

KLASIFIKASI TINGKAT PENJUALAN LAPTOP DI *E-COMMERCE* MENGUNAKAN ALGORITMA *CLASSIFICATION AND REGRESSION TREE* (CART)

Neli Erliani¹, Kris Suryowati², Maria Titah Jatipaningrum³

^{1,2,3}Jurusan Statistika, Fakultas Sains Terapan, Institut Sains & Teknologi AKPRIND

Email: nelierliani1379@gmail.com

Abstract

Along with the development of the laptop era, it has become a basic need for people in their daily activities, such as carrying out learning, teaching, working, and even shopping activities, this is what has led to an increase in buying laptops through E-commerce. This study aims to determine the characteristics of the level of laptop sales in E-commerce so that it can provide an overview to prospective buyers in determining the laptop to be purchased according to the budget provided and the desired criteria and can provide an overview to shop owners in increasing laptop sales in E-commerce. In this research, the classification method used is classification and regression tree (CART). CART is a decision tree algorithm that can be used to classify using a hierarchical or tree structure. The CART algorithm has the advantages that are in accordance with research data, namely CART can be used for classification with a large amount of data with many factors and can perform classification analysis on nominal, ordinal, and continuous response variables. The results of this study indicate that the decision tree model formed from the CART algorithm produces a depth of five, the variables that affect the level of laptop sales in E-commerce are the number of reviews, price and number of product views, the variable number of reviews becomes the root or is the most important variable on the level of laptop sales in E-commerce. The classification and regression tree (CART) algorithm produces a very high accuracy value of 95.10% so that the CART algorithm can be used to classify the level of laptop sales in E-commerce.

Keywords: *Classification, CART, C5.0, and the level of laptop sales*

Abstrak

Seiring perkembangan zaman laptop sudah menjadi kebutuhan dasar bagi masyarakat dalam kegiatan sehari-hari, seperti melakukan kegiatan belajar, mengajar, bekerja, bahkan berbelanja, hal inilah yang menyebabkan meningkatnya pembelian laptop melalui *E-commerce*. Penelitian ini bertujuan untuk mengetahui karakteristik tingkat penjualan laptop di *E-commerce* sehingga dapat memberikan gambaran kepada calon pembeli dalam menentukan laptop yang akan dibeli sesuai dengan *budget* yang disediakan dan kriteria yang diinginkan serta dapat memberikan gambaran kepada pemilik toko dalam meningkatkan penjualan laptop di *E-commerce*. Pada penelitian ini metode klasifikasi yang digunakan adalah *classification and regression tree* (CART). CART merupakan salah satu algoritma *decision tree* yang dapat digunakan untuk melakukan klasifikasi menggunakan struktur hirarki atau pohon. Algoritma CART memiliki kelebihan yang sesuai dengan data penelitian yaitu CART dapat digunakan untuk klasifikasi dengan jumlah data yang cukup besar dengan banyak faktor serta dapat melakukan analisis klasifikasi pada peubah respon baik nominal, ordinal, maupun kontinu. Hasil dari penelitian ini menunjukkan bahwa model pohon keputusan yang terbentuk dari algoritma CART menghasilkan kedalaman lima, variabel yang berpengaruh terhadap tingkat penjualan laptop di *E-commerce* yaitu jumlah ulasan, harga dan jumlah produk dilihat, variabel jumlah ulasan menjadi akar atau merupakan variabel yang paling penting terhadap tingkat penjualan laptop di *E-commerce*. Algoritma *classification and regression tree* (CART) menghasilkan nilai akurasi yang sangat tinggi yaitu sebesar 95,10% sehingga algoritma CART dapat digunakan untuk melakukan klasifikasi tingkat penjualan laptop di *E-commerce*.

