

# Komparasi *Multiple Linear Regression* dan *Decision Tree* dalam Memprediksi Penetasan Penyu Jenis *Chelonioidea Sp* di Pulau Mangkai

Agustriono<sup>1</sup>, Rapindra Septia<sup>2</sup>, Rahmaddeni<sup>3\*</sup>

<sup>1,2,3</sup>Teknik Informatika, Prodi Teknik Informatika, STMIK Amik Riau

<sup>1</sup>agustriono.neckoe@gmail.com, <sup>2</sup>rapindrasedia@gmail.com, <sup>3</sup>rahmaddeni@sar.ac.id\*

## Abstract

Mangkai Island is located in Anambas Islands Regency, Riau Islands Province. Geographically, Mangkai Island is located at the coordinates 03005'32' LU and 105035'00" BT with an area of + 2.27 km. Sea turtles are often found on the northern coast of Mangkai Island, which is a turtle nesting site. Conservation is one of the steps to suppress the decline in sea turtle populations, integrated area management while taking into account the ecology and ecosystems, and working together the existence of endangered species, communities around the area through special interest ecotourism mechanisms. The role of machine learning is needed to analyse the hatching time of sea turtle eggs in the management of protected areas. The aim is to find the best algorithm in predicting the length of time required in hatching turtle eggs to see the availability of turtle nest relocation and become a turtle watching tour package. The modelling multiple linear regression algorithm obtained an RMSE value of 3.96387 on training data and 4.95446 on test data, while the decision tree algorithm obtained an RMSE value of 4.29728 on training data and 4.82765 on test data. Testing on actual data for prediction models on Multiple Linear Regression algorithm and Decision Tree algorithm with nest depth = 47, number of eggs = 173 eggs, turtle species = hawksbill, laying date = 27 April 2022. The Multiple Linear Regression algorithm predicts that the eggs will hatch in 47.99 days, rounded to 48 days, while the Decision Tree algorithm predicts that the eggs will hatch in 48 days.

**Keywords:** multiple linear regression, decision tree, machine learning, ecotourism, sangkai island.

## Abstrak

Pulau Mangkai yang terletak di Kabupaten Kepulauan Anambas Provinsi Kepulauan Riau. Secara geografis Pulau Mangkai terletak pada titik koordinat 03005'32' LU dan 105035'00"BT dengan luas + 2,27 km. Pantai bagian utara di Pulau Mangkai merupakan tempat penetasan penyu. Konservasi salah satu langkah untuk menekan menurunnya populasi penyu, pengelolaan kawasan yang terintegrasi dengan tetap mempertimbangkan ekologi dan ekosistem serta mengkolaborasi keberadaan spesies yang terancam punah, masyarakat di sekitar kawasan melalui mekanisme ekowisata minat khusus. Peran serta *Machine Learning* diperlukan dalam menganalisis lama penetasan telur penyu pada pengelolaan kawasan konservasi. Tujuannya adalah untuk mencari algoritma yang terbaik dalam memprediksi lama waktu yang dibutuhkan dalam penetasan telur penyu untuk melihat ketersediaan relokasi sarang penyu dan menjadi paket wisata *Turtle Watching*. Pemodelan algoritma *Multiple Linear Regression* diperoleh nilai RMSE 3,96387 pada data *training* dan 4,95446 pada data *testing*, sementara pada Algoritma *Decision Tree* nilai RMSE pada data *training* 4,29728 dan 4,82765 pada data *testing*. Pengujian pada data aktual untuk model prediksi pada algoritma *Multiple Linear Regression* dan algoritma *Decision Tree* dengan kedalaman sarang = 47, jumlah telur = 173 butir, jenis penyu = sisik, tanggal bertelur = 27 April 2022. Algoritma *Multiple Linear Regression* memprediksi telur akan menetas selama 47,99 dibulatkan menjadi 48 hari, sementara Algoritma *Decision Tree* memprediksi telur akan menetas selama 48 hari.

