

# Penerapan Algoritma Decision Tree dalam Segmentasi Customer

Ni Putu Vina Amandari<sup>1</sup>, Dr. Ngurah Agus Sanjaya ER<sup>2</sup>

<sup>1,2</sup>Program Studi Informatika  
Fakultas Matematika dan Ilmu Pengetahuan Alam  
Universitas Udayana

<sup>1</sup>vinamandari1@gmail.com, <sup>2</sup>agus\_sanjaya@unud.ac.id

## Abstract

*Segmentasi customer merupakan proses pembagian customer yang dilakukan oleh suatu bisnis guna mengetahui target pasar yang sesuai dengan usaha yang dijalankan. Customer akan dibagi menjadi beberapa kelompok sesuai dengan karakteristiknya seperti usia, frekuensi pembelian, jenis kelamin, pekerjaan, dan lain sebagainya. Tujuan dilakukannya segmentasi customer yaitu mengembangkan hubungan yang lebih baik dengan cara memahami kebutuhan setiap segmen pelanggan, meningkatkan profitabilitas dengan cara membuat strategi pemasaran yang lebih efektif, serta mengidentifikasi customer yang kemungkinan dapat meningkatkan pendapatan suatu usaha. Pada penelitian ini akan dilakukan klasifikasi customer menggunakan algoritma decision tree dengan tujuan untuk mengetahui tingkat akurasi dari algoritma decision tree ketika digunakan untuk mengklasifikasi customer menjadi beberapa segmen. Dataset yang digunakan pada penelitian kali ini didapatkan dari website Kaggle.com dimana dataset ini terdiri dari data training dan data testing.*

**Keywords:** Decision Tree, Segmentasi Customer

## 1. Pendahuluan

Perkembangan teknologi yang semakin pesat menyebabkan hampir seluruh aspek kehidupan seperti aspek ekonomi, kesehatan, industri, dan aspek lainnya bergantung pada teknologi. Dalam persaingan antar bisnis, suatu perusahaan dituntut agar dapat memanfaatkan semaksimal mungkin sumber daya yang ada agar mampu bersaing dengan perusahaan lain. Sejalan dengan peradaban manusia, terjadi perubahan terhadap kebutuhan pelanggan. Dengan berubahnya kebutuhan pelanggan maka diperlukan perubahan juga dalam bidang pemasaran. Bagian marketing dalam suatu perusahaan sangat berperan besar terhadap keberhasilan suatu perusahaan dalam mendapatkan customer. Ditambah lagi ketatnya persaingan antara perusahaan satu dengan perusahaan lainnya membuat suatu perusahaan harus memiliki strategi pemasaran yang baik. Untuk menciptakan strategi pemasaran yang baik kita perlu mengetahui target pasar kita serta mempertimbangkan karakteristik dari setiap customer, maka dari itu segmentasi customer sangat penting dilakukan. Customer segmentation merupakan membagi-bagi pasar menjadi beberapa kelompok pembeli berbeda yang mungkin memerlukan produk atau jasa yang berbeda pula [1]. Dengan melakukan segmentasi customer kegiatan pemasaran suatu perusahaan lebih terarah mengingat banyaknya customer yang memiliki keinginan dan kebutuhan yang berbeda. Adapun beberapa tujuan dilakukannya segmentasi customer yaitu mengembangkan hubungan yang lebih baik dengan cara memahami kebutuhan setiap segmen pelanggan, meningkatkan profitabilitas dengan cara membuat strategi pemasaran yang lebih efektif, serta mengidentifikasi customer yang kemungkinan dapat meningkatkan pendapatan suatu usaha. Untuk melakukan segmentasi customer dibutuhkan suatu algoritma yang dapat membantu dalam mengklasifikasikan customer ke dalam beberapa kelompok.