Kata Kunci: Klasifikasi, CART, C5.0, dan tingkat penjualan laptop

1. Pendahuluan

Perkembangan *E-commerce* telah menjadi terobosan dalam strategi bisnis moderen yang dapat menghasilkan kenaikan pendapatan secara virtual. Seiring perkembangan zaman laptop sudah menjadi kebutuhan dasar bagi masyarakat dalam kegiatan sehari-hari, seperti melakukan kegiatan belajar, mengajar bahkan berbelanja, hal inilah yang menyebabkan meningkatnya pembelian laptop melalui *E-commerce* (Muhammad & Imam, 2021). Laptop memiliki beberapa merek dan banyak tipe, setiap merek memiliki keunggulan masing-masing serta kisaran harga yang berbeda-beda, sehingga banyak calon pembeli yang bingung dalam menentukan laptop yang akan dibeli sesuai dengan *budget* yang disediakan serta kriteria yang diinginkan, maka dari itu perlu dilakukan klasifikasi tingkat penjualan laptop di *E-commerce* untuk mengetahui karakteristik tingkat penjualan laptop di *E-commerce* agar dapat memberikan gambaran kepada calon pembeli dalam menentukan laptop yang akan dibeli sesuai dengan *budget* yang disediakan dan kriteria yang diinginkan serta dapat memberikan gambaran kepada pemilik toko dalam meningkatkan penjualan laptop di *E-commerce*.

Data yang digunakan pada penelitian ini adalah data tingkat penjualan laptop dari berbagai toko yang ada di *E-commerce*, baik tingkat penjualan laptop baru maupun laptop bekas, data ini diperoleh dengan cara melakukan *Scraping data* penjualan laptop pada salah satu *E-commerce* yaitu Tokopedia. Beberapa faktor yang dapat digunakan dalam mengklasifikasikan tingkat penjualan laptop di *E-commerce* adalah harga laptop, jumlah ulasan, jumlah produk dilihat dan *rating*. Tingkat penjualan laptop dalam penelitian ini yaitu pengukuran volume penjualan yang diukur berdasarkan unit produk yang terjual. Harga laptop yaitu nilai jual laptop yang telah ditetapkan oleh pemilik toko, jumlah ulasan yaitu jumlah penilaian yang diberikan pembeli terhadap pemilik toko baik dari segi kepuasan terhadap produk (laptop) maupun pelayanan, jumlah produk dilihat yaitu berapa kali produk (laptop) dilihat. *Rating* atau skor performa toko merupakan rata-rata nilai kepuasan yang pembeli berikan terhadap produk dan pelayanan toko.

Metode yang digunakan untuk melakukan klasifikasi pada penelitian ini adalah *decision tree*. Menurut Septiyan, Jonson, & Susilo (2019), *Decision tree* merupakan metode yang dapat digunakan dalam pengambilan keputusan yang menggunakan struktur hirarki atau pohon dengan membentuk cabang-cabang dari setiap keputusan. Algoritma *classification and regression tree* merupakan salah satu metode yang dapat digunakan untuk melakukan klasifikasi. Algoritma *classification and regression tree* (CART) memiliki kelebihan yang sesuai dengan data penelitian yaitu CART dapat melakukan klasifikasi jumlah data yang cukup besar, dengan banyak faktor serta dapat dikembangkan untuk melakukan analisis klasifikasi pada peubah respon baik yang nominal maupun ordinal (Darmawan, Yuniarti, & Nasution, 2019).

Pada penelitian ini akan dibentuk model pohon keputusan menggunakan Algoritma *classification and regression tree* (CART) yang dapat menggali informasi dan pola data tingkat penjualan laptop di *E-commerce*. Diharapkan keluaran yang dihasilkan dapat memberikan gambaran kepada calon pembeli dalam menentukan laptop yang akan dibeli sesuai dengan *budget* yang disediakan dan kriteria yang diinginkan serta dapat memberikan gambaran kepada pemilik toko dalam meningkatkan penjualan laptop di *E-commerce*. Selain itu dapat mengetahui tingkat keakuratan dari metode yang digunakan yaitu algoritma *classification and regression tree*.

2. Metode

Penelitian ini menggunakan jenis desain penelitian *predictive*, yaitu menggambarkan dan mendeskripsikan data tingkat penjualan laptop di *E-commerce* serta memprediksi hasil klasifikasinya. Metode pengumpulan data yang dilakukan pada penelitian ini adalah menggunakan data primer yang diperoleh dengan cara melakukan *scraping data* penjualan laptop pada salah satu *E-commerce* yaitu

Tokopedia (<https://www.tokopedia.com>.) pada bulan januari 2023. Populasinya adalah toko yang ada di tokopedia, sampelnya menggunakan 811 toko di Tokopedia. Sampelnya diambil secara sampling acak sederhana.

Metode yang digunakan dalam penelitian ini adalah *classification and regression tree* (CART). CART merupakan salah satu metode atau algoritma dari salah satu teknik eksplorasi data yaitu teknik pohon keputusan. CART bertujuan untuk mendapatkan suatu kelompok data yang akurat sebagai penciri dari suatu pengklasifikasian, selain itu CART digunakan untuk menggambarkan hubungan antara variabel respon dengan satu atau lebih variabel prediktor (Pratiwi & Ismaini, 2019). Berikut adalah tahapan klasifikasi dari algoritma CART.

