

## Pemodelan Prediktif Keterlambatan Bicara pada Balita Terkait dengan Penggunaan Smartphone Menggunakan Data Mining

Uci Suryani<sup>1</sup>, Ilsa Palingga Ninditama<sup>2</sup>, Wahyudi Syaputra<sup>3</sup>

Program Studi Teknologi Rekayasa Multimedia, Politeknik Prasetya Mandiri, Palembang,  
Indonesia

Email: uci.suryani1@gmail.com

### Abstrak

Penelitian ini bertujuan untuk memprediksi tingkat keterlambatan bicara pada balita yang sudah terindeksi *smartphone* dengan menggunakan metode klasifikasi dan algoritma *Decision Tree* C4.5. Atribut yang digunakan dalam penelitian ini mencakup umur (usia), durasi, dan jenis aplikasi yang digunakan. Pengelolaan data untuk prediksi tingkat keterlambatan bicara pada balita menggunakan tahapan *Knowledge Discovery in Database* (KDD) dengan alat bantu *tools RapidMiner*. Proses penghitungan data dengan algoritma *Decision Tree* menunjukkan bahwa tingkat keterlambatan bicara (*Speech Delayed*) yang terlambat lebih rendah dibandingkan dengan tingkat balita yang tidak mengalami keterlambatan bicara (Normal). Hasil akurasi prediksi sebesar 89.59%. Evaluasi dengan metrik AUC juga menunjukkan nilai 89.59%, mengindikasikan bahwa model ini memiliki kemampuan klasifikasi yang hampir sempurna. Temuan ini memverifikasi bahwa model mampu memprediksi tingkat keterlambatan Bicara pada balita dengan tingkat akurasi yang tinggi.

**Kata Kunci:** Keterlambatan Bicara, *Decision Tree*, Balita

### Abstract

This research aims to predict the level of speech delay in toddlers who have been exposed to smartphones using classification methods and the *Decision Tree* C4.5 algorithm. The attributes used in this study include age, duration, and type of application used. Management of data for predicting speech delay levels in toddlers using the *Knowledge Discovery in Database* (KDD) stages with the aid of *RapidMiner* tools. The data calculation process using the *Decision Tree* algorithm shows that the rate of speech delay (*Speech Delayed*) is lower compared to the rate of toddlers who do not experience speech delay (Normal). The prediction accuracy result is 89.59%. Evaluation with the AUC metric also shows a value of 89.59%, indicating that this model has an almost perfect classification capability. These findings verify that the model can predict speech delay in toddlers with a high level of accuracy.

**Keywords:** Speech Delays, Decision Trees, Toddlers



## 1. PENDAHULUAN

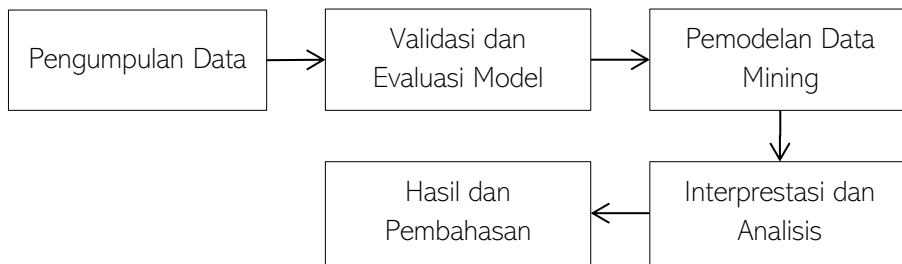
Kemajuan teknologi informasi dan komunikasi telah membawa perubahan signifikan dalam kehidupan manusia, termasuk dalam pola pengasuhan anak. Salah satu fenomena yang sering diamati adalah peningkatan penggunaan smartphone oleh balita. [1]. Keterlambatan bicara pada balita menjadi salah satu isu penting yang perlu mendapatkan perhatian serius. Keterlambatan ini dapat mempengaruhi perkembangan sosial, emosional, dan kognitif anak di masa depan.[2] [3]. Oleh karena itu, penting untuk memahami faktor-faktor yang berkontribusi terhadap keterlambatan bicara ini, salah satunya adalah penggunaan smartphone. [4],[5]. Kemampuan anak untuk berkomunikasi dimulai dengan reaksinya terhadap bunyi atau suara ibu bapaknya, bahkan di usia 2 bulan anak sudah menunjukkan senyum sosial pada semua orang yang berinteraksi dengannya. Diusia 18 bulan anak sudah mampu memahami dan mengeluarkan sekitar 20 kosa kata yang bermakna. Sedangkan di usia 2 tahun sudah mampu mengucapkan 1 kalimat yang terdiri dari 2 kata, misalnya “mama pergi”, “aku pipis”. Jika anak tidak mengalami hal tersebut bisa dikategorikan anak tersebut mengalami keterlambatan berbicara (speech delayed).[6]. Namun, untuk memahami sejauh mana penggunaan smartphone berkontribusi terhadap keterlambatan bicara dan faktor-faktor apa saja yang memediasi hubungan tersebut.