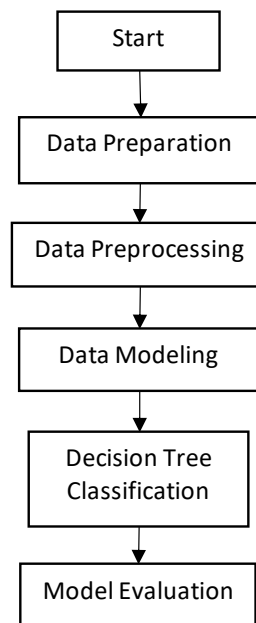
Pada penelitian yang Alkahfi Madani, dkk (2022) dengan judul Segmentasi Pelanggan pada BC HNI 2 Pekanbaru dengan Menerapkan Algoritma K-Medoids dan Model Recency, Frequency, Monetary (RFM), dilakukan segmentasi pelanggan pada BC HNI 2 Pekanbaru dengan menerapkan algoritma K-Medoids dan Model RFM. Setelah melakukan klasifikasi didapatkan 2 kelompok customer dimana customer pada kelompok 1 merupakan pelanggan dengan kategori pelanggan utama yang recent

transaction time yang rendah dengan artian bahwa baru-baru ini dan frekuensinya juga tinggi hal ini diartikan bahwa pelanggan sering bertransaksi antara pelanggan dan perusahaan serta rendahnya nilai monetary yang berarti total uang yang dipakai tidak terlalu besar [2]. Sedangkan customer pada kelompok 2 merupakan customer yang tergolong bertransaksi hanya pada awal bulan dan bertransaksi ketika sangat membutuhkan barang tersebut. Kemudian pada penelitian yang dilakukan oleh Nana Suryana (2017) dengan judul penelitian Prediksi Churn dan Segmentasi Pelanggan TV Berlangganan (Studi Kasus Transvision Jawa Barat), dilakukan prediksi churn rate dan segmentasi pelanggan menggunakan algoritma K-Means. Berdasarkan penelitian tersebut tingkat akurasi yang didapatkan sebesar 90,89%.

Pada penelitian ini, proses klasifikasi dilakukan dengan menggunakan algoritma Decision Tree pada suatu dataset. Dataset yang digunakan diperoleh dari website Kaggle.com dengan nama dataset “Customer Segmentation Classification”. Dataset ini terdiri dari data training dan data testing dimana masing-masing set data memiliki 11 atribut. Tujuan dari penelitian ini yaitu untuk mengetahui akurasi dari algoritma Decision Tree jika digunakan untuk mengklasifikasikan customer menjadi beberapa segmen.

## 2. Metode Penelitian

Pada penelitian ini akan menggunakan library python mulai dari tahap data preparation sampai model evaluation. Berikut ini tahapan dalam penelitian ini :



**Gambar 1.** Tahapan Penelitian

### 2.1 Data Preparation

Pada tahap ini dilakukan proses pengumpulan data serta informasi yang berkaitan dengan penelitian yang akan dilakukan. Dataset yang digunakan dalam penelitian yaitu “Customer Segmentation Classification” yang didapatkan dari website *Kaggle.com*. Dataset ini terdiri dari data training dan data testing dimana masing-masing data terdiri dari 11 atribut. Data training digunakan untuk melatih algoritma dalam mencari model yang sesuai sedangkan data testing digunakan untuk menguji performa dari algoritma yang sudah dilatih sebelumnya.

**Tabel 1.** Atribut Dataset

No	Atribut	Penjelasan
1.	ID	Unique ID
2.	Gender	Gender dari customer
3.	Ever_Married	Status pernikahan customer (sudah atau belum)

4.	Age	Usia customer
5.	Graduated	Sudah lulus sekolah atau belum
6.	Profession	Profesi customer
7.	Work_Experience	Pengalaman kerja
8.	Spending_Score	Skor yang diberikan customer
9	Family_Size	Jumlah anggota keluarga customer
10.	Var_1	Kategori anonim customer
11.	Segmentation	Segmentasi pelanggan (A,B,C,D)

## 2.2 Data Preprocessing

### 2.2.1 Cleaning Dataset

Cleaning dataset dilakukan dengan tujuan memastikan bahwa tidak terdapat missing value dalam dataset sehingga tidak menimbulkan error ketika tahap klasifikasi berlansung. Missing value merupakan informasi yang tidak tersedia dalam suatu data dimana hal ini dapat terjadi karena beberapa hal diantaranya responden menolak untuk menjawab, informasi tidak tersedia atau sulit untuk dicari, kesalahan ketika mengumpulkan data, dan lain sebagainya. Pada penelitian ini ketika ditemukan missing value pada dataset maka data tersebut nantinya akan dihapus.

