika (JSON)
Volume 4, Nomor 2, Desember 2022
e-ISSN 2685-998X
DOI 10.30865/json.v4i2.5299

Agar memudahkan mendapatkan informasi dari pemodelan *decision tree* yang dibuat, akan dibuat penjelasan atau penjabaran dari model yang sudah dilakukan. Hal itu menggunakan cara export_text yang ada pada library sklearn.tree, maka hasilnya akan menjadi seperti ini:

```
from sklearn.tree import export_text
r = export_text(clf, feature_names=feature_cols)
print(r)
```

```
|--- Jenis-Kelamin <= 1.50
|   |--- Jarak Rumah-Sekolah <= 1.50
|   |   |--- class: 1
|   |--- Jarak Rumah-Sekolah >  1.50
|   |   |--- Penghasilan Orang Tua <= 2.50
|   |   |   |--- class: 1
|   |   |--- Penghasilan Orang Tua >  2.50
|   |   |   |--- class: 1
|--- Jenis-Kelamin >  1.50
|   |--- Jarak Rumah-Sekolah <= 2.50
|   |   |--- Penghasilan Orang Tua <= 1.50
|   |   |   |--- class: 1
|   |   |--- Penghasilan Orang Tua >  1.50
|   |   |   |--- class: 1
|   |--- Jarak Rumah-Sekolah >  2.50
|   |   |--- Penghasilan Orang Tua <= 1.50
|   |   |   |--- class: 0
|   |   |--- Penghasilan Orang Tua >  1.50
|   |   |   |--- class: 0
```

Gambar 9. Penjabaran Model *Decision Tree*

3.5 Evaluasi

Pada tahap ini dilakukan uji kelayakan apakah algoritma *decision tree* bisa digunakan sebagai algoritma untuk memprediksi pada penelitian ini. Untuk mendapatkan jawaban tersebut, maka diperlukan beberapa fungsi dalam klasifikasi. Dimana pada pemodelan *decision tree* ini akan dilakukan prediksi terlebih dahulu, kemudian akan dilihat *confusion matrix*, *classification report*, nilai akurasi, nilai AUC dan juga kurva AUC. Semua proses yang dilakukan menggunakan bahasa pemrograman python. Maka hasil dari pengujian ini adalah sebagai berikut:

```
#Accuracy
from sklearn.metrics import accuracy_score
y_pred = clf.predict(X_test)
print("Accuracy : ",accuracy_score(y_test,y_pred))

#Confusion_matrix
from sklearn.metrics import confusion_matrix
print("Confusion Matrix :\n",confusion_matrix(y_test,y_pred))

#Classification Report
from sklearn.metrics import classification_report
print(classification_report(y_test,y_pred))
```

```
Accuracy :  0.6875
Confusion Matrix :
 [[ 5 13]
 [ 2 28]]
              precision    recall  f1-score   support

           0       0.71      0.28      0.40        18
           1       0.68      0.93      0.79        30

    accuracy                           0.69        48
   macro avg       0.70      0.61      0.59        48
weighted avg       0.69      0.69      0.64        48
```

Gambar 10. Nilai Akurasi, *Confusion Matrix* dan *Classification Report*

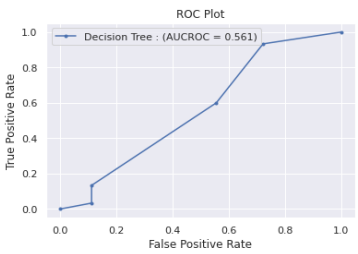
```
#Nilai AUC
tree_probs = clf.predict_proba(X_test)
tree_probs = tree_probs[:,1]
from sklearn.metrics import roc_curve, roc_auc_score
tree_auc = roc_auc_score(y_test, tree_probs)
print('AUCROC : %.3f' % (tree_auc))

tree_fpr, tree_tpr, _ = roc_curve(y_test, tree_probs)

#plot ROC curve
plt.plot(tree_fpr, tree_tpr, marker='.', label='Decision Tree : (AUCROC = %.3f)' % tree_auc)

plt.title('ROC Plot')
plt.xlabel('False Positive Rate')
plt.ylabel('True Positive Rate')
plt.legend()
plt.show()
```

```
AUCROC : 0.561
```