Maka dengan memanfaatkan kemajuan teknologi informasi yang menghasilkan berbagai metode dalam data science yang mempermudah pengambilan keputusan di berbagai bidang, salah satunya adalah metode Decision Tree. Decision Tree (Pohon Keputusan) adalah metode pembelajaran yang membangun model prediktif dalam bentuk struktur pohon. Pohon ini terdiri dari simpul-simpul (nodes) yang mewakili keputusan berdasarkan nilai-nilai atribut, serta cabang-cabang yang mewakili hasil dari keputusan tersebut.[7]. Dengan demikian, banyak peneliti telah melakukan penelitian yang memanfaatkan kemampuan-kemampuan ini. contohnya, Terdapat penelitian yang berfokus Deteksi Kejadian Speech Delayed Pada Anak dengan menggunakan 100 data sample. menggunakan algoritma ID3 (algoritma pembelajaran pohon keputusan) dengan jenis desain retrospektif terdapat 51 anak speech delayed dan 49 anak tidak speech delayed. [8]. Selain itu, sebuah penelitian melakukan intensitas penggunaan gadget dan keterlambatan dalam perkembangan aspek bicara dan bahasa menggunakan 100 responden, 53 responden (53%) memiliki intensitas penggunaan gadget yang rendah, 29 responden (29%) memiliki intensitas penggunaan gadget yang sedang, dan 18 responden (18%) memiliki intensitas penggunaan gadget yang tinggi. Sebanyak 39 responden (39%) mengalami keterlambatan dalam perkembangan aspek bicara dan

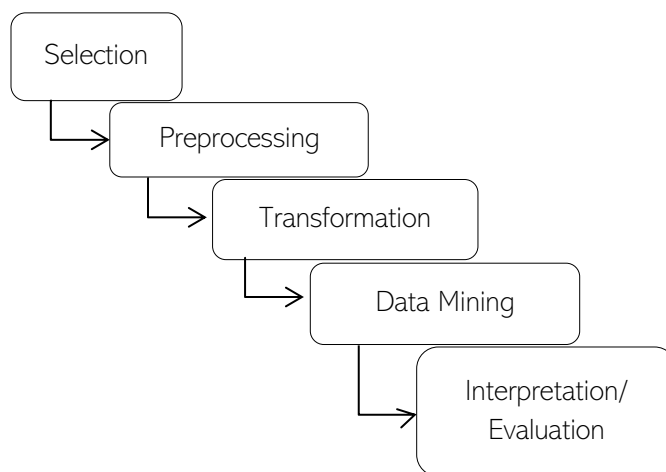
bahasa, sementara 61 responden lainnya (61%) tidak mengalami keterlambatan. Analisis Chi-square menunjukkan hubungan yang signifikan antara intensitas penggunaan gadget dan keterlambatan dalam perkembangan aspek bicara dan bahasa ( $p=0.002$ ). Penelitian ini bertujuan untuk menerapkan algoritma *DecisionTree* C4.5 dalam data mining guna memprediksi tingkat keterlambatan bicara pada balita yang sudah terindeksi *smartphone*.

## 2. METODE

Data yang digunakan dalam penelitian ini adalah data balita pengguna *smartphone* dengan menggunakan metode *Decision Tree* untuk menganalisis dan memprediksi data. Tahapan penelitian pemodelan prediktif keterlambatan bicara pada balita terkait dengan penggunaan *smartphone* menggunakan *Decision Tree*, berikut pada Gambar 1 adalah langkah-langkah penelitian.