### 2.2.2 Encoding Nominal Data

Model machine learning hanya dapat memahami data berupa data numerikal sedangkan dataset yang digunakan terdiri dari data numerikal dan data kategorikal. Maka dari itu dibutuhkan sebuah metode yang dapat mengubah data kategorikal menjadi data numerikal. Encoding nominal data merupakan transformasi data kategorikal menjadi data numerikal. Bentuk transformasi data menggunakan mekanisme sederhana yaitu mengubah semua data kategorikal dengan kode numerik 0 - n, dimana n adalah varian terakhir dari data kategorikal pada atribut tersebut [3]. Tahap ini dapat dilakukan dengan menggunakan salah satu library python yaitu *sci-kit LabelEncoder* kemudian dilakukan proses *fit\_transform()* terhadap kolom yang ingin dirubah.

### 2.2.3 Handle Unwanted Column

Unwanted Column merupakan kolom atau fitur yang tidak relevan dengan hasil klasifikasi nantinya. Maka dari itu untuk meningkatkan kinerja dari model klasifikasi maka fitur-fitur yang tidak relevan akan dihapus dengan cara melihat korelasinya terlebih dahulu dengan fitur lainnya.

### 2.2.4 Handle Outliers

Outlier merupakan data yang memiliki nilai-nilai yang jauh berbeda dibandingkan dengan kelompoknya baik itu terlalu tinggi maupun terlalu rendah. Pada penelitian ini data outlier dideteksi menggunakan metode boxplot. Konsep metode ini adalah menggunakan nilai dari jangkauan interkuartil atau *Interquartile Range (IQR)* yang merupakan selisih antara kuartil 1 terhadap kuartil 3 [4]. Data yang dapat dikatakan outlier adalah data yang nilainya lebih dari  $Q3 + 1.5 * IQR$  dan data yang nilainya kurang dari  $Q1 - 1.5 * IQR$  dimana  $Q3$  adalah kuartil 3 dan  $Q1$  merupakan kuartil 1. Nilai yang dinyatakan sebagai outlier nantinya akan diganti dengan nilai *mean* dari kelompok datanya.

## 2.3 Modeling Data

### 2.3.1 X and y definision

Pada tahap ini data dibagi ke dalam 2 variabel yaitu x dan y dimana x sebagai fitur atau variabel yang mempengaruhi dan y sebagai label atau variabel yang dipengaruhi.

### 2.3.2 Feature Scalling

Scalling Features merupakan suatu cara untuk membuat data numerik pada dataset supaya memiliki jangkauan nilai (scale) yang sama [5]. Terdapat 3 scaler pada library scikit-learn yang sering digunakan untuk feature scalling dataset yaitu *StandardScaler*, *MinMaxScaler*, dan *RobustScaler* namun pada penelitian ini digunakan *StandardScaler*. *StandardScaler* merupakan suatu metode dimana metode tersebut akan melakukan standarisasi fitur dengan menghapus rata-

rata dan menskalakan unit varian [6]. Rumus dari Standard Scaler ditunjukkan pada persamaan di bawah, dimana  $\bar{X}$  adalah rata-rata nilai sampel dan  $\sigma$  adalah standar deviasi.

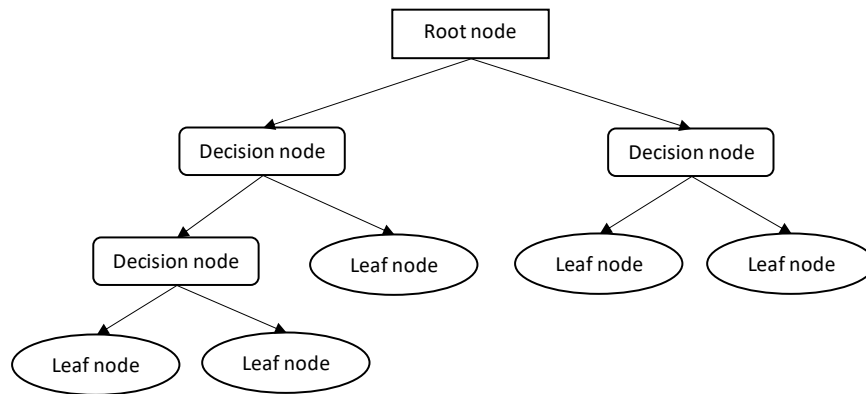
$$Z = \frac{X_i - \bar{X}}{\sigma}$$

### 2.3.3 Split Data

Pada tahap ini dengan menggunakan library python yaitu sklearn, data akan dibagi menjadi 2 yaitu data training dan data testing dengan perbandingan 80% untuk data training dan 20% untuk data testing.

