E-ISSN: 3031-8629
P-ISSN: 3032-3258
Halaman 1-13
Volume 2 Nomor 2 Tahun 2025

Clustering Penyakit Demam Berdarah Menggunakan Algoritma K-Means Di Kabupaten Landak

Yulianus Anggi[1], Putri Yuli Utami[2], Istikoma[3]
[1,] Program Studi Sistem Informasi, Universitas Muhammadiyah Pontianak
e-mail: yulianusanggi00@gmail.com putriyuli@unmuhpnk.ac.id
istikoma@unmuhpnk.ac.id

Abstrak

Demam berdarah *dengue* (DBD) merupakan penyakit menular melalui nyamuk *aedes aegypti* yang terjadi di daerah tropis dan sub tropis. Penyebaran penyakit DBD di kabupaten Landak sulit di deteksi karena belum adanya informasi daerah dengan kasus DBD terendah dan tertinggi. Tahapan penelitian dimulai dari *business understanding, data understanding, data preparation, modelling, evaluation*, dan *deployment*. Tujuan penelitian ini mengelompokkan daerah yang memiliki penyebaran penyakit DBD terendah dan tertinggi, menerapkan Algoritma K-*Means*. Hasil penelitian ini mendapatkan nilai k yang optimal dengan 3 *cluster* (*Cluster* 0, *Cluster* 1 dan *Cluster* 2). *Cluster* 0 adalah daerah yang memiliki kasus DBD rendah, *Cluster* 1 adalah daerah yang memiliki kasus DBD sedang dan *Cluster* 2 adalah yang memiliki kasus DBD paling tinggi. Evaluasi yang dilakukan menggunakan *silhouette coefficient* menunjukan hasil yaitu *shilouette coefficient Cluster* 0 yaitu 0.638 masuk pada kategori *Medium structure*, *shilouette coefficient cluster* 1 yaitu 0.491 masuk pada kategori *Weak structure* dan *shilouette coefficient Cluster* 2 yaitu 0.423 masuk pada kategori *Weak structure.*
Kata kunci: *Aedes Aegypti, Clustering, DBD, K-Means, silhouette coefficient.*

Abstract

Dengue fever is a disease transmitted through the aedes aegypti mosquito that occurs in tropical and subtropical areas. The spread of dengue fever in Landak district is difficult to detect because there is no information on areas with the lowest and highest dengue fever cases. The research stages start from business understanding, data understanding, data preparation, modeling, evaluation, and deployment. The purpose of this study is to

group areas with the lowest and highest spread of dengue fever, applying the K-Means Algorithm. The results of this study obtained an optimal k value with 3 clusters (Cluster 0, Cluster 1 and Cluster 2). Cluster 0 is an area with low dengue fever cases, Cluster 1 is an area with moderate dengue fever cases and Cluster 2 is the one with the highest dengue fever cases. The evaluation carried out using the silhouette coefficient showed the results, namely the shilouette coefficient Cluster 0, which is 0.638, is included in the Medium structure category, the shilouette coefficient Cluster 1, which is 0.491, is included in the Weak structure category and the shilouette coefficient Cluster 2, which is 0.423, is included in the Weak structure category.

Keywords: *Aedes Aegypti, Clustering, DBD, K-Means, silhouette coefficient.*

LATAR BELAKANG

Demam berdarah *dengue* (DBD) adalah penyakit yang disebabkan oleh virus *dengue* yang dibawa oleh nyamuk *Aedes Aegypti* betina lewat air liur gigitan nyamuk saat menghisap darah manusia yang menyebabkan gangguan pada sistem pembekuan darah, sehingga mengakibatkan pendarahan (M. A. Sembiring et al., 2021). Penyebar penyakit DBD yaitu *Aedes Aegypti* menyukai lingkungan yang kurang sinar matahari, lembap, gelap dan dekat dengan sungai (Fatmawati & Windarto, 2018).