Menurut Breiman (1993) klasifikasi algoritma CART dilakukan dengan tahapan sebagai berikut:

- 1) Membagi data menjadi *training* dan *testing*.
- 2) Membentuk pohon klasifikasi maksimal dilakukan dengan langkah-langkah sebagai berikut:

a. Pemilihan pemilah

Pemilihan pemilah bertujuan untuk memperoleh pemilah yang dapat menghasilkan simpul dengan tingkat kehomogenan nilai peubah respon tertinggi. Aturan pemilihan pemilah menggunakan kriteria *goodness of split* dengan fungsi indeks Gini menggunakan persamaan berikut:

$$i(t) = 1 - \sum_{j=1}^N p^2(j|t) \quad (1)$$

$i(t)$: indeks Gini dari simpul t

$p(j|t)$: proporsi kelas j pada simpul t dimana $j = 1, 2, 3, \dots, N$ dengan $p(j|t) = \frac{N(j|t)}{N(t)}$

dimana $N(j|t)$ adalah banyaknya pengamatan kelas j pada simpul t

$N(t)$: banyak pengamatan pada simpul t

$$\phi(s, t) = \Delta i(s, t) = i(t) - p_L i(t_L) - p_R i(t_R) \quad (2)$$

Keterangan:

$\Delta i(s, t)$: *goodness of split*

$i(t)$: indeks Gini simpul t

p_L : proporsi pengamatan dari simpul t menuju simpul kiri

p_R : proporsi pengamatan dari simpul t menuju simpul kanan

$i(t_L)$: indeks Gini pada simpul anak kiri

$i(t_R)$: indeks Gini pada simpul anak kanan

b. Penentuan simpul terminal

Pohon akan berhenti dikembangkan apabila pada simpul terdapat pengamatan berjumlah $n \leq 5$. Proses pembentukan pohon juga berhenti apabila telah mencapai batasan level yang telah ditentukan atau tingkat kedalaman (*depth*) pohon mencapai maksimal.

c. Penandaan label kelas

Penentuan label kelas merupakan proses identifikasi tiap node pada kelas tertentu. Penentuan label kelas pada simpul terminal dilakukan berdasarkan aturan jumlah terbanyak, yaitu:

$$P(j|t) = \max_j \frac{N(j|t)}{N(t)} \quad (3)$$

Keterangan:

$P(j|t)$: proporsi kelas j pada simpul t

$N(j|t)$: jumlah pengamatan kelas j pada simpul t

$N(t)$: jumlah pengamatan pada simpul t

d. Menghitung ketepatan klasifikasi menggunakan persamaan berikut:

$$\text{Sensitivity} = \frac{n_{11}}{N_1} \quad (4)$$

$$\text{Specificity} = \frac{n_{22}}{N_2} \quad (5)$$

$$\text{Akurasi} = \frac{n_{11} + n_{22}}{N} \quad (6)$$

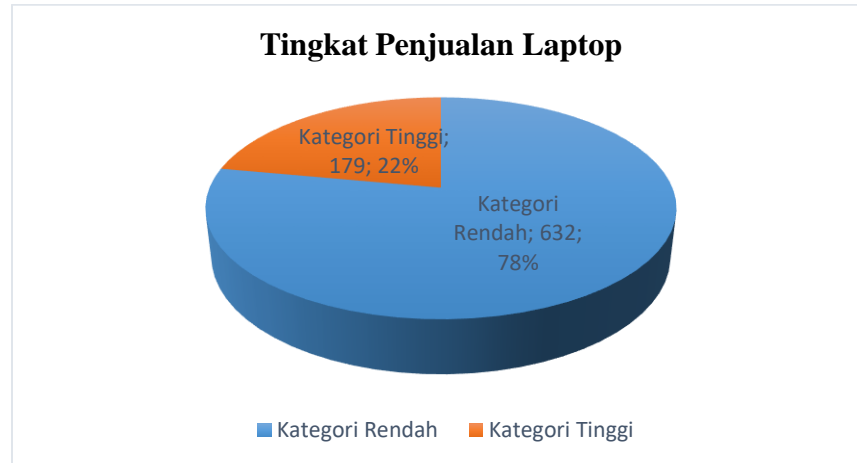
3. Hasil dan Pembahasan

Dalam melakukan klasifikasi data perlu dibagi menjadi dua yaitu *data training* dan *data testing*. Data training akan digunakan untuk melatih algoritma dalam membangun dan mencari model yang sesuai, sedangkan data testing akan digunakan untuk mengukur keakuratan atau mengetahui performa model yang didapatkan. Pembagian *data training* dan *data testing* pada penelitian ini menggunakan proporsi 50:50, 60:40, 70:30, 75:25, 80:20 dan 90:10. Diperoleh tingkat akurasi paling baik menggunakan proporsi 80:20, sehingga pada penelitian ini akan menggunakan proporsi 80:20.