## 2.4 Decision Tree Classification

Proses klasifikasi akan menggunakan algoritma Decision Tree. Algoritma Decision Tree merupakan model prediksi terhadap suatu keputusan menggunakan struktur hirarki atau pohon [7]. Setiap pohon memiliki cabang dimana cabang tersebut mewakili sebuah atribut yang harus dipenuhi agar dapat menuju ke cabang selanjutnya, hal tersebut berulang hingga tidak ada cabang lagi. Konsep data dalam decision tree adalah data dinyatakan dalam bentuk tabel yang terdiri dari atribut dan record dimana atribut digunakan sebagai parameter yang dibuat sebagai kriteria dalam pembuatan pohon [7].



**Gambar 2.** Konsep Decision Tree

Tahapan algoritma decision tree dalam membentuk pohon keputusan yaitu [8] :

- Menentukan akar/root
- Menghitung Gain Information yang memiliki nilai terbesar sebagai splitting atribut yang nantinya dipilih sebagai cabang
- Ulangi langkah 2 dan 3 hingga terdapat leaf node.

Gain information mengukur nilai impurity dari suatu partisi, dengan perhitungan :

$$\text{Gini}(D) = 1 - \sum_{i=1}^m P_i^2 \quad (1)$$

Keterangan :

Gini(D) = nilai impurity dari partisi D

M = jumlah indeks

$P_i$  = peluang sebuah *tuple* D pada indeks ke i

Nilai *Average Gini Impurity* dapat dihitung dengan :

$$\text{Gini}_A(D) = \frac{|D1|}{|D|} \text{Gini}(D1) + \frac{|D2|}{|D|} \text{Gini}(D2) \quad (2)$$

Keterangan :

D = *tuple* D

D1 = partisi pertama *tuple* D  
 D2 = partisi kedua *tuple* D  
 $Gini_A(D)$  = impurity dari partisi D pada atribut A  
 $Gini(D1)$  = impurity dari partisi pertama *tuple* D  
 $Gini(D2)$  = impurity dari partisi kedua *tuple* D

Kemudian penurunan tingkat impurity, bisa dihitung dengan :

$$\Delta Gini(A) = Gini(D) - Gini_A(D) \quad (3)$$

Dimana :

$\Delta Gini(A)$  = tingkat impurity

## 2.5 Model Evaluation

Pada tahap ini akan dilakukan evaluasi terhadap performa dari model yang telah dijalankan menggunakan confusion matrix. Confusion matrix adalah alat ukur berbentuk matrix yang digunakan untuk mendapatkan jumlah ketepatan klasifikasi terhadap kelas dengan algoritma yang dipakai [9]. Hasil evaluasi yang akan ditampilkan mencakup accuracy, precision, recall, dan f1-score pada basis per kelasnya.

		Predicted	
		0	1
Actual	0	TN	FP
	1	FN	TP

**Gambar 3.** Confusion Matrix

Confusion Matrix memiliki 4 istilah diantaranya :

1. True Negative (TN) : Ketika data berada di kelas negatif dan model memprediksi data ada di kelas negative
2. True Positive (TP) : Ketika data berada di kelas positif dan model memprediksi data ada di kelas positif
3. False Negative (FN) : Ketika data berada di kelas positif namun model memprediksi data ada di kelas negatif
4. False Positive (FP) : Ketika data berada di kelas negatif namun model memprediksi data ada di kelas negatif

Precision merupakan ketepatan informasi yang diminta oleh user dengan jawaban yang diberikan oleh sistem. Nilai precision didapatkan dari hasil perbandingan antara True Positive (TP) dengan banyaknya data yang diprediksi positif.