Demam berdarah *dengue* merupakan salah satu permasalahan di negara tropis, khususnya di negara Indonesia baik di perkotaan ataupun pedesaan. Berdasarkan data dari kementrian kesehatan terdapat 35.694 kasus pada tahun 2023, Dinas Kesehatan Provinsi Kalimantan Barat mencatat sejak Januari hingga November tahun 2023 terdapat 5.845 kasus dengan 67 kasus kematian akibat DBD. Kasus kematian paling banyak lebih kepada anak-anak di Kabupaten Landak ada 498 kasus demam berdarah dengan 9 kasus kematian. Dari 13 Kecamatan yang ada di Kabupaten Landak terdapat 2 Kecamatan yang paling banyak kasus penyebaran DBD yaitu, Kecamatan Ngabang dan Kecamatan Sengah Temila yang masing-masing berjumlah 141 dan 60 kasus.

Penyebaran penyakit DBD di kabupaten Landak sulit di deteksi karena belum adanya informasi untuk daerah dengan kasus DBD terendah dan tertinggi. Untuk mengenali prioritas wilayah penyakit DBD terendah dan tertinggi di kabupaten Landak dapat menggunakan metode *data mining*. *K-Means* adalah salah satu metode *clustering* dalam *data mining*. *K-Means* adalah metode data yang mempartisi data ke dalam bentuk satu atau lebih *cluster* sehingga data yang mempunyai karakteristik yang sama dikelompokkan ke dalam satu *cluster* yang sama yang memiliki karakteristik yang berbeda dikelompokkan ke dalam kelompok yang lain (Metisen & Sari, 2015). Algoritma K-*Means* dimulai dengan pemilihan secara acak K, K disini merupakan banyaknya *cluster* yang ingin dibentuk. Lalu tetapkan nilai K secara acak, sementara nilai itu menjadi pusat dari *cluster* atau biasa disebut dengan *centroid*, *mean* atau "*Means*". Hitung jarak tiap data yang ada, masing – masing *centroid* menggunakan rumus *Euclidean* hingga

ditemukan jarak yang paling dekat dari setiap data dengan *centroid*. Klasifikasikan berdasarkan kedekatannya dengan *centroid* (*distance space*) (Aras et al., 2016

Berdasarkan permasalahaan di atas, dilakukan sebuah penelitian yang berjudul *Clustering* Penyakit Demam Berdarah Di Kabupaten Landak Menggunakan *Algoritma K-Means.* Algoritma K-*Means Clustering* adalah algoritma yang layak karena merupakan salah satu metode *data clustering* nonhirarki yang mengelompokan data dalam bentuk satu atau beberapa *cluster*. Penelitian ini bertujuan untuk membuat klasifikasi penyakit BDB berdasarkan daerah yang mengalami penyebaran penyakit DBD terendah dan tertinggi di Kabupaten Landak. Hasil penelitian ini mengelompokan daerah penyebaran DBD berdasarkan *cluster* terendah dan tertinggi. Metode ini adalah sebuah metode pengelompokkan yang direkomendasikan dalam pentuan daerah yang sering terjangkit DBD.

METODE

1. *Business understanding*

Tahapan awal dari CRISP-DM adalah mendefinisikan masalah. Tahapan ini dijelaskan bahwa masalah yang dihadapi saat ini adalah belum adanya pemetaan daerah yang tepat dalam pemberantasan penyakit demam berdarah.

2. *Data understanding*

Tahapan kedua di metode CRISP-DM adalah memahami data yang diperlukan dan yang akan dianalisis. Dalam hal ini data didapatkan dari Dinas Kesehatan Kabupaten Landak dengan jumlah data 807 dan proses analisisnya menggunakan *software RapidMiner*.

3. Data preparation

Sebelum data siap untuk dibuat modelnya, maka harus melakukan seleksi data terlebih dahulu, Pada tahap ini peneliti melakukan seleksi data dengan menentukan *database* yang akan digunakan yang sesuai atau berkaitan dalam proses *clustering*. Dimana data yang digunakan adalah data penyakit DBD yang ada di Kabupaten Landak. Seleksi data bertujuan untuk menyeleksi data karena tidak semuanya yang akan digunakan, oleh karena itu hanya data yang sesuai untuk dianalisa yang akan

diambil dari *database*.