Gambar 11. Nilai AUC dan Kurva ROC

Jurnal Sistem Komputer dan Informatika (JSON)
Volume 4, Nomor 2, Desember 2022
e-ISSN 2685-998X
DOI 10.30865/json.v4i2.5299

3.6 Pembahasan

Dalam analisis prestasi siswa, bisa dilihat dari model algoritma *decision tree* yang sudah dibuat, bahwa penghasilan orang tua dan juga jarak rumah ke sekolah memiliki peranan tersendiri dalam prestasi siswa. Hal itu bisa dilihat jika dalam pemodelan *export_text* yang sudah dilakukan, akar utama dari pengklasifikasian dataset siswa adalah jenis kelamin, penjelasannya yaitu:

a. Jenis kelamin <=1.5 (Lelaki) dan jarak rumah ke sekolah <= 1.5 (Dekat) akan masuk ke dalam kategori 1 (rata-rata), sedangkan jika jarak rumah ke sekolah > 1.5 (Lumayan dan Jauh) maka akan menghasilkan cabang lagi, yaitu penghasilan orang tua <= 2.5 (Lumayan dan Sedikit), tetapi akan masuk ke dalam kategori 1 (rata-rata) juga, begitu juga dengan penghasilan orang tua > 2.5 (Banyak dan Sangat Banyak) akan masuk ke dalam kategori 1 (rata-rata).
b. Jenis kelamin > 1.5 (Perempuan) dan jarak rumah ke sekolah <=2.5 (Lumayan dan dekat) akan menghasilkan cabang penghasilan orang tua <= 1.5 (Sedikit) dan penghasilan orang tua > 1.5 (Lumayan, Banyak dan Sangat Banyak) yang akan masuk ke dalam kategori 1 (rata-rata). Sedangkan jika jarak rumah ke sekolah > 2.5 (Jauh) yang menghasilkan cabang penghasilan orang tua <= 1.5 (Sedikit) dan penghasilan orang tua > 1.5 (Lumayan, Banyak dan Sangat Banyak) akan masuk ke dalam kategori 0 (tinggi).

Maka dengan itu, jarak rumah ke sekolah seorang siswa tidak berpengaruh besar terhadap prestasi mereka, hal itu bisa dibuktikan bahwa dalam pengklasifikasian menggunakan model *decision tree* semua cabang jarak rumah ke sekolah dengan akar utama jenis kelamin lelaki akan masuk ke dalam kategori 1(rata-rata), tetapi hal berbeda terdapat pada akar utama jenis kelamin perempuan dimana ketika jarak rumah ke sekolah yang masuk kedalam kategori dekat dan lumayan akan masuk ke dalam kategori 1 (rata-rata) sedangkan jika jarak rumah ke sekolah yang masuk ke dalam kategori jauh akan masuk ke dalam kategori 0 (tinggi). Lalu, penghasilan orang tua tidak berpengaruh terhadap prestasi seorang siswa, karena baik itu penghasilan orang tua sedikit, lumayan, banyak dan sangat banyak akan menghasilkan kategori prestasi yang sama dengan akar yang sama. Lalu dalam hasil tahap pengujian atau evaluasi, nilai akurasi yang dihasilkan algoritma *decision tree* dalam melakukan prediksi adalah 68% dan dengan nilai AUC 0.561. Dengan itu, algoritma *decision tree* dalam penelitian ini masuk ke dalam kategori *failure* sesuai dengan kelompok performa AUC yang dijelaskan sebelumnya [16]. Maka, algoritma *decision tree* tidak cocok dilakukan untuk dilakukan prediksi dalam penelitian ini. Jadi, Penelitian ini selaras dengan penelitian yang dilakukan oleh Raja dan teman-temannya [10], bahwa penelitian ini sama-sama melakukan analisis terhadap prestasi siswa menggunakan algoritma *decision tree* tetapi dengan atribut yang berbeda. Pada penelitian mereka pun, algoritma *decision tree* tidak cocok untuk dilakukan prediksi. Lalu topik penelitian ini juga sama dengan penelitian yang dilakukan oleh Akon dan teman-temannya, tetapi dengan cara atau metode yang berbeda. Hasil dari penelitiannya pun sama dalam beberapa hal dengan hasil penelitian ini, bahwa penghasilan orang tua tidak berpengaruh penting terhadap prestasi siswa [9].