$$\text{Precision} = \frac{TP}{(TP+FP)} \quad (1)$$

Recall merupakan kemampuan classifier dalam menemukan kembali informasi. Nilai recall didapatkan dari hasil perbandingan antara setiap kelas yang didefinisikan sebagai rasio True Positive (TP) dengan jumlah True positive (TP) dan False Negative (FN).

$$\text{Recall} = \frac{TP}{(TP+FN)} \quad (2)$$

F1-Score merupakan rata-rata harmonik dari precision dan recall. Nilai terbaik dari F1-Score yaitu 1.0 dan nilai terburuknya adalah 0.

$$\frac{1}{F1} = \frac{1}{2} \left( \frac{1}{precision} + \frac{1}{recall} \right) \quad (3)$$

### 3. Hasil dan Pembahasan

Tahapan awal dalam penelitian ini dimulai dengan import dataset. Setelah melakukan import dataset selanjutnya penulis perlu mengetahui data-data serta nilai apa saja yang terkandung di dalam dataset. Pada Gambar 4 merupakan dataset yang digunakan dalam penelitian ini.

index	ID	Gender	Ever_Married	Age	Graduated	Profession	Work_Experience	Spending_Score	Family_Size	Var_1	Segmentation
0	462809	Male	No	22	No	Healthcare	1.0	Low	4.0	Cat_4	D
1	462643	Female	Yes	38	Yes	Engineer	NaN	Average	3.0	Cat_4	A
2	466315	Female	Yes	67	Yes	Engineer	1.0	Low	1.0	Cat_6	B
3	461735	Male	Yes	67	Yes	Lawyer	0.0	High	2.0	Cat_6	B
4	462669	Female	Yes	40	Yes	Entertainment	NaN	High	6.0	Cat_6	A
5	461319	Male	Yes	56	No	Artist	0.0	Average	2.0	Cat_6	C
6	460156	Male	No	32	Yes	Healthcare	1.0	Low	3.0	Cat_6	C
7	464347	Female	No	33	Yes	Healthcare	1.0	Low	3.0	Cat_6	D
8	465015	Female	Yes	61	Yes	Engineer	0.0	Low	3.0	Cat_7	D
9	465176	Female	Yes	55	Yes	Artist	1.0	Average	4.0	Cat_6	C

Gambar 4. Dataset Awal

Dataset ini memiliki 11 atribut dengan jumlah data sebanyak 8068 baris. Tipe data yang digunakan dalam dataset ini terdiri dari tipe data int64, object, dan float64 seperti yang ditunjukkan pada gambar 5. Jenis data dalam dataset ini terbagi ke dalam dua jenis yaitu data kategorikal dan data numerikal. Tipe data yang termasuk ke dalam data kategorikal yaitu tipe data object sedangkan tipe data yang termasuk ke dalam data numerikal diantaranya tipe data int64 dan float64.

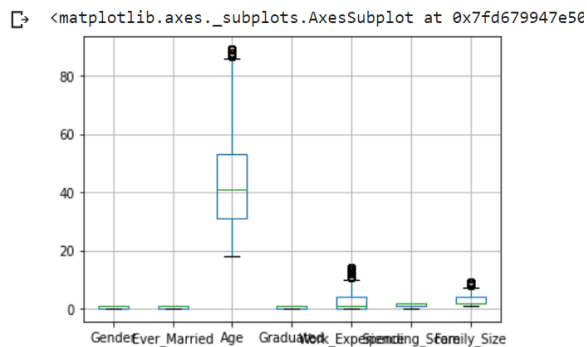
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8068 entries, 0 to 8067
Data columns (total 11 columns):
#   Column          Non-Null Count  Dtype
---  -
0   ID               8068 non-null  int64
1   Gender           8068 non-null  object
2   Ever_Married    7928 non-null  object
3   Age             8068 non-null  int64
4   Graduated       7990 non-null  object
5   Profession       7944 non-null  object
6   Work_Experience  7239 non-null  float64
7   Spending_Score  8068 non-null  object
8   Family_Size     7733 non-null  float64
9   Var_1           7992 non-null  object
10  Segmentation    8068 non-null  object
dtypes: float64(2), int64(2), object(7)
memory usage: 693.5+ KB
```

Gambar 5. Tipe Data Atribut

Tahap selanjutnya yaitu tahap preprocessing, seperti yang ditunjukkan pada gambar 6 terdapat missing value di beberapa atribut diantaranya pada atribut Ever\_Married terdapat 140 missing value, atribut Graduated memiliki 78 missing value, atribut Profession memiliki 124 missing value, atribut Work\_Experience memiliki 829 missing value, atribut Family\_Size memiliki 335 missing value, dan atribut Var\_1 memiliki 76 missing value. Untuk menghindari error ketika proses klasifikasi maka data yang mengandung missing value dihapus. Selain mengandung missing value, dataset yang digunakan juga mengandung data outliers. Data outliers ditemukan di beberapa column diantaranya column age, work experience, dan family size.