4. Modeling

Pada tahap *modelling* penulis melakukan tahapan algoritma K-*Means clustering* sebagai berikut :

a. Tentukan jumlah *cluster* (K) yang terdapat pada dataset
b. Tentukan titik pusat (*centroid*) secara acak pada langkah awal
c. Hitung setiap jarak terdekat data dengan *centroid*. Dengan rumus dibawah ini:

$$De = \sqrt{(x_i - y_i)^2 + \ldots + (x_n - y_n)^2}$$

Keterangan:

De = *Euclidean Distance*

(x) = Koordinat Objek

(y) = Koordinat *Centroid*

d. Hitung lagi titik *cluster* dengan anggota *cluster* yang saat ini. Titik *cluster* merupakan rata-rata dari seluruh data dalam sebuah *cluster*. Dapat dihitung menggunakan rumus:

Keterangan:

$$c_{kj} = \frac{\sum_{i=1}^{p} x_{ij}}{p}$$

X_{ij} = *claster* ke-n

P = banyaknya anggota klaster ke-n

Hitung lagi tiap objek dengan titik *Cluster* (*Centroid* baru). Bila perhitungan kelompok tidak terjadi perubahan lagi, maka perhitungan *Clustering* dinyatakan

selesai. Namun bila perhitungan kelompok masih terjadi perubahan, maka perhitungan dilakukan lagi seperti langkah ke-c sampai anggota *Cluster* tidak terjadi perpindahan lagi (Gunawan & Purwayoga, 2022).

5. Evaluation.

Proses evaluasi dari model *clustering* dapat dilakukan dengan menilai skor *silhouette coefficient*. Pada saat menggunakan widget K-*Means* dapat melihat skor *silhouette* untuk *Clustering* yang dihasilkan.

Tabel 1 Nilai *silhouette coefficient*

Nilai *Silhouette Coefficient*	Keterangan
0,71-1,00	*Strong strucuture*
0,51-0,70	*Medium strucuture*
0,25-0,50	*Weak strucuture*
≤ 0,25	*No strucuture*

6. Deployement.

Proses *deployment* yang paling sederhana adalah dengan menuliskan laporan *(report)* dari keseluruhan proses modeling dari awal hingga akhir. Dalam hal ini *deployment* dituangkan dalam bentuk laporan penelitian.

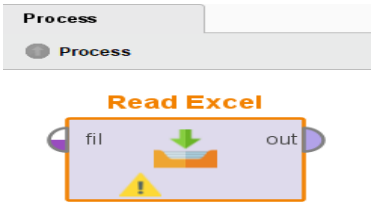
HASIL DAN PEMBAHASAN

Hasil penelitian yang telah dilakukan menggunakan algoritma K-*Means* untuk mengelompokan jumlah daerah mana saja yang memiliki penyebaran penyakit DBD terendah dan tertinggi di kabupaten Landak menggunakan *software RapidMiner*. Berikut adalah proses pengerjaan dalam melakukan *clustering* penyebaran penyakit DBD di Kabupaten Landak menggunakan algoritma K-*Means*. Sampel data yang digunakan ini adalah data yang bersumber dari dinas kesehatan Kabupaten Landak dari tahun 2020-2023 di 13 kecamatan dengan jumlah data 807.

1. Pengolahan Data

Instrumen Pengolahan adalah sebuah alat yang digunakan untuk mengumpulkan data atau mengukur objek dari suatu variabel penelitian. Untuk mendapatkan data yang benar dan sesuai dengan keadaan sebenarnya, maka dibutuhkan suatu instrumen yang valid dan konsisten dalam memberikan data hasil penelitian *reliabel* (Yusup, 2018). Pada penelitian ini akan digunakan beberapa penelitian, antara lain perangkat lunak *software* yang digunakan dalam pengolahan data ini adalah *software RapidMiner* dan data yang digunakan dalam penelitian ini menggunakan data penderita penyakit DBD di Kabupaten Landak. Pada tahap ini merupakan tahap untuk memastikan data penyakit DBD yang dipilih telah layak untuk dilakukan proses pengolahan data.