4. KESIMPULAN

Dari penelitian yang sudah dilakukan dalam menerapkan metode *data mining* dalam analisis prestasi siswa Sekolah Menengah Kejuruan dengan menggunakan metode *clustering* tepatnya dengan menggunakan *K-Means* dan juga metode klasifikasi tepatnya algoritma *decision tree* menggunakan dataset siswa yang berisi tentang jenis kelamin, prestasi, jarak antara rumah ke sekolah, dan juga penghasilan orang tua. Menghasilkan bahwa jarak rumah ke sekolah sedikit berpengaruh terhadap prestasi siswa, tetapi penghasilan orang tua tidak berpengaruh terhadap prestasi siswa. Karena yang menjadi faktor utama pembeda dari klasifikasi yang sudah dilakukan adalah jenis kelamin. Bahwa semua siswa yang ber jenis kelamin lelaki memiliki nilai rata-rata, sedangkan siswa yang berjenis kelamin perempuan ada yang memiliki nilai rata-rata dan juga tinggi. Siswa yang memiliki nilai tinggi mempunyai jarak rumah ke sekolah yang jauh. Maka bisa didapatkan kesimpulan bahwa jika jarak rumah ke sekolah jauh, mungkin meningkatkan motivasi belajar seorang siswa yang berjenis kelamin perempuan. Lalu untuk nilai akurasi dan juga nilai AUC yang dihasilkan algoritma *decision tree* dalam melakukan prediksi adalah 68% dan 0.561, dengan itu bisa disimpulkan bahwa algoritma *decision tree* tidak cocok digunakan untuk memprediksi prestasi siswa. Maka, untuk pihak sekolah SMK Tarbiyatul Ulum disarankan untuk memperhatikan atau memberikan perhatian lebih kepada siswa, terutama yang ber jenis kelamin lelaki. Kemudian memberikan motivasi kepada semua siswa untuk belajar lebih giat lagi dan lebih memperhatikan prestasi mereka, terutama yang mempunyai jarak rumah ke sekolah nya dekat, karena mungkin mereka jadi menganggap mudah dan gampang karena rumahnya dekat dengan sekolah, sehingga motivasi mereka dalam belajar tidak maksimal. Untuk penelitian selanjutnya, disarankan untuk menambahkan faktor dan atribut lainnya dalam melakukan analisis prestasi siswa. Lebih baik juga untuk menambahkan jumlah data pada dataset yang akan dilakukan penelitian agar hasil klasifikasi dan analisis prestasi siswa semakin akurat.

REFERENCES

[1] K. Kartono, Kamus Lengkap Psikologi, Jakarta: P.T. Raja Grafindo Persada, 1996.
[2] M. Z. Rosyid, M. Mansyur dan A. R. Abdullah, Prestasi Belajar, Malang: Literasi Nusantara, 2019.

Jurnal Sistem Komputer dan Informatika (JSON)
Volume 4, Nomor 2, Desember 2022
e-ISSN 2685-998X
DOI 10.30865/json.v4i2.5299

[3] A. Syafi'i, T. Marfiyanto dan S. K. Rodiyah, "Studi Tentang Prestasi Belajar Siswa Dalam Berbagai Aspek dan Faktor Yang Mempengaruhi," *Jurnal Komunikasi Pendidikan,* vol. 2, no. 2, pp. 115-123, 2018.

[4] Elisa, Sulistyarini dan H. Syahrudin, "Rendahnya Hasil Belajar Siswa Pada Mata Pelajaran Ekonomi Di Sekolah Menengah Atas," *Jurnal UNTAN,* vol. 7, no. 5, 2018.

[5] Wagiman, "Hubungan Bimbingan Orang Tua dan Metode Mengajar Guru dengan Prestasi Belajar Siswa di SMP Penda Tawangmangu," *Universitas Muhammadiyah Surakarta,* 2012.

[6] K. F. Irnanda, D. Hartama dan A. P. Windarto, "Analisa Klasifikasi C4.5 Terhadap Faktor Penyebab Menurunnya Prestasi Belajar Mahasiswa Pada Masa Pandemi," *Jurnal Media Informatika Budidarma,* vol. 5, no. 1, pp. 327-331, 2021.

[7] A. Sani, "PENERAPAN METODE K-MEANS CLUSTERING PADA PERUSAHAAN," *Journal Teknologika,* 2018.