Gambar 1. Tahapan Penelitian



Gambar 2. Knowledge Discovery in Databases (KDD)

Penelitian ini mengimplementasikan pengolahan data mining dengan mengikuti tahapan terstruktur dalam *Knowledge Discovery in Databases* (KDD). Penelitian ini bertujuan untuk menghasilkan informasi yang relevan sesuai dengan urutan proses yang telah ditentukan. Proses penelitian ini dapat dilihat pada Gambar 2.

### 2.1. Data Selection

Pada tahap seleksi data (*data selection*) dalam pemodelan prediktif keterlambatan bicara pada balita terkait dengan penggunaan *smartphone*. Tahap awal yang dilakukan ialah menentukan kriteria atau parameter untuk pemilihan variabel dan data yang relevan sesuai dengan tujuan *Knowledge Discovery in Databases* (KDD). Pada tahap ini, dilakukan identifikasi dan penghapusan data yang duplikat atau redundan. Menghilangkan data duplikat adalah langkah penting untuk menjaga integritas dan konsistensi data dalam analisis.

### 2.2. Data Preprocessing

Tahapan ini merupakan langkah penting dalam analisis data dan data mining yang bertujuan untuk mempersiapkan data mentah menjadi data yang berkualitas dan siap untuk digunakan dalam pemodelan atau analisis lebih lanjut. [9]. Proses ini melibatkan beberapa tahapan untuk memastikan bahwa data yang akan digunakan akurat, konsisten, dan relevan. Langkah yang dilakukan pada tahap preprocessing data dalam penelitian ini, pembersihan data, pengurangan data, dan normalisasi data.

### 2.3. Transformation

Transformasi data ini dilakukan untuk memperbaiki kualitas data serta memberikan pemahaman yang lebih mendetail tentang faktor-faktor yang berpengaruh terhadap tingkat keterlambatan bicara pada balita pengguna *smartphone*. Teknik-teknik yang diterapkan pada tahapan ini termasuk transformasi atribut, transformasi skala, penciptaan fitur baru, serta diskritisasi data. [10].

### 2.4. Data Mining

Tahapan ini merupakan proses menemukan pola, tren, dan hubungan dalam kumpulan data besar menggunakan metode statistik, matematika, dan algoritma pembelajaran mesin. Data yang digunakan dalam analisis ini adalah data training yang telah melalui proses transformasi, sehingga siap untuk proses data mining. Proses pembentukan data training dalam data mining melibatkan beberapa

langkah penting untuk memastikan bahwa data yang digunakan memiliki kualitas yang baik dan siap untuk digunakan dalam model pembelajaran mesin. Pertama, Mengumpulkan data dari berbagai sumber. Kemudian, data harus disaring (seleksi) untuk menentukan atribut-atribut yang dapat mempengaruhi keterlambatan bicara pada balita, yang disebut sebagai data target. Data target berisi atribut-atribut kunci yang mendukung proses data mining dalam memprediksi tingkat Keterlambatan bicara pada balita. Langkah-langkah ini memastikan data training terbentuk dengan baik dan siap dipakai dalam proses data mining.

Tabel 1. Parameter Data

Atribut	Keterangan
Usia (Tahun)	1-2
	2-3
	3-4
	4-5
Jenis Kelamin	Laki-laki
	Perempuan
Durasi Per Hari (Menit)	<30 (Rendah)
	>30 (Sedang)
	>60 (Tinggi)
Aplikasi yang digunakan	Youtube
	Aplikasi Game
	Aplikaasi Pembelajaran
Perkembangan bicara	Normal (Tidak)
	Terlambat (Ya)

Pada Table 1 menunjukkan parameter data, menunjukkan atribut yang akan digunakan pada data training. Data training merupakan data yang sudah siap untuk diolah dalam data mining. Adapun beberapa contoh *data training* dapat dilihat pada Tabel 2)

Tabel 2. Data Training

No	Usia (Tahun)	Jenis Kelamin	Durasi Penggunaan Per Hari	Jenis Aplikasi yang digunakan	Perkembangan Bicara
1	1-2	Laki-laki	Rendah	Youtube	Tidak
2	2-3	Laki-laki	Sedang	Aplikasi Game	Ya
3	2-3	Laki-laki	Rendah	Aplikasi Game	Tidak
4	3-4	Perempuan	Rendah	Youtube	Tidak