Tabel 1. Pembagian Data *Training* dan *Testing*

Proporsi	Data <i>training:testing</i>	Nilai akurasi
50:50	406:405	95,80
60:40	486:325	96,70
70:30	567:244	97,17
75:25	608:203	97,36
80:20	648:163	97,53
90:10	729:82	96,84

Karakteristik tingkat penjualan laptop di *E-commerce* terbagi menjadi dua kategori yaitu rendah dan tinggi. Diperoleh tingkat penjualan laptop dengan kategori rendah sebanyak 632 atau sebesar 78% dan kategori tinggi sebanyak 179 atau sebesar 22% dengan jumlah total sebanyak 811 data.



Gambar 1. Pohon Keputusan Algoritma CART

Tabel 2. Deskriptif Variabel

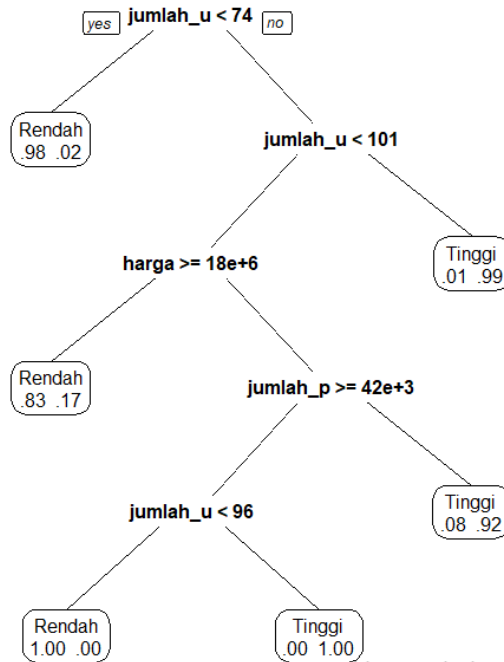
No	Variabel	N	Mean	Minimum	Maksimum
1	Harga	811	8.037.602	1.220.000	38.999.000
2	Jumlah Ulasan	811	73.19	0	2.092
3	Jumlah yang melihat	811	23.632	79	674.368
4	<i>Rating</i>	811	4,6	0	5

Berdasarkan tabel 1 diperoleh jumlah data setiap variabel adalah 811. Untuk penjelasan terperinci mengenai analisis deskriptif setiap variabel sebagai berikut:

- 1) Harga laptop, yaitu nilai jual laptop yang telah ditetapkan oleh pemilik toko. Variabel harga memiliki rata-rata sebesar Rp. 8.037.602, harga tertinggi sebesar Rp. 38.999.000 yaitu laptop Apple Macbook Promax 14 di toko Applewatchstuff dan harga terendah sebesar Rp. 1.220.000 yaitu laptop Dell 3180 Intel Inside di toko King's Macbook (laptop bekas).
- 2) Jumlah ulasan, yaitu jumlah penilaian yang diberikan pembeli terhadap toko baik dari segi kepuasan terhadap produk (laptop) maupun pelayanan toko. Variabel jumlah ulasan memiliki rata-rata sebesar 73.19, jumlah ulasan tertinggi sebesar 2.092 dan jumlah ulasan terendah sebesar 0.
- 3) Jumlah produk laptop dilihat, yaitu berapa kali produk (laptop) dilihat. Variabel jumlah produk dilihat memiliki rata-rata sebesar 23.632, jumlah tertinggi sebesar 674.368 dan jumlah terendah sebesar 79.
- 4) *Rating* atau skor performa toko merupakan rata-rata nilai kepuasan yang pembeli berikan terhadap produk dan pelayanan toko. Variabel *rating* memiliki rata-rata sebesar 4,6, *rating* tertinggi sebesar 5 dan terendah sebesar 0.

Klasifikasi Tingkat Penjualan Laptop di *E-commerce* Menggunakan Algoritma *classification and regression tree* (CART).