```

↳ Missing Value pada setiap atribut:
ID                0
Gender            0
Ever_Married     140
Age              0
Graduated        78
Profession       124
Work_Experience  829
Spending_Score  0
Family_Size     335
Var_1           76
Segmentation     0
dtype: int64
    
```



**Gambar 6.** Missing Value Setiap Atribut

**Gambar 7.** Outliers Setiap Atribut

Hasil dari tahap preprocessing dapat dilihat pada Gambar 8. Jumlah baris pada dataset berkurang dari data awal 8068 baris menjadi 6665 baris karena hasil dari proses cleaning data, nilai pada dataset yang awalnya bertipe kategorikal seperti kolom Gender, Ever\_Married, Graduated, dan lainnya diganti menjadi data numerikal dengan rentang nilai 0 – n, serta jumlah kolom pada dataset bertambah yang awalnya 11 menjadi 24 kolom karena hasil preprocessing data kategori.

	Gender	Ever_Married	Age	Graduated	Work_Experience	Spending_Score	Family_Size	Segmentation	Profession_Artist	Profession_Doctor	...	Profe
0	1	0	22	0	1.0	2	4.0	3	0	0	...	
2	0	1	67	1	1.0	2	1.0	1	0	0	...	
3	1	1	67	1	0.0	1	2.0	1	0	0	...	
5	1	1	56	0	0.0	0	2.0	2	1	0	...	
6	1	0	32	1	1.0	2	3.0	2	0	0	...	
...	...	...	...	...	...	...	...	...	...	...	...	...
8062	1	1	41	1	0.0	1	5.0	1	1	0	...	
8064	1	0	35	0	3.0	2	4.0	3	0	0	...	
8065	0	0	33	1	1.0	2	1.0	3	0	0	...	
8066	0	0	27	1	1.0	2	4.0	1	0	0	...	
8067	1	1	37	1	0.0	0	3.0	1	0	0	...	

6665 rows x 24 columns

**Gambar 8.** Hasil Preprocessing Data

Setelah melewati tahap preprocessing hingga klasifikasi, dilakukan evaluasi terhadap performa dari model yang telah dikembangkan. Dalam penelitian ini, digunakan confusion matrix dalam melakukan evaluasi terhadap performa dari model. Hasil evaluasi yang akan ditampilkan mencakup accuracy, precision, recall, dan f1-score pada basis per kelasnya. Hasil evaluasi ditampilkan dalam bentuk classification report yang yang mencakup accuracy, precision, recall, dan f1-score pada basis per kelasnya.

```

↳
           precision    recall  f1-score   support

    A         0.38       0.41       0.39         320
    B         0.30       0.32       0.31         302
    C         0.46       0.42       0.44         369
    D         0.58       0.56       0.57         342

 accuracy          0.43         1333
 macro avg         0.43         1333
 weighted avg      0.44         1333

[[130  78  43  69]
 [ 75  98  97  32]
 [ 70 105 155  39]
 [ 65  44  42 191]]
    