[8] A. E. Putriku, "PENGARUH TINGKAT PENDIDIKAN ORANG TUA, PENGHASILAN ORANGTUA, DAN MINAT BELAJAR MAHASISWA TERHADAP PRESTASI MAHASISWA JURUSAN MANAJEMEN FAKULTAS EKONOMI STAMBUK 2014 UNIVERSITAS HKBP NOMMENSEN," *NIAGAWAN,* vol. 7, no. 1, pp. 50-58, 2018.

[9] Akon, Mashudi dan Y. Thomas, "PENGARUH PENGHASILAN DAN MOTIVASI ORANG TUA TERHADAP Hasil Belajar Siswa," *Jurnal Pendidikan dan Pembelajaran Untan,* vol. 4, 2015.

[10] R. A. Simarmata dan Y. T. Tarihoran, "Analisa Pengaruh Penggunaan Gadget Terhadap Nilai Akhir Siswa SMA Secara Umum Menggunakan Metode Data mining (Decision Tree)," *TeIKa,* vol. 11, no. 1, pp. 15-28, 2021.

[11] A. Wanto, M. N. H. Siregar, A. P. Windarto, D. Hartama, N. L. W. S. R. Ginantra, D. Napitupulu, E. S. Negara, M. R. Lubis, S. V. Dewi dan C. Prianto, Data Mining: Algoritma dan Implementasi, Yayasan Kita Menulis, 2020.

[12] C. A. Sugianto, A. H. Rahayu dan A. Gusman, "Algoritma K-Means Untuk Pengelompokkan Penyakit Pasien Pada Puskesmas Cigugur Tengah," *JOINT (Journal of Information Technology),* vol. 2, no. 2, pp. 39-44, 2020.

[13] D. Setiawati, I. Taufik, Jumaidi dan W. B. Z, "KLASIFIKASI TERJEMAHAN AYAT AL-QURAN TENTANG ILMU SAINS MENGGUNAKAN ALGORITMA DECISION TREE BERBASIS MOBILE," *JOIN,* vol. 1, no. 1, pp. 24-27, 2016.

[14] A. Susanto, S. R. Riady, S. D. Ranti dan R. Mandala, "Penerapan Perhitungan Metode Decision Tree Menggunakan Algoritma Iterative Dichotomiser 3 (ID3) Berbasis Website," *JSI (Jurnal Sains Indonesia),* vol. 1, no. 2, pp. 59-68, 2020.

[15] S. T. Rizaldi dan Mustakim, "Perbandingan Teknik Pembagian Data untuk Klasifikasi Sarana Akses Air pada Algoritma K-Nearest Neighbor Sarana Akses Air pada Algoritma K-Nearest Neighbor," *Seminar Nasional Teknologi Informasi, Komunikasi dan Industri (SNTIKI),* vol. 12, pp. 130-137, 2020.

[16] A. Ridwan, "Penerapan Algoritma Naïve Bayes Untuk Klasifikasi Penyakit Diabetes Mellitus," *Jurnal Sistem Komputer dan Kecerdasan Buatan ,* vol. 4, no. 1, pp. 15-21, 2020.

[17] R. Takdirillah, "Penerapan Data Mining Menggunakan Algoritma Apriori Terhadap Data Transaksi Sebagai Pendukung Informasi Strategi Penjualan," *Edumatic: Jurnal Pendidikan Informatika,* vol. 4, no. 1, pp. 37-46, 2020.

[18] H. E. Simanjutak dan Windarto, "Analisa Data Mining Menggunakan Frequent Pattern Growth pada Data Transaksi Penjualan PT Mora Telematika Indonesia untuk Rekomendasi Strategi Pemasaran Produk Internet," *Jurnal Media Informatika Budidarma,* vol. 4, no. 4, pp. 914-923, 2020.

[19] M. G. A. Sutanto, SISTEM VISUALISASI DATA PKH KOTA MALANG BERBASIS GIS MENGGUNAKAN MULTI OBJECTIVE OPTIMIZATION ON THE BASIS OF RATIO ANALYSIS, Malang: Universitas Islam Negeri Maulana Malik Ibrahim, 2018.

[20] J. P. Gultom dan A. Rikki, "Implementasi Data Mining menggunakan Algoritma C-45 pada Data Masyarakat Kecamatan Garoga untuk Menentukan Pola Penerima Beras Raskin," *KAKIFIKOM (Kumpulan Artikel Karya Ilmiah Fakultas Ilmu Komputer),* vol. 2, no. 1, pp. 11-19, 2020.