**Kata kunci:** regresi linier berganda, *decision tree*, *machine learning*, ekowisata, pulau Mangkai

©This work is licensed under a Creative Commons Attribution - ShareAlike 4.0 International License

## 1. Pendahuluan

Indonesia Negara Kepulauan yang sebagian wilayahnya air, hampir 2/3 wilayah Indonesia adalah perairan. Indonesia memiliki ribuan pulau terbentang dari Sabang hingga Marauke. Salah satu Pulau di dalamnya adalah Pulau Mangkai yang terletak di Kabupaten Kepulauan Anambas Provinsi Kepulauan Riau. Secara geografis Pulau Mangkai terletak pada titik koordinat 03<sup>00</sup>5'32' LU dan 105<sup>03</sup>5'00"BT dengan luas ± 2,27 km. Kondisi perairan dengan tingkat kecerahan 5 – 10 meter memiliki parameter kualitas air dengan suhu 31<sup>0</sup>C, pH 7,8, dengan salinitas 31 ‰, *Dissolved Oxygen* (DO) 12,4 mg/l, *Biological Oxygen Demand* (BOD) 6,40 mg/l, *Amonia* (NH<sub>3</sub>) 0,035 mg/l, *Nitrat* (NO<sub>3</sub>) 0,088 mg/l, *Nitrit* (NO<sub>2</sub>) 0,015 mg/l dan *Sulfida* (S<sub>2</sub>) < 0,01 mg/l. Vegetasi Pantai yang tumbuh

di Pulau Mangkai meliputi kelapa, Semak, rumput dan tanaman perdu, tekstur pulau berbatuan dan pasir putih di Pantai [1]. Biota yang sering di jumpai di Pulau Mangkai adalah Penyu, hal ini dikarenakan pantai bagian utara yang berada di Pulau Mangkai merupakan tempat penetasan penyu. Tingkat kemunculan penyu di pulau Mangkai 73-100% terjadi pada bulan April-Juni dengan jumlah 619 ekor yang terdiri dari 571 ekor penyu hijau (*Chelonia mydas*) dan 48 ekor penyu sisik (*Eretmochelys imbricata*). Penyu merupakan hewan reptile yang hidupnya hampir 2/3 nya adalah berada di air, namun untuk bertelur hewan ini membutuhkan pasir putih di tepian Pantai [2].

Pertumbuhan manusia yang pesat menyebabkan banyaknya pergeseran fungsi lahan, khususnya lahan pesisir pantai menjadi wilayah pemukiman dan

pembangunan bahkan pemanfaatan alam sebagai objek wisata yang tidak ramah lingkungan mengakibatkan habitat alami penyu mengalami gangguan. Gangguan lain juga dapat berasal dari salah penangkapan ikan (salah tangkap), eksploitasi telur penyu di pasar gelap, penyakit penyu dan pengaruh perubahan iklim dunia. Hal ini disebabkan karena siklus hidup penyu yang panjang untuk menghasilkan penyu yang siap bertelur, waktu yang dibutuhkan yakni 30-40 tahun. Oleh karena itu diperlukan [3]. Penyu merupakan salah satu biota langka yang keberadaannya dilindungi skala nasional dan internasional, hal ini dikarenakan populasi penyu mengalami tingkat penurunan yang signifikan setiap tahunnya. Biota penyu terancam punah keberadaannya di dunia yang disebabkan oleh faktor internal dan faktor eksternal [4].

Untuk menekan laju penurunan populasi penyu, langkah yang dapat dilakukan adalah dengan konservasi dan habitat penyu itu sendiri. Konservasi jenis merupakan salah satu cara menekan tingkat penurunan populasi biota. Pengelolaan kawasan yang terintegrasi dengan tetap mempertimbangkan ekologi dan ekosistem serta mengkolaborasi keberadaan spesies yang terancam punah, masyarakat di sekitar kawasan melalui mekanisme ekowisata minat khusus. Ekowisata merupakan salah satu wisata yang lebih mengutamakan keberlangsungan dan keseimbangan ekosistem untuk kelestarian alam serta memberikan manfaat secara finansial tanpa mengurangi keutuhan suatu ekosistem termasuk didalamnya kebudayaan masyarakat tempatan [5]. Konservasi juga merupakan salah satu teknik untuk mengurangi dampak penurunan populasi biota tertentu oleh tindakan yang merugikan [6].

Salah satu pemanfaatan sumberdaya alam yang tetap mempertahankan ekosistem dan ekologi adalah dengan ekowisata *turtle watching* atau pengamatan penyu bertelur dan pelepasan tukik (anak penyu). Pariwisata adalah pemanfaatan keunikan sumberdaya alam pada suatu daerah dan dapat meningkatkan data tarik orang luar untuk berkunjung serta dapat menambah penghasilan bagi masyarakat tempatan [7]. Kegiatan ini perlu dijalankan sebagai langkah pengendalian dan mengurangi akibat dari perubahan dan peralihan fungsi lahan serta masifnya pemahaman orang terhadap pelestarian biota penyu [8].

Pengamatan dan penggabungan serta peran serta teknologi diperlukan dalam menganalisis tingkat keberhasilan penetasan dan lama penetasan telur penyu pada pengelolaan kawasan konservasi. Tujuannya adalah untuk memprediksi lama waktu yang dibutuhkan dalam penetasan telur penyu. Hal ini diperlukan untuk melihat ketersediaan lahan untuk relokasi sarang penyu, selanjutnya hal ini dapat dimanfaatkan bagi operator wisata untuk dijadikan paket wisata atraksi wisata *Turtle watching* kepada wisatawan dengan tetap menjalankan prosedur yang berlaku [6]. Pendekatan yang dilakukan dalam penelitian ini adalah memprediksi lama waktu yang

dibutuhkan untuk penyu menetas menggunakan pendekatan *Machine Learning* dengan menggunakan *Algoritma Multiple Linear Regression* dan *Decission Tree*.

*Machine Learning* merupakan salah satu metode pendekatan teknologi berbasis komputer yang dapat digunakan dalam memprediksi dari sebuah data set [9]. *Machine Learning* dapat di implementasikan dalam memprediksi lama waktu yang diperlukan untuk penetasan telur penyu dengan menggunakan *Algoritma Multiple Linear Regression* dan *Decission Tree*.

Analisa Regresi Linier Berganda atau *Multiple Linear Regression* digunakan untuk menghitung lebih dari satu variabel independen. Hal ini digunakan untuk menentukan arah dan pengaruh variabel independen terhadap variabel dependen [10]. Penelitian terdahulu dalam pengimplementasian *Algoritma Multiple Regression* atau Regresi Linier Berganda dalam memprediksi kunjungan pasien di rumah sakit berdasarkan pelayanan dan umur pasien menghasilkan akurasi sebesar 80% [11]. Penelitian lain dalam implementasi Regresi Linier Berganda dalam memprediksi jumlah pelanggan dan persediaan barang pada Bali Orchid, bahwa Regresi Linier Berganda dapat dijadikan salah satu metoda dalam dasar untuk mengambil kesimpulan pada Bali Orchid dengan nilai *Mean Absolute Deviation* (MAD) 0,00002, dengan kata lain tingkat error kecil sehingga Regresi Linier Berganda sangat baik untuk memprediksi [12]. Penelitian lain yang mengimplementasikan Regresi Linier Berganda untuk pengambilan keputusan dalam hubungan dua variabel *dependent* kedalaman sarang, lama inkubasi terhadap tingkat penetasan [8]. Regresi linier ganda merupakan cara yang yang dapat digunakan dalam mengetahui keterkaitan antara dua buah variabel atau lebih yakni variabel *dependent* dan variabel *independent* [13]. Penelitian ini menguji variabel *dependent* tanggal bertelur, tanggal menetas, kedalaman sarang, jumlah telur dan jenis penyu terhadap variabel *independent* yakni lama menetas.

*Algoritma Decission Tree* dalam memprediksi persediaan obat dari data set pada Apotek FranchFarma sebanyak 50 data dengan variabel nama barang, jenis obat, persediaan, stok awal, stok akhir mendapat hasil prediksi sebanyak 26 jenis obat yang harus tersedia dalam satu siklus dengan keakurasian 0,983 [14]. *Algoritma Decission Tree* juga telah diimplementasikan pada penelitian untuk mengetahui hubungan variabel-variabel dalam prediksi kesuksesan pendidikan karakter dengan nilai akurasi 60,91%, dan menunjukkan bahwa *Algoritma* ini layak diimplementasikan pada prediksi kesuksesan pada pendidikan karakter [15]. *Decission Tree* algoritma yang merujuk kepada pohon keputusan untuk mengambil kesimpulan dari analisis faktor-faktor keputusan pembelian konsumen, penerapan algoritma *Decission Tree* mempunyai kelebihan dalam menganalisa dan mempunyai tingkat akurasi 83,33%

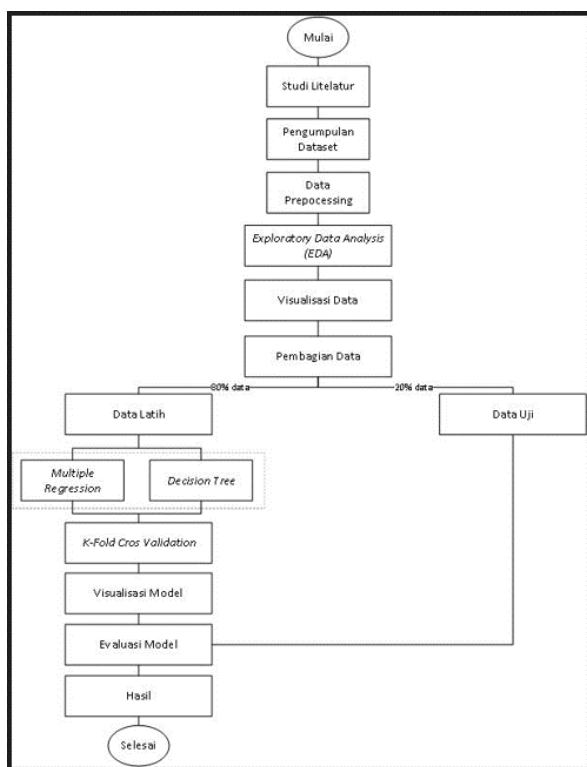
dalam analisis pembelian program pada perkumpulan pengiat programmer Indonesia [16].

Berdasarkan pertimbangan di atas maka dianggap perlu untuk melakukan penelitian dibidang perikanan dan kelautan melibatkan *Machine Learning* yang menggunakan *Algoritma Multiple Linear Regression* dan *Decision Tree* untuk memprediksi penetasan penyu. Tujuannya adalah untuk melihat kemampuan masing-masing algoritma dalam memprediksi lama waktu yang dibutuhkan dalam penetasan telur penyu dengan implementasi *Machine Learning*.

## 2. Metode Penelitian

### 2.1. Tahapan Penelitian

Tahapan penelitian yang dilakukan pada penelitian ini melalui beberapa tahapan, tahapan tersebut meliputi mulai studi litelatur hingga evaluasi model. Secara rinci analisis data yang dilakukan terlihat pada tahapan berikut dibawah ini:



Gambar 1. Diagram Tahapan Penelitian

Secara rinci keterangan gambar dapat dilihat dibawah ini:

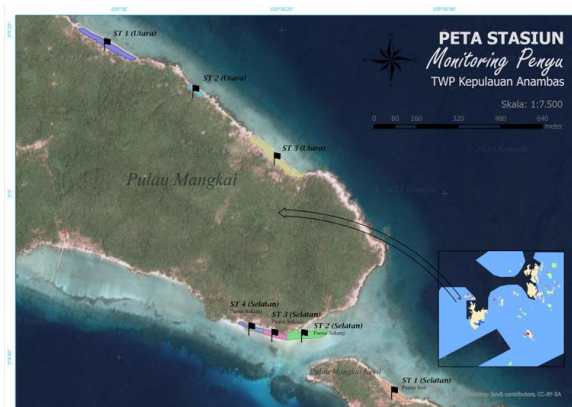
#### a. Studi litelatur.

Mengumpulkan sumber referensi bacaan dari jurnal, buku dan bahan bacaan lain yang berkaitan dengan penelitian ini. Teori-teori yang mendasari dan relevan terhadap judul yang diangkat dalam penelitian ini. Tujuan studi litelatur agar tidak terjadinya duplikasi penelitian, serta dapat memanfaatkan seluruh informasi dan pemikiran-pemikiran yang relevan.

#### b. Pengumpulan data set

Pengumpulan *dataset* atau pengambilan *dataset* yakni proses memilih sejumlah komponen dari data yang terdiri dari variabel yang akan dijadikan data olah dan diteliti lebih lanjut. Data olah ini kemudian akan dibagi menjadi data fitur (variabel x) dan target (variabel y). Tujuan pengumpulan atau pengambilan *dataset* ini untuk memperoleh keterangan dan informasi yang dibutuhkan yang mendukung dalam rangka mencapai tujuan penelitian.

Pada penelitian ini, pengumpulan *dataset* dilaksanakan pada bulan Desember 2023 sampai dengan Januari 2024 dengan menggunakan *dataset* pengamatan pendaratan penyu di Pulau Mangkai di tahun 2022. Pengambilan data sekunder berupa *dataset* dilakukan mulai bulan Januari 2022 sampai dengan bulan Desember 2022, lokasi di Pulau Mangkai, Kawasan Konservasi Perairan Nasional Kepulauan Anambas, Provinsi Kepulauan Riau. Lokasi pengamatