

**PENILAIAN PENGELOMPOKAN DATA PRESTASI SISWA  
MENGUNAKAN METODE K-MEANS UNTUK MENGENALI SISWA  
BERPRESTASI**

**Jihan Ramadhani\***, Muhammad DinNawar, Nurmala Plorensia Aritonang,  
Mahasiswa Program Studi Sistem Informasi  
Sekolah Tinggi Manajemen Informatika dan Komputer Royal  
*\*email:* [ramadhanijihan753@gmail.com](mailto:ramadhanijihan753@gmail.com)

**Abstract:** Not a few efforts made by the school and parties in the field of education to improve student achievement in the academic field which is useful for achieving national education standards. The level of student success is a reflection of the quality of the world of education. Data mining technology is a data processing technology that can find and identify hidden data patterns. The results of data processing are useful in future decision making. Given the large amount of student data, finding out outstanding students is not an easy thing for teachers and schools to do, as well as at MTs Hidayatul Ulumiyah Ujung Kubu. This is seen in terms of mastery of the subject matter both in theory and practice. Not only that, the value of attitudes and attendance lists are also important to consider. This research uses the K-Means Clustering method. The closer to 1, the better the model. The application of the K-Means Clustering algorithm produces 3 clusters with a silhouette\_score value of 0.41, namely with cluster 0 with the lowest number of Student Scores of 80, cluster 1 with the highest number of Student Scores as many as 49, cluster 2 with the number of Best Student Grades of 82.

**Keywords:** Clustering; Data Mining; K-means; Student; Value

**Abstrak:** Tidak sedikit upaya yang dilakukan pihak sekolah maupun pihak di bidang pendidikan untuk meningkatkan prestasi siswa di bidang akademik yang berguna untuk mencapai standar pendidikan nasional. Hasil pengolahan data tersebut berguna dalam pengambilan keputusan ke depannya. Mengingat banyaknya data siswa, maka untuk mengetahui siswa berprestasi bukanlah hal mudah yang dilakukan oleh guru maupun sekolah, sama halnya di Mts Hidayatul Ulumiyah Ujung Kubu. Hal tersebut dilihat dari segi penguasaan materi pelajaran baik secara teori maupun praktik. Bukan hanya itu, nilai sikap dan daftar hadir juga penting untuk dipertimbangkan. Penelitian ini menggunakan metode *K-Means Clustering*. Semakin mendekati 1, maka modelnya semakin bagus. Penerapan algoritma K-Means Clustering menghasilkan 3 cluster dengan nilai silhouette\_score sebesar 0.41, yaitu dengan cluster 0 dengan jumlah Nilai Siswa Terendah sebanyak 80, cluster 1 dengan jumlah Nilai Siswa Tertinggi sebanyak 49, cluster 2 dengan jumlah Nilai Siswa Terbaik yaitu 82.

**Kata kunci:** *Clustering; Data Mining; K-means; Nilai; Siswa*

## PENDAHULUAN

Dunia pendidikan saat ini dituntut untuk memiliki kemampuan bersaing dengan memanfaatkan semua sumberdaya yang sudah ada. Mulai dari sumber daya sarana, prasarana, dan juga manusia, serta sumber system informasi yang merupakan salah satu sumber daya yang digunakan untuk meningkatkan kemampuan bersaing dengan yang lain[1]. Tingginya tingkat keberhasilan siswa merupakan cerminan daripada kualitas dunia pendidikan. Belakangan ini, dunia pendidikan dituntut untuk memiliki daya saing yang tinggi dengan memanfaatkan kualitas serta kuantitas sumber daya manusia (SDM) yang ada, seperti halnya siswa berprestasi. Secara umum prestasi siswa dapat dinilai dari mata pelajaran yang diambil, baik teori maupun praktik[2].

Tidak sedikit upaya yang dilakukan pihak sekolah maupun pihak di bidang pendidikan untuk meningkatkan prestasi siswa di bidang akademik yang berguna untuk mencapai standar pendidikan nasional. Tingkat keberhasilan siswa merupakan cerminan dari kualitas dunia pendidikan[3]. Teknologi data mining merupakan sebuah teknologi pemrosesan data yang dapat menemukan dan mengidentifikasi pola data yang tersembunyi. Hasil pengolahan data tersebut berguna dalam pengambilan keputusan ke depannya[4].