Berikut adalah pohon keputusan yang terbentuk menggunakan algoritma CART:



Gambar 2. Pohon Keputusan Algoritma CART

Hasil dari model pohon keputusan algoritma CART dengan pembagian data 80%:20% adalah sebagai berikut:

- 1) Terdapat tiga variabel yang berpengaruh dalam tingkat penjualan laptop di *E-commerce* yaitu jumlah ulasan, harga dan jumlah produk dilihat. Terdapat satu variabel yang kurang berpengaruh yaitu *rating*.
- 2) Variabel jumlah ulasan menjadi akar atau merupakan variabel paling penting dalam pembentukan pohon.
- 3) Terdapat 6 *leaf nodes* yang mengidentifikasi:
 - a. Tingkat penjualan laptop di *E-commerce* dikategorikan rendah jika jumlah ulasan < 74. Beberapa contoh produk laptop yang termasuk ke dalam klasifikasi ini antara lain: Acer Aspire 5 Slim A514 di Toko Alienware Official, Acer Aspire 3 Slim A314 di Toko Paradise Store Indonesia, Macbook Air Retina 13 di Toko Wawit, Apple Macbook Pro 14 Inch di Toko Onestop Gaming, Asus Vivobook Pro 14x di Toko Protech Computer, Asus Vivobook 15 Touch di Toko Intel Store ID, Dell Inspiron 3501 di Toko Dell Authorized Store Yolori, Dell Vostro 3515 di Toko AMD Official Store, Hp 14S Ryzen 3 di Toko Home Of Technology ID, Hp 240 G9 Intel di Toko Digital Zone dan Lenovo Thinkpad L380 Yoga 2 di Toko Ocean Laptop.
 - b. Tingkat penjualan laptop di *E-commerce* dikategorikan tinggi jika jumlah ulasan \geq 101. Beberapa contoh produk laptop yang termasuk ke dalam klasifikasi ini antara lain: Acer Aspire 5 Slim A514-54 di Toko Collins Official, Acer Aspire 3 A314-22 di Toko Protech Computer, Apple Macbook Air13 inch di Toko Studio Ponsel, Macbook Air 13 M1 Chip di Toko Putra Group, Asus Vivobook A416jao di Toko Gateway Indonesia Comp, Asus Vivobook Pro 14X di Toko Collins Official, Dell Inspiron 15 3511 di Toko Top Tech, Dell Vostro 14 V3401 Core I3 di Toko Rogs Store, HP 14s-Dq0508tu Celeron N4120 di Toko Sinarmulia Sukses Makmur, HP

- 14s Dq0508tu Celeron N4120 di Toko Distributor Laptop ID dan Lenovo IdeaPad Slim 3i i3 di Toko Lenovo Authorized Tangerang.
- c. Tingkat penjualan laptop di *E-commerce* dikategorikan rendah jika jumlah ulasan ≥ 74 dan jumlah ulasan < 101 , harga \geq Rp. 18.000.000. Beberapa contoh produk laptop yang termasuk ke dalam klasifikasi ini antara lain: Apple Macbook Pro 16 Inch M1 di Toko winselindo dan Apple Macbook Pro M2 Chip 13 di Toko Big Berry Cellular.
 - d. Tingkat penjualan laptop di *E-commerce* dikategorikan tinggi jika jumlah ulasan ≥ 74 dan jumlah ulasan < 101 , harga $<$ Rp. 18.000.000, jumlah produk dilihat < 42.000 . Beberapa contoh produk laptop yang termasuk ke dalam klasifikasi ini antara lain: Acer Aspire 5 Slim A514 I3 di Toko royalltech, MacBook Air Apple M1 di Toko Onestop Gaming, Asus Vivobook 14x M1403qa di Toko Protech Computer, Dell Latitude 7390 Core i7 di Toko Brankas Laptop, HP 14s-fq1036AU di Toko Distributor Laptop ID dan Lenovo Ideapad Slim 3 intel i3 di Toko Distributor Laptop ID.
 - e. Tingkat penjualan laptop di *E-commerce* dikategorikan rendah jika jumlah ulasan ≥ 74 dan jumlah ulasan < 101 , harga $<$ Rp. 18.000.000, jumlah produk dilihat ≥ 42.000 dan jumlah ulasan < 96 . Beberapa contoh produk laptop yang termasuk ke dalam klasifikasi ini antara lain: Apple MacBook Air M1 Chip di Toko winselindo dan Macbook Air 13 inch di Toko Prxloved Store.
 - f. Tingkat penjualan laptop di *E-commerce* dikategorikan tinggi jika jumlah ulasan ≥ 74 dan jumlah ulasan < 101 , harga $<$ Rp. 18.000.000, jumlah produk dilihat ≥ 42.000 dan jumlah ulasan ≥ 96 . Produk laptop yang termasuk ke dalam klasifikasi ini adalah Lenovo Ideapad Slim 5 Pro di Toko Top Tech.
- 4) Urutan variabel yang paling penting terhadap tingkat penjualan laptop di *E-commerce* yaitu: variabel jumlah ulasan, jumlah produk dilihat, harga.