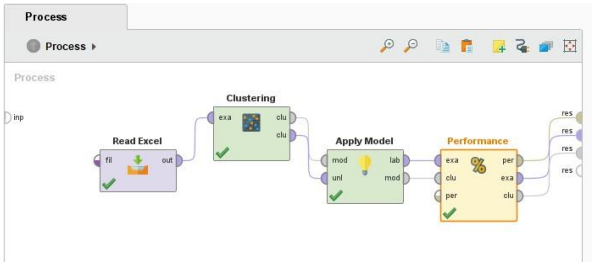
Pada tahap permodelan ini dilakukan dengan menggunakan software rapidminer 10.0 dengan langkah sebagai berikut. Buka *software RapidMiner* lalu pilih *menu file* kemudian *new proses* dan masuk kedalam panel operator seperti yang terlihat pada gambar 1 kemudian ketik *read excel*, lalu *double klik* atau *drag* pada *read excel* agar operator tersebt muncul di panel proses dan bisa digunakan untuk memasukan data *excel* yang telah melalui tahap *data cleansing* dan *data transformation.*



Gambar 1 Read Excel

Dari *dataset* yang ada kemudian dilakukan *preprocessing* data, salah satunya yaitu mengubah nama atribut Kecamatan menjadi kode unik (id) dengan tipe data adalah *polynominal*. Untuk atribut puskesmas tipe data adalah *polynominal* dan untuk atribut bulan adalah i*nteger*. Hasi preposesing data dengan menggunakan *RapidMiner*, tidak

terdapat *missing value* untuk ketiga atribut. Selanjutnya membentuk permodelan *Clustering* dengan operator K-*means,* Pada tahap ini dilakukan perancangan dan implementasi data. Data yang telah ditransformasikan selanjutnya dibuat rancangan dataset menggunakan *RapidMiner*. Rancangan dataset ini menggunakan algoritma K-*Means*. Tahapan yang dilakukan pada *Rapidminer* pertama *read excel* atau input data proses K-*Means*, selanjutnya menghubungkan port-port dari operator *clustering, apply model* dan *performance*.



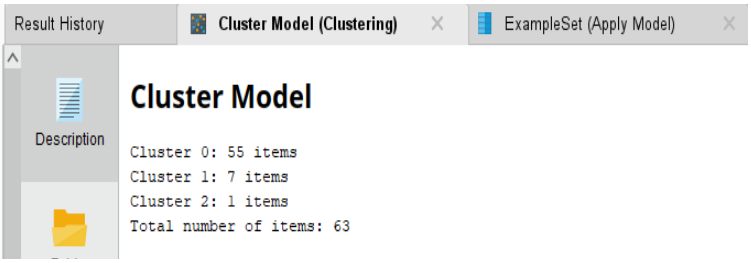
Gambar 2 **Pemodelan Clustering dengan Operator K-Means**

Berdasarkan hasil visualisasi *data excampel set* di bawah menunjukan bahwa *cluster* 0 terdiri atas Kecamatan Menjalin, Air Besar, Mandor, Meranti, Kuala Behe, Sebangki, Jelimpo, Sompak, Dan Banyuke Hulu masuk kedalah jumlah kasus DBD yang rendah, sedangkan *cluster* 1 yang terdiri atas kecamatan Mempawah Hulu, Menyuke, dan Sengah Temila masuk kategori jumlah kasus DBD sedang dan *cluster* 2 adalah Kecamatan Ngabang yang memiliki jumlah kasus DBD tertinggi.

Row No.	KECAMATAN	cluster	2020.0	2021.0	2022.0	2023.0
1	Menjalin	cluster_0	0	0	7	31
2	Mempawah ...	cluster_1	0	1	3	51
3	menyuke	cluster_1	5	8	15	84
4	Air Besar	cluster_0	3	7	4	15
5	Ngabang	cluster_2	21	98	55	141
6	Sengah Temila	cluster_1	18	4	18	60
7	Mandor	cluster_0	0	5	6	14
8	Meranti	cluster_0	0	0	4	19
9	Kuala Behe	cluster_0	0	3	1	13
10	Sebangki	cluster_0	0	0	1	26
11	Jelimpo	cluster_0	0	10	4	18
12	Sompak	cluster_0	1	0	0	3
13	Banyuke Hulu	cluster_0	6	1	0	23

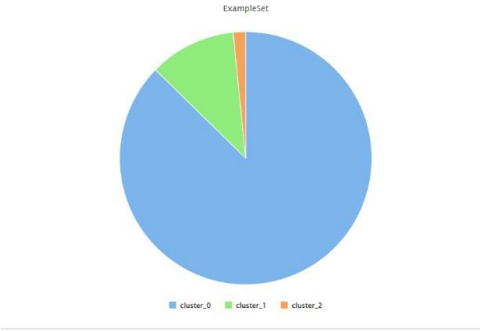
Gambar 3 **Data Excample Set**

Pada Gambar 4 dibawah hasil data terurai menjadi 3 *cluster* yaitu *cluster* 0, *cluster* 1, *cluster* 2 dengan setiap *cluster* mengantongi hasil pengelompokkan data *cluster*_0 terdiri 55 *items*, *cluster*_1 terdiri dari 7 *items*, *cluster*_2 terdiri dari 1 *items*.