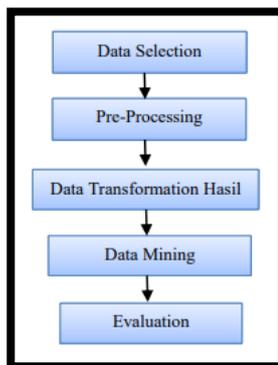
Mengingat banyaknya data siswa, maka untuk mengetahui siswa berprestasi bukanlah hal mudah yang dilakukan oleh guru maupun sekolah, sama halnya diMts Hidayatul Ulumiyah Ujung Kubu. Hal tersebut dilihat dari segi penguasaan materi pelajaran baik secara teori maupun praktik. Bukan hanya itu, nilai sikap dan daftar hadir juga penting untuk dipertimbangkan. Oleh sebab itu, pihak sekolah memerlukan sebuah metode data mining clustering, yaitu K-Means untuk menyelesaikan permasalahan tersebut dengan cepat, tepat dan akurat.

Clustering merupakan sebuah metode yang digunakan untuk mencari dan mengelompokkan data yang memiliki kesamaan karakteristik antara data satu dengan data lainnya. Terdapat beberapa algoritma clustering, salah satunya K-Means. Algoritma K-Means merupakan algoritma pengklasteran yang cukup sederhana, yaitu dengan mempartisi database kedalam beberapa klasteran[5].

Dengan dilakukannya penelitian ini, dapat membantu pihak sekolah dalam menentukan Siswa yang Berprestasi dengan Nilai yang terbaik secara lebih akurat.

## METODE

Metode analisis data yang digunakan pada penelitian ini adalah *Knowledge Discovery Database (KDD)*. *Knowledge Discovery in Database (KDD)* adalah proses yang bertujuan untuk menggali dan menganalisis sejumlah besar himpunan data dan mengekstrak informasi serta pengetahuan yang berguna[6]. Adapun tahapan dari metode *Knowledge Discovery in Database* yang digunakan pada penelitian ini adalah sebagai berikut:



Gambar 1. Alur Penelitian Knowledge Discovery Database (KDD)

**Data Selection**

Pemilihan (seleksi) data dari sekumpulan data operasional perlu dilakukan sebelum tahap penggalian informasi dalam KDD dimulai. Data hasil seleksi yang akan digunakan untuk proses Data mining disimpan dalam suatu berkas, terpisah dari basis data operasional[7]. Data yang diseleksi yaitu data Nilai siswa di MTs Hidayatul Ulumiyah Ujung Kubu. Data yang akan digunakan pada penelitian ini ditunjukkan pada gambar 2:

```

In [2]: #Load Dataset
df=pd.read_excel('ML4_Kelompok7_NilaiUM23.xlsx')
df
    
```

	Kelas	Nama Siswa	AL-QUR'AN HADITS	AQIDAH AKHLAK	Fiqh	SKI	PPKN	Bahasa Indonesia	Bahasa Arab	Matematika	IPA	IPS	Bahasa Inggris	Seni Budaya	Penjaskes	Prakarya	Keterampilan Agama
0	VIII-1	Aff Rizky Nasution	85	88	84	85	84	85	87	82	84	84	78	84	84	85	85
1	VIII-1	Airin Permata Sari	85	84	84	84	84	85	88	79	84	84	84	84	85	85	85
2	VIII-1	Aisa Riata Acha Nasution	84	92	84	85	84	85	87	80	84	84	83	84	84	85	85
3	VIII-1	Akyas Zuhaz	83	79	80	86	85	85	76	56	84	86	80	84	70	85	55
4	VIII-1	Alifna Zahra	87	94	87	84	85	85	75	85	84	84	85	85	85	85	85
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
206	VIII-6	Muhammad Alwi	77	88	87	88	74	87	88	90	85	86	80	76	78	87	76
207	VIII-6	Nada Hilya arisa	75	78	92	88	7	87	85	89	84	87	92	66	78	85	77

Gambar 2. Data Nilai Siswa

**Data Pre-Processing**

Preprocessing data merupakan teknik awal data mining untuk mengubah data mentah atau biasa dikenal dengan raw data yang dikumpulkan dari berbagai sumber menjadi informasi yang lebih bersih dan bisa digunakan untuk pengolahan selanjutnya. Dari dataset dilakukannya proses pembersihan data atau pemilihan data, dimana hanya data-data tertentu yang akan digunakan pada proses data mining[8].

**Data Transformation**

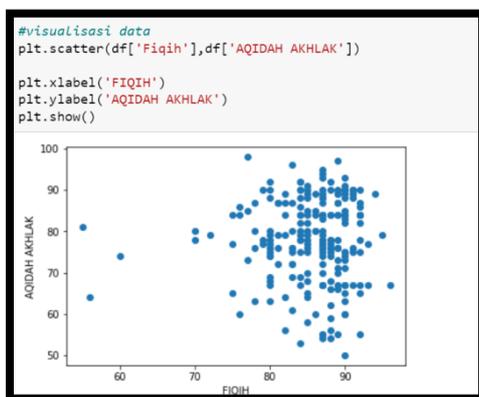
Transformasi data dilakukan untuk mengubah data tujuannya adalah agar data dapat diolah dengan menggunakan metode K-Means Clustering. Pada tahap ini data ditransformasikan dan atribut-atribut yang akan digunakan[9]. Tabel atribut yang digunakan pada dataset jumlah Nilai Siswa dapat dilihat di gambar 3:

```
df.info()
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 211 entries, 0 to 210
Data columns (total 20 columns):
 #   Column                Non-Null Count  Dtype
---  -
 0   Kelas                 211 non-null   object
 1   Nama Siswa           211 non-null   object
 2   AL-QUR'AN HADITS     211 non-null   int64
 3   AQIDAH AKHLAK        211 non-null   int64
 4   Fiqih                211 non-null   int64
 5   SKI                  211 non-null   int64
 6   PPKN                 211 non-null   int64
 7   Bahasa Indonesia     211 non-null   int64
 8   Bahasa Arab          211 non-null   int64
 9   Matematika           211 non-null   int64
10  IPA                  211 non-null   int64
11  IPS                  211 non-null   int64
12  Bahasa Inggris       211 non-null   int64
13  Seni Budaya         211 non-null   int64
14  Penjasokes          211 non-null   int64
15  Prakarya             211 non-null   int64
16  Keterampilan Agama  211 non-null   int64
17  Nahu/Shorof         211 non-null   int64
18  Baca,Tulis Alquran  211 non-null   int64
19  Hasil              211 non-null   float64
dtypes: float64(1), int64(17), object(2)
```

Gambar 3. Atribut Dataset

### Data Mining

Data mining adalah metode pengolahan data untuk mencari pola tersembunyi dalam data tersebut sehingga pola tersebut dapat dipakai sebagai pengetahuan[10]. Salah satu jenis algoritma data mining adalah k-means yang dapat dipakai untuk mengelompokkan data ke dalam beberapa kluster yang lebih homogen. Algoritma yang digunakan pada penelitian ini adalah algoritma *k-means* yang merupakan salah satu algoritma *clustering*[11]. Dataset yang telah diperoleh, selanjutnya diproses datanya. Proses *clustering* pada data ini menggunakan metode *KMeans Clustering*.



Gambar 4. Visualisasi Dataset

### Knowledge Interpretation/Evaluation.

Tahap ini dilakukan dalam rangka mencari pengetahuan, termasuk memeriksa apakah pola atau informasi yang didapatkan bertentangan atau sudah sesuai dengan fakta atau asumsi dari data yang sudah ada sebelumnya.

### HASIL DAN PEMBAHASAN

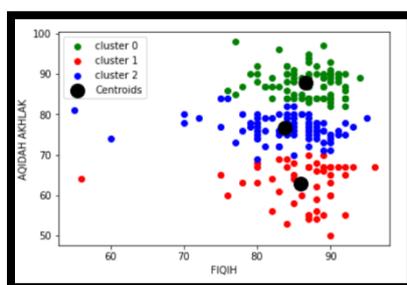
Langkah pertama dalam metode *K-Means Clustering* yaitu menentukan jumlah *cluster* terlebih dahulu, yaitu sebanyak 3 cluster. *Cluster* yang ditentukan yaitu

untuk *cluster* dengan jumla Nilai siswa terendah (C0), *cluster* dengan jumla Nilai Siswa Tertinggi (C1), *cluster* dengan jumla Nilai Terbaik (C2)[12]. Selanjutnya menentukan pusat *cluster* atau *centroid* dengan acak. Data yang memiliki jarak pendek atau terdekat dengan centroid akan membentuk sebuah cluster.

```
#melihat centroid
km.cluster_centers_

array([[86.6375 , 87.9      ],
       [85.95918367, 62.89795918],
       [83.7195122 , 76.67073171]])
```