Tabel 3. Kesalahan Klasifikasi Algoritma CART

Aktual	Prediksi		Total
	Rendah	Tinggi	
Rendah	128	5	133
Tinggi	3	27	30
Total	133	30	163

$$Sensitivity = \frac{128}{133} \times 100\% = 96,24\%$$

$$Specivity = \frac{27}{30} \times 100\% = 90\%$$

$$Akurasi = \frac{128+27}{163} \times 100\% = 95,10\%$$

Ukuran tingkat ketepatan klasifikasi model pohon keputusan sebesar 95,10% menunjukkan bahwa model dapat memprediksi klasifikasi dengan akurasi yang sangat tinggi. Nilai *sensitivity* yaitu nilai yang digunakan untuk mengukur klasifikasi pada kelas tingkat penjualan laptop rendah adalah sebesar 96,24%. Nilai *specificity* yaitu nilai yang digunakan untuk mengukur klasifikasi pada kelas tingkat penjualan laptop tinggi adalah sebesar 90%.

4. Kesimpulan

Berdasarkan hasil analisis yang dilakukan diperoleh model pohon keputusan yang terbentuk dari algoritma CART menghasilkan kedalaman lima, variabel yang berpengaruh terhadap tingkat penjualan laptop di *E-commerce* yaitu jumlah ulasan, harga dan jumlah produk dilihat, variabel jumlah ulasan menjadi akar atau merupakan variabel yang paling penting terhadap tingkat penjualan laptop di *E-commerce*. Algoritma *classification and regression tree* (CART) menghasilkan nilai akurasi yang sangat tinggi yaitu sebesar 95,10% sehingga algoritma CART dapat digunakan untuk melakukan klasifikasi tingkat penjualan laptop di *E-commerce*.

Daftar Pustaka

- Pratiwi, R., Hayati, M. N., & Prangga, S. (2020). Perbandingan Klasifikasi Algoritma C5.0 dengan Classification and Regression Tree . *Jurnal Ilmu Matematika dan Terapan* , 267 – 278 .
- M. S., & I. S. (2021). Sistem Pendukung Keputusan Pemilihan Laptop Untuk Kebutuhan Kuliah Metode Simple Additive Weighting (SAW). *Jurnal ICTEE*, 28-34.
- Darmawan, H. D., Yuniarti, D., & Nasution, Y. N. (2017). Klasifikasi Lama Masa Studi Mahasiswa Menggunakan Perbandingan Metode Algoritma C.45 dan Algoritma Classification and Regression Tree. *Jurnal Eksponensial*, 151-160.
- Ayani, D. D., Pratiwi , H. S., & H. M. (2019). Implementasi Web Scraping untuk Pengambilan Data pada Situs Marketplace. *Jurnal Sistem dan teknologi Informasi* , 257-262.
- Tanjung, R. H., & Kartiko. (2018). Penerapan Metode CART (Classification and Regression Tree) untuk Menentukan Faktor-faktor yang Mempengaruhi Pembayaran Kredit oleh Nasabah. *Jurnal Statistika Industri dan Komputasi* , 78-83.
- Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1993). *Classification and Regression Tree*. New York: Chapman and Hall.
- H. S. (2020). Data Mining untuk Memprediksi Prestasi Siswa. *Jurnal Pendidikan Vokasi*, 222-231.ja
- Putra, F. D. (2020). Pengaruh Volume Penjualan dan Biaya Produksi Kalung. *Jurnal Pendidikan Ekonomi Undiksha*, 462-473.
- S. F., J. G., & Susilo, S. B. (2019). Perbandingan Klasifikasi SVM dan Decision Tree untuk Pemetaan Mangrove. *Jurnal Pengelolaan Sumber Daya Alam dan Lingkungan*, 746-757.