5	3-4	Laki-laki	Rendah	youtube	Tidak
6	2-3	Perempuan	Sedang	Aplikasi Game	Ya
7	3-4	Perempuan	Tinggi	Youtube	Ya
8	3-4	Perempuan	Rendah	Aplikasi Pembelajaran	Tidak
9	3-4	Laki-laki	Sedang	Aplikasi Game	Ya
10	3-4	Laki-laki	Tinggi	Youtube	Ya
11	2-3	Perempuan	Rendah	Aplikasi Game	Ya

Pada Table 2 terdapat Data training, dimana data yang sudah siap untuk diolah atau di proses untuk membuat sebuah pohon keputusan menggunakan metode *Decision Tree* dengan algoritma C4.5. beberapa tahapan dalam membuat sebuah pohon keputusan dalam algoritma C4.5 yaitu:

1. Menyiapkan data pelatihan (Data Training). Data training biasanya berasal dari data historis yang telah terjadi sebelumnya, atau disebut juga data masa lalu, dan telah dikelompokkan dalam kelas-kelas tertentu.
2. Menentukan akar pohon, Akar dipilih dari atribut yang ada dengan menghitung nilai gain masing-masing atribut, di mana atribut dengan nilai gain tertinggi akan menjadi akar pertama. Sebelum menghitung nilai gain dari atribut, hitung terlebih dahulu nilai entropy. Entropy adalah parameter untuk mengukur tingkat keberagaman (heterogenitas) dari kumpulan data. Semakin besar nilai entropy, semakin tinggi tingkat keberagaman kumpulan data tersebut. Untuk menghitung nilai entropy, digunakan rumus:

$$Entropy(S) = - \sum_{i=1}^n p_i * \log_2(p_i) \quad (1)$$

Di mana S adalah himpunan kasus, A adalah fitur, n adalah partisi dari S, dan Pi adalah proporsi jumlah Si terhadap S.

3. Menghitung nilai Gian menggunakan persamaan sebagai berikut:

$$Gaint(S,A)=Entropy(S)- \sum_{i=1}^n \frac{s_i}{s} Entrophy(s_i) \quad (2)$$

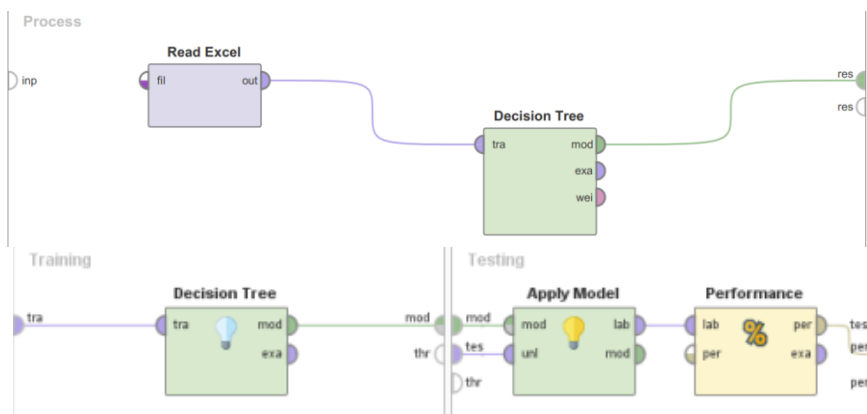
4. Ulangi langkah ke 2 dan langkah ke 3 hingga semua *record* terpartisi
5. Proses partisi pohon keputusan akan berhenti saat semua *record* dalam simpul N mendapat kelas yang sama dan tidak ada atribut di dalam *record* yang dipartisi lagi. Pada tahap ini, pembentukan pohon keputusan selesai karena telah mencapai kondisi akhir yang diinginkan.

## 2.5. Interpretation

Tahapan ini merupakan proses pemeriksaan pola yang diidentifikasi melalui data mining dapat divisualisasikan sebagai pohon keputusan apakah sesuai dengan fakta atau eksperimen sebelumnya. [11]. Pola informasi yang dihasilkan dari data mining harus disampaikan dalam bentuk yang jelas dan mudah dipahami oleh pihak terkait. Pada metode *Decision Tree*, pola atau informasi ini berupa aturan-aturan (*rules*) yang dihasilkan dari decision tree yang telah dibuat. Interpretasi yang akurat memungkinkan pihak-pihak terkait untuk merumuskan langkah-langkah yang tepat berdasarkan pola dan informasi yang diungkap oleh data mining.