Gambar 5. Centroid Pada Dataset



Gambar 6. Visualisasi Hasil Clusters Dan Centroid

Selanjutnya mengelompokan data ke dalam cluster, data yang mempunyai nilai yang minim atau kecil akan masuk ke dalam cluster yang tersedia seperti pada gambar 7 di bawah ini.

```
In [16]: #menampilkan cluster Penjualan Tinggi
df_tinggi=df[df["cluster"]=="NILAI TERTINGGI"]
df_tinggi
```

N S	AQIDAH AKHLAK	Fiqih	SKI	PPKN	Bahasa Indonesia	Bahasa Arab	Matematika	IPS	Bahasa Inggris	Seni Budaya	PenjasKes	Prakarya	Keterampilan Agama	Nahu/Shorof	Baca,Tulis Alquran	Hasil	cluster
8	54	87	85	78	65	85	85	87	81	87	85	57	85	86	87	79.294118	NILAI TERTINGGI
5	65	85	84	78	85	84	82	84	89	84	84	84	85	84	85	83.000000	NILAI TERTINGGI
0	67	84	78	99	75	89	80	85	93	83	77	84	45	78	58	78.941176	NILAI TERTINGGI
5	55	89	80	88	55	89	98	80	77	85	80	89	88	93	80	79.178471	NILAI TERTINGGI
9	67	90	88	78	67	90	84	65	70	87	90	90	90	88	88	82.941176	NILAI TERTINGGI
0	65	90	89	86	54	87	89	89	90	90	75	90	90	85	89	84.470588	NILAI TERTINGGI

Gambar 7. Tampilan Nilai Siswa Sesuai Klaster

```
In [17]: #menampilkan cluster Penjualan Rendah
df_rendah=df[df["cluster"]=="NILAI TERENDAH"]
df_rendah
```

N S	AQIDAH AKHLAK	Fiqih	SKI	PPKN	Bahasa Indonesia	Bahasa Arab	Matematika	IPS	Bahasa Inggris	Seni Budaya	PenjasKes	Prakarya	Keterampilan Agama	Nahu/Shorof	Baca,Tulis Alquran	Hasil	cluster
5	88	84	85	84	85	87	82	84	78	84	84	85	85	84	85	84.294118	NILAI TERENDAH
5	84	84	84	84	85	88	79	84	84	84	85	85	85	84	85	84.294118	NILAI TERENDAH
4	92	84	85	84	85	87	80	84	83	84	84	85	85	84	85	84.647059	NILAI TERENDAH
7	94	87	84	85	85	75	85	84	85	85	85	85	85	86	87	85.178471	NILAI TERENDAH
7	87	87	85	89	85	89	85	84	90	87	85	85	85	86	87	86.411765	NILAI TERENDAH
5	90	84	85	82	84	56	83	88	89	84	55	84	84	83	85	79.058824	NILAI TERENDAH

Gambar 8. Tampilan Nilai Siswa Sesuai Klaster

```
In [18]: #menampilkan cluster Penjualan Rendah
df_terbaik=df[df["cluster"]=="NILAI TERBAIK"]
df_terbaik
```

AL- JAN NTS	AQIDAH AKHLAK	Fiqih	SKI	PPKN	Bahasa Indonesia	Bahasa Arab	Matematika	IPS	Bahasa Inggris	Seni Budaya	Penjaskes	Prakarya	Keterampilan Agama	Nahu/Shorof	Baca,Tulis Alquran	Hasil	cluster
83	79	80	85	85	85	76	56	86	80	84	70	85	55	82	80	78.588235	NILAI TERBAIK
76	82	85	86	79	55	84	55	84	85	85	67	86	85	85	85	79.294118	NILAI TERBAIK
88	80	87	84	94	85	90	86	84	90	84	85	85	85	86	88	86.176471	NILAI TERBAIK
85	78	84	84	87	83	92	81	83	81	84	84	85	84	83	85	83.764706	NILAI TERBAIK
86	79	85	84	89	85	82	85	85	80	84	85	85	85	86	86	84.411765	NILAI TERBAIK
87	77	87	84	88	85	80	85	84	96	85	86	85	85	86	87	85.352941	NILAI TERBAIK

Gambar 9. Tampilan Nilai Siswa Sesuai Klaster

```
#Evaluasi Model silhouette_score (Nilainya Antara -1 hingga 1)
#Semakin mendekati 1, maka modelnya semakin bagus
from sklearn.metrics import silhouette_samples, silhouette_score

score = silhouette_score (df[['Fiqih','AQIDAH AKHLAK']], km.labels_)

print('silhouette_score: %.2f' % score)

silhouette_score: 0.41
```

Gambar 10. Evaluasi Hasil Cluster

Pada gambar 10 hasil evaluasi parameter dalam mengukur kinerja atau akurasi algoritma *K-means* dilakukan dengan menghitung *Model silhouette\_score* dengan hasil score yaitu 0.41[13]. Pemilihan metode *silhouette* digunakan untuk melihat kualitas cluster dimana seberapa baik suatu objek ditempatkan dalam suatu cluster. Evaluasi Model *silhouette\_score* (Nilainya Antara -1 hingga 1). Semakin mendekati 1, maka modelnya semakin bagus[14].

**SIMPULAN**

Penerapan algoritma *K-Means Clustering* menghasilkan 3 cluster dengan nilai *silhouette\_score* sebesar 0.41, yaitu dengan cluster 0 dengan jumlah Nilai Siswa Terendah sebanyak 80, cluster 1 dengan jumlah Nilai Siswa Tertinggi sebanyak 49, cluster 2 dengan jumlah Nilai Siswa Terbaik yaitu 82.

Dalam hasil tersebut masih ada beberapa jumlah nilai siswa yang jumlahnya nilai nya masih rendah. Dengan demikian hal tersebut dapat dijadikan sebagai acuan bagi sekolah untuk mengevaluasi dalam menentukan jumlah nilai siswa yang sesuai dengan yang di ajukan dalam pemilihan siswa berprestasi. Untuk penelitian lebih lanjut dapat melakukan pengoptimalan dengan teknik *preprocessing* data untuk menghasilkan hasil yang berkualitas dan menggunakan algoritma yang lain agar mendapatkan perbandingan. Ada juga metode lain seperti algoritma *X-means* bisa digunakan sehingga didapatkan perbandingan dan hasil keputusan yang baik dan akurat[15].

## UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada MTs Hidayatul Ulumiyah Ujung Kubu atas bantuan dan kerja sama yang luar biasa selama penelitian ini. Tanpa bantuan dan partisipasi mereka, penelitian ini tidak akan berhasil. Kami berharap kerja sama ini dapat berlanjut di masa depan untuk mengembangkan solusi yang lebih baik untuk menganalisis siswa berprestasi di masa depan karena penelitian sebelumnya telah memberikan wawasan yang berharga dan menjadi pijakan bagi kami untuk merancang solusi yang lebih baik. Saya ingin mengucapkan terima kasih atas dukungan Anda, kerja sama, dan kesempatan ini.

## DAFTAR PUSTAKA

- [1] Y. Prayoga, A. Mahmudi, and H. Zulfia Zahro', "Penerapan Metode K Means Pada Sistem Informasi Akademik Untuk Pengelompokkan Siswa Berprestasi Di Upt Sma Negeri 3 Kota Pasuruan Berbasis Web," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 5, no. 2, pp. 822–828, 2021, doi: 10.36040/jati.v5i2.3740.
- [2] N. Ismaya *et al.*, "Penentuan Siswa Berprestasi Menggunakan Metode K-Means Clustering Di Smp Takhassus Al Qur ' an," *J. Tek. Inform.*, vol. 1, no. 1, pp. 64–68, 2022.
- [3] D. O. Dacwanda and Y. Nataliani, "Implementasi k-Means Clustering untuk Analisis Nilai Akademik Siswa Berdasarkan Nilai Pengetahuan dan Keterampilan," *Aiti*, vol. 18, no. 2, pp. 125–138, 2021, doi: 10.24246/aiti.v18i2.125-138.
- [4] E. A. Saputra and Y. Nataliani, "Analisis Pengelompokan Data Nilai Siswa untuk Menentukan Siswa Berprestasi Menggunakan Metode Clustering K-Means," *J. Inf. Syst. Informatics*, vol. 3, no. 3, pp. 424–439, 2021, doi: 10.51519/journalisi.v3i3.164.
- [5] . Novi and A. Mubarak, "Penerapan Algoritma K-Means untuk Menentukan Kelas Unggulan di SMP Pelita Bandung," *Infomatek*, vol. 32, no. 2, pp. 97–106, 2021, doi: 10.23969/infomatek.v23i2.4351.
- [6] S. Oktarian, S. Defit, and Sumijan, "Clustering Students' Interest Determination in School Selection Using the K-Means Clustering Algorithm Method," *J. Inf. dan Teknol.*, vol. 2, pp. 68–75, 2020, doi: 10.37034/jidt.v2i3.65.
- [7] S. D. Agosto, "... Data Nilai Siswa Kelas 8 Berbasis Nilai Pengetahuan Untuk Menentukan Siswa Berprestasi Menggunakan Metode K-Means Clustering (Studi Kasus: SMP Negeri 4 ...," vol. 7, no. September, pp. 630–638, 2023, [Online]. Available: [https://repository.uksw.edu/handle/123456789/30778%0Ahttps://repository.uksw.edu/bitstream/123456789/30778/7/T1\\_672018414\\_Isi.pdf](https://repository.uksw.edu/handle/123456789/30778%0Ahttps://repository.uksw.edu/bitstream/123456789/30778/7/T1_672018414_Isi.pdf)
- [8] F. P. Dewi, P. S. Aryni, and Y. Umaidah, "Implementasi Algoritma K-Means Clustering Seleksi Siswa Berprestasi Berdasarkan Keaktifan dalam Proses Pembelajaran," *JISKA (Jurnal Inform. Sunan Kalijaga)*, vol. 7, no. 2, pp. 111–121, 2022, doi: 10.14421/jiska.2022.7.2.111-121.
- [9] M. A. Fitria, "Bab Ii Landasan Teori," *J. Chem. Inf. Model.*, vol. 53, no. 9, pp. 1–15, 2021.
- [10] N. Butarbutar, A. P. Windarto, D. Hartama, and S. Solikhun, "Komparasi Kinerja

- Algoritma Fuzzy C-Means Dan K-Means Dalam Pengelompokan Data Siswa Berdasarkan Prestasi Nilai Akademik Siswa,” *Jurasik (Jurnal Ris. Sist. Inf. dan Tek. Inform.*, vol. 1, no. 1, p. 46, 2017, doi: 10.30645/jurasik.v1i1.8.
- [11] T. Widyanti, S. S. Hilabi, A. Hananto, Tukino, and E. Novalia, “Implementasi K-Means dan K-Nearest Neighbors pada Kategori Siswa Berprestasi,” *J. Inf. dan Teknol.*, vol. 5, no. 1, pp. 75–82, 2023, doi: 10.37034/jidt.v5i1.255.
- [12] N. Putu, E. Merliana, and A. J. Santoso, “Analisa Penentuan Jumlah Cluster Terbaik pada Metode K-Means,” pp. 978–979.
- [13] H. Yuwafi, F. Marisa, and I. D. Wijaya, “Implementasi Data Mining Untuk Menentukan Santri Berprestasi Di Pp . Manaarulhuda Dengan Metode Clustering Algoritma K-Means,” *J. SPIRIT*, vol. 11, no. 1, pp. 22–29, 2019.
- [14] N. Anisah, “Seminar Nasional Informatika (SENATIKA) Prosiding SENATIKA 2021 Metode Clustering K-Means dalam Memprediksi Prestasi Siswa,” pp. 2003–2006, 2021.
- [15] A. A. I. I. P. Nengah Widya Utami, “Penerapan Data Mining Untuk Mengetahui Pola Pemilihan Program Studi Di Stmik Primakara Menggunakan Algoritma K-Means ...,” *J. Teknol. Inf. dan ...*, vol. 3, pp. 456–463, 2021, [Online]. Available: <http://jurnal.undhirabali.ac.id/index.php/jutik/article/view/1540>