```

**Gambar 7.** Classification Report

Berdasarkan output dari classification report didapatkan bahwa classification model mendapat akurasi 41% saat memprediksi segmen A, 32% saat memprediksi segmen B, 42% saat memprediksi segmen

C, dan 56% saat memprediksi segmen D. Secara keseluruhan, tingkat akurasi dari classification model menggunakan algoritma Decision Tree ketika digunakan untuk melakukan segmentasi customer ke dalam 4 segmen hanya 43%.

#### 4. Kesimpulan

Pada penelitian ini dilakukan klasifikasi menggunakan Algoritma Decision Tree terhadap suatu dataset dimana dataset tersebut terdiri data training dan data testing dimana masing-masing set data memiliki 11 atribut. Keluaran yang diharapkan dengan melakukan penelitian ini yaitu untuk mengetahui tingkat akurasi algoritma Decision Tree dalam mengklasifikasikan customer menjadi beberapa segmen. Setelah melalui tahap klasifikasi didapatkan bahwa secara keseluruhan tingkat akurasi dari classification model saat digunakan untuk mengklasifikasikan dataset ini hanya 43% dengan rincian akurasi 41% untuk segmentasi A, 32% untuk segmentasi B, 42% untuk segmentasi C, dan 56% untuk segmentasi D.

#### Daftar Pustaka

- [1] TIRIS SUDRARTONO, "Pengaruh Segmentasi Pasar Terhadap Tingkat Penjualan Produk Fashion Umk," *Coopetition J. Ilm. Manaj.*, vol. 10, no. 1, pp. 53–64, 2019, doi: 10.32670/coopetition.v10i1.40.
- [2] A. Madani, A. Rahmah, F. Nurunnisa, and A. Elia, "SENTIMAS: Seminar Nasional Penelitian dan Pengabdian Masyarakat Customer Segmentation at BC HNI 2 Pekanbaru by Applying the K-Medoids Algorithm and Recency, Frequency, Monetary (RFM) Model Segmentasi Pelanggan pada BC HNI 2 Pekanbaru dengan Menerapkan Alg," pp. 179–186, 2022, [Online]. Available: <https://journal.irpi.or.id/index.php/sentimas>.
- [3] I. Pratama, A. Y. Chandra, and P. T. Presetyaningrum, "Seleksi Fitur dan Penanganan Imbalanced Data menggunakan RFECV dan ADASYN," *J. Eksplora Inform.*, vol. 11, no. 1, pp. 38–49, 2022, doi: 10.30864/eksplora.v11i1.578.
- [4] Generosa Lukhayu Pritalia, "Analisis Komparatif Algoritme Machine Learning dan Penanganan Imbalanced Data pada Klasifikasi Kualitas Air Layak Minum," *KONSTELASI Konvergensi Teknol. dan Sist. Inf.*, vol. 2, no. 1, pp. 43–55, 2022, doi: 10.24002/konstelasi.v2i1.5630.
- [5] A. Rahmawati, I. Yulianti, Y. Yuliani, N. Nurhadianto, and H. B. Novitasari, "Analisis Algoritma KNN Berbasis Feature Selection untuk Memprediksi Nasabah Pengguna Deposito Melalui Pemasaran Langsung," *Swabumi*, vol. 8, no. 1, pp. 29–36, 2020, doi: 10.31294/swabumi.v8i1.7581.
- [6] R. Vincentius, M. Mirella, A. Anasthasya, S. Lauren, and Budiarjo, "Prediksi Rating Film Pada Website Imdb Menggunakan Metode Neural Network Film Rating Prediction on Imdb Website Using Neural Network," vol. 7, no. 1, 2022.
- [7] D. Sartika and D. I. Sensuse, "Perbandingan Algoritma Klasifikasi Naive Bayes, Nearest Neighbour, dan Decision Tree pada Studi Kasus Pengambilan Keputusan Pemilihan Pola Pakaian," *J. Tek. Inform. Dan Sist. Inf.*, vol. 1, no. 2, pp. 151–161, 2017, [Online]. Available: <https://doi.org/10.35957/jatisi.v3i2.78>.
- [8] M. A. Hasanah, S. Soim, and A. S. Handayani, "Implementasi CRISP-DM Model Menggunakan Metode Decision Tree dengan Algoritma CART untuk Prediksi Curah Hujan Berpotensi Banjir," *J. Appl. Informatics Comput.*, vol. 5, no. 2, pp. 103–108, 2021, doi: 10.30871/jaic.v5i2.3200.
- [9] Laila Qadrini, Andi Seppewali, and Asra Aina, "Decision Treedan Adaboost Pada Klasifikasi Penerima Program Bantuan Sosial," *J. Inov. Penelit.*, vol. 2, no. 7, pp. 1959–1965, 2021.