Gambar 4 Cluster Model

Pada tampilan *pie* di bawah warna biru merupakan merupakan *cluster* 0 yang terdiri dari 55 i*tems*, warna hijau adalah *cluster* 1 yang terdiri dari 7 *items*, dan yang berwarna jingga adalah *cluster* 2 yang terdiri hanya 1 *items.*



Gambar 5 **Tampilan Pie Pada Hasil Cluster**

2. Menentukan jumlah *Cluster* paling optimal

Untuk menentukan jumlah *cluster* paling optimal penulis menggunakan program bantu *Rapidminer Studio* untuk mendapatkan hasil yang lebih cepat dan akurat. Pengujian dilakukan menggunakan beberapa *cluster* yang berbeda untuk mendapatkan jumlah *cluster* paling optimal untuk karakteristik data kasus DBD yang akan di implementasikan pada proses selanjutnya.

Data yang diuji pada penelitian ini sebanyak 807 data. Dimana data tersebut merupakan data kasus DBD periode tahun 2020-2023 yang terjadi di Kabupaten Landak. Dari hasil pengujian performa *cluster* menggunakan *Davies Bouldin Index* akan dibandingkan dan dicari nilai paling kecil atau nilai yang mendekati 0, Dimana semakin kecil nilai *DBI* maka semakin optimal *Cluster* yang dibutuhkan.

Cluster	Nilai DB
2	0.484
3	0.319
4	0.431
5	0.555

Dari beberapa kali pengujian menggunakan 4 *Cluster* yang berbeda

menunjukkan bahwa nilai *DBI* terkecil terdapat pada pengujian menggunakan 3 jumlah *Cluster* yaitu 0.319 yang berarti *Clustering* penyakit demam berdarah di Kabupaten Landak menggunakan Algoritma K-*Means* ini paling optimal menggunakan 3 *cluster.*

KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan menggunakan algoritma K-*Means Clustering* diperoleh kesimpulan sebagai berikut :

1. Hasil *clustering* menunjukan bahwa *cluster* 0 terdiri atas Kecamatan Menjalin, Air Besar, Mandor, Meranti, Kuala Behe, Sebangki, Jelimpo, Sompak, Dan Banyuke Hulu masuk kedalah jumlah kasus DBD yang rendah, sedangkan *cluster* 1 yang terdiri atas kecamatan Mempawah Hulu, Menyuke, dan Sengah Temila masuk kategori jumlah kasus DBD sedang dan *cluster* 2 adalah Kecamatan Ngabang yang memiliki jumlah kasus DBD tertinggi.
2. Dari beberapa kali pengujian menggunakan 4 *Cluster* yang berbeda menunjukkan bahwa nilai DBI terkecil terdapat pada pengujian menggunakan 3 jumlah *Cluster* yaitu 0.034 yang berarti *Clustering* penyakit demam berdarah di Kabupaten Landak menggunakan Algoritma K-*Means* ini paling optimal menggunakan 3 *cluster.*
3. Dari hasil proses perhitungan *shilouette coefficient* terhadap data penyakit DBD di Kabupaten Landak maka hasil *shilouette coefficient Cluster* 0 yaitu 0.638 termasuk pada kategori *Medium structure*. Hasil *shilouette coefficient cluster* 1 yaitu 0.491 termasuk kategori *Weak structure* dan hasil *shilouette coefficient Cluster* 2 yaitu 0.423 termasuk kategori *Weak structure.*

DAFTAR PUSTAKA

Amelia, D., Padilah, T. N., & Jamaludin, A. (2022). Optimasi Algoritma K-*Means*

Menggunakan Metode Elbow dalam Pengelompokan Penyakit Demam Berdarah Dengue (DBD) di Jawa Barat. *Jurnal Ilmiah Wahana Pendidikan*, *8*(11), 207–215. https://doi.org/10.5281/zenodo.6831380

Aras, Z. Z., Studi Magister Sistem Informasi, P., Dinamika Bangsa, S., & Jl Jend Sudirman Thehok -jambi, J. (2016). ANALISIS DATA MINING UNTUK MENENTUKAN KELOMPOK PRIORITAS PENERIMA BANTUAN BEDAH RUMAH MENGGUNAKAN METODE *CLUSTERING* K-*MEANS* (Studi Kasus : KANTOR KECAMATAN BAHAR UTARA). *Jurnal Manajemen Sistem Informasi*, *1*(2).

Elvitaria, L., & Havenda, M. (n.d.). *RABIT(Jurnal Teknologi dan Sistem Informasi Univrab) MEMPREDIKSI TINGKAT PEMINAT EKSTRAKURIKULER PADA SISWA SMK ANALISIS KESEHATAN ABDURRAB MENGGUNAKAN ALGORITMA C4.5 (STUDI KASUS: SMK ANALIS KESEHATAN ABDURRAB)*.

Fatmawati, K., & Windarto, A. P. (2018). *DATA MINING: PENERAPAN RAPIDMINER DENGAN K-MEANS CLUSTER PADA DAERAH TERJANGKIT DEMAM BERDARAH DENGUE (DBD) BERDASARKAN PROVINSI* (Vol. 3, Issue 2). https://www.depkes.go.id/.

Gunawan, H., & Purwayoga, V. (2022). DATA MINING MENGGUNAKAN ALGORITMA K-*MEANS CLUSTERING* UNTUK MENGETAHUI POTENSI PENYEBARAN VIRUS CORONA DI KOTA CIREBON. *Jurnal Sisfokom (Sistem Informasi Dan Komputer)*, *11*(1), 1–8. https://doi.org/10.32736/sisfokom.v11i1.1316

Hardiani, T. (2022). Analisis *Clustering* Kasus Covid 19 di Indonesia Menggunakan Algoritma K-*Means*. *Jurnal Nasional Pendidikan Teknik Informatika (JANAPATI)*, *11*(2), 156–165. https://doi.org/10.23887/janapati.v11i2.45376

Nyoman, I., & Adiputra, M. (2021). *CLUSTERING* PENYAKIT DBD PADA RUMAH SAKIT DHARMA KERTI MENGGUNAKAN ALGORITMA K-*MEANS*. *INSERT: Information System and Emerging Technology Journal*, *2*(2), 99.

Purba, N., & Satria Tambunan, H. (2021). Penerapan Algoritma K-*Means Clustering*

Pada Penyebaran Penyakit Infeksi Saluran Pernapasan Akut (ISPA) di Provinsi Riau. In *Journal of Information System Research* (Vol. 2, Issue 3).

Rosiana, P. S., Mohsa, A. A., & Umaidah, Y. (2023). IMPLEMENTASI K-*MEANS* DALAM PENGELOMPOKAN PENYEBARAN PENYAKIT DBD DI JAWA BARAT. *Jurnal Informatika Dan Teknik Elektro Terapan*, *11*(3). https://doi.org/10.23960/jitet.v11i3.3344

Sembiring, M. A., Agus, T. A., Fitri, M., Sibuea, L., Tinggi, S., Informatika, M., & Royal, D. K. (2021). PENERAPAN METODE ALGORITMA K-*MEANS CLUSTERING* UNTUK PEMETAAN PENYEBARAN PENYAKIT DEMAM BERDARAH DENGUE (DBD). In *Journal of Science and Social Research* (Issue 3). http://jurnal.goretanpena.com/index.php/JSSR

Supriyadi, A., Triayudi, A., Sholihati, I. D., Fakultas,), Komunikasi, T., Informatika, D., Nasional, U., Sawo Manila, J., Minggu, P., Selatan, J., Khusus, D., Jakarta, I., Penulis,)

*, Abstrak, K., & Kunci, K. (n.d.). *PERBANDINGAN ALGORITMA K-MEANS DENGAN K-MEDOIDS PADA PENGELOMPOKAN ARMADA KENDARAAN TRUK BERDASARKAN PRODUKTIVITAS*.

Widiastuti, S. H., & Jumardi, R. (2022). Pengelompokan Daerah Rawan Demam Berdarah dengan Metode K-*Means Clustering*. *Jurnal Informasi Dan Teknologi*. https://doi.org/10.37034/jidt.v4i4.213