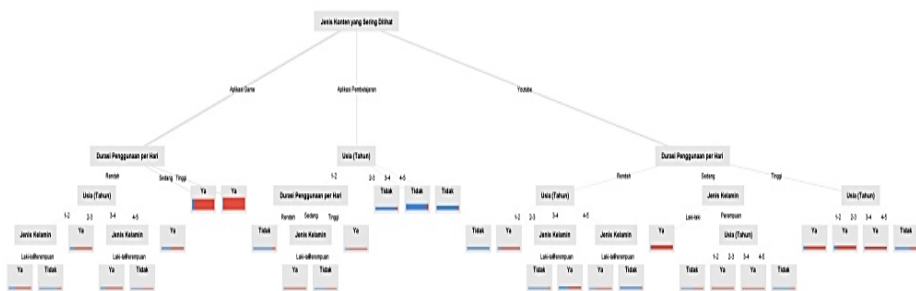
## 3. HASIL DAN PEMBAHASAN

Penelitian ini melakukan eksperimen dengan memanfaatkan *software RapidMiner* untuk analisis data serta menggunakan algoritma C4.5 sebagai pemodelan. Pengolahan data menggunakan metode KDD. Data yang digunakan sebanyak 701 *record* data balita pengguna *smartphone*, terdiri dari 463 balita yang mengalami keterlambatan bicara dan 238 balita yang tidak mengalami keterlambatan bicara (normal). Untuk mengidentifikasi pola balita yang mengalami keterlambatan bicara dengan pemodelan data dan diproses menggunakan *software Rapidminer*. *RapidMiner* akan menghasilkan pohon keputusan yang dapat dilihat secara visual. Pohon keputusan ini akan menampilkan cabang-cabang keputusan berdasarkan atribut-atribut yang paling penting. Pemodelan klasifikasi menggunakan algoritma *Decision Tree* C4. 5 menggunakan *RapidMiner* dapat dilihat pada Gambar 3.



Gambar 3. Model diklasifikasikan melalui algoritma *Decision Tree* C4.5.

Akurasi dan model yang dihasilkan oleh algoritma C4.5 dipengaruhi oleh penggunaan parameter. Parameter *Cross-validation* digunakan untuk membentuk pohon Keputusan untuk mengevaluasi kinerja model, untuk membantu menentukan seberapa baik performa model pada data baru. Dalam proses data mining untuk *Classification*, *Decision Tree* berfungsi sebagai algoritma operator. Selain itu, operator *Apply Model* dan Operator *Performance* adalah bagian dari proses data mining yang digunakan untuk menghasilkan *Decision Tree*. Untuk Hasil Pemodelan Algoritma *Decision Tree* C4.5 terdapat pada Gambar 4.



Gambar 4. Hasil Pemodelan Algoritma *Decision Tree* C 4.5

Dari eksperimen yang dilakukan, informasi yang didapatkan berupa pola pemodelan dan informasi tentang keakuratan data yang digunakan. Jenis aplikasi yang digunakan menjadi simpul akar (*root node*). Dimana pada *node 5* terdapat 60 balita yang belum diketahui status keterlambatan bicara atau tidak. *Node 6* terdapat 11 balita yang tidak mengalami keterlambatan bicara dan 110 balita yang mengalami keterlambatan sedangkan *node 7* terdapat 3 balita tidak mengalami keterlambatan bicara dan 135 balita yang mengalami keterlambatan bicara.

Selain itu pada *node 10* terdapat 99 balita yang belum diketahui berapa yang mengalami keterlambatan bicara, *node 24* terdapat 3 balita tidak mengalami keterlambatan bicara 25 yang mengalami keterlambatan bicara. *Node 25* ada 4 balita tidak mengalami keterlambatan dan 29 yang mengalami keterlambatan. Dengan hasil akurasi data sebesar 89.59% dengan menggunakan *confusion matrix*. Dapat dilihat pada Gambar 5.



● Table View ○ Plot View

accuracy: 89.59% +/- 2.33% (micro average: 89.59%)

	true Tidak	true Ya	class precision
pred. Tidak	186	21	89.86%
pred. Ya	52	442	89.47%
class recall	78.15%	95.46%	

Gambar 5. Hasil *Accuracy* Algoritma *Decision Tree* C 4.5

Penelitian ini pun memanfaatkan metrik evaluasi AUC (*Area Under the Curve*) guna menilai efektivitas model klasifikasi, terutama dalam konteks klasifikasi biner. AUC (*Area Under the Curve*) menilai sejauh mana model dapat membedakan antara kelas positif dan negatif. Evaluasi metrik dengan AUC (*Area Under the Curve*) menunjukkan nilai 89,6%.

#### 4. KESIMPULAN

Penggunaan *smartphone* yang berlebihan pada balita berhubungan erat dengan peningkatan risiko keterlambatan bicara. Model *Decision Tree* C4.5 yang digunakan dalam penelitian ini berhasil memprediksi keterlambatan bicara dengan tingkat akurasi yang tinggi, menunjukkan bahwa faktor-faktor seperti durasi penggunaan *smartphone*, usia balita saat mulai menggunakan *smartphone*, dan aplikasi yang diakses merupakan prediktor penting. Akurasi mencapai 89,59% ini menandakan bahwa model yang dibuat mampu melakukan prediksi dengan akurat. Penting untuk diperhatikan bahwa tingginya akurasi mungkin disebabkan oleh rendahnya kompleksitas data, yang membuat model dapat memprediksi dengan akurasi yang cukup baik.

#### REFERENSI

- [1] Y. Rismala, Aguswan, D. E. Priyantoro, and Suryadi, "Dampak Penggunaan Gadget Terhadap Perkembangan Sosial Anak Usia Dini," *El-Athfal J. Kaji. Ilmu Pendidik. Anak*, vol. 1, no. 01, pp. 46–55, 2021, doi: 10.56872/elathfal.v1i01.273.
- [2] A. D. Parahita, K. A. Harras, and J. Nurhadi, "Studi Kasus Bahasa Lisan Anak Terlambat Bicara: Kajian Psikolinguistik," *J. PESONA*, 2022, doi: 10.52657/jp.v8i1.1651.
- [3] N. Hasanah and S. Sugito, "Analisis Pola Asuh Orang Tua terhadap Keterlambatan Bicara pada Anak Usia Dini," *J. Obs. J. Pendidik. Anak Usia Dini*, vol. 4, no. 2, p. 913, 2020, doi: 10.31004/obsesi.v4i2.456.

- [4] E. Deviany Widyawaty and M. Jannah, "Faktor-Faktor Yang Mempengaruhi Kejadian Keterlambatan Bicara Di House of Fatima Child Center Kota Malang," *Heal. Care Media*, vol. 5, pp. 3–6, 2021.
- [5] K. B. Suryawan and L. T. Merijanti, "Bermain aplikasi gadget berhubungan dengan keterlambatan perkembangan bicara dan bahasa pada balita," *J. Biomedika dan Kesehatan*, 2021, doi: 10.18051/jbiomedkes.2021.v4.157-163.
- [6] A. N. Istiqlal, "Gangguan Keterlambatan Berbicara (Speech Delay) Pada Anak Usia 6 Tahun," *Preschool*, vol. 2, no. 2, pp. 206–216, 2021, doi: 10.18860/preschool.v2i2.12026.
- [7] M. R. Qisthiano, P. A. Prayesy, and I. Ruswita, "Penerapan Algoritma Decision Tree dalam Klasifikasi Data Prediksi Kelulusan Mahasiswa," *G-Tech J. Teknol. Terap.*, vol. 7, no. 1, pp. 21–28, Jan. 2023, doi: 10.33379/gtech.v7i1.1850.
- [8] N. Hestiyana *et al.*, "Deteksi Kejadian Speech Delayed Pada Anak Dengan Algoritma Id3," *Din. Kesehat. J. Kebidanan dan Keperawatan*, vol. 12, no. 2, pp. 2549–4058, 2021, doi: 10.33859/dksm.v12i2.
- [9] A. Muzakir, H. Syaputra, and F. Panjaitan, "A Comparative Analysis of Classification Algorithms for Cyberbullying Crime Detection: An Experimental Study of Twitter Social Media in Indonesia," *Sci. J. Informatics*, vol. 9, no. 2, pp. 133–138, Oct. 2022, doi: 10.15294/sji.v9i2.35149.
- [10] F. A. Rahman, M. I. Desa, A. Wibowo, and N. A. Haris, "Knowledge discovery database (KDD)-data mining application in transportation," *Int. Conf. Electr. Eng. Comput. Sci. Informatics*, vol. 1, no. 978, pp. 116–119, 2014, doi: 10.11591/eecsi.1.357.
- [11] F. SLN, "Buku Dasar Data Mining from A to Z," no. January, pp. 15–15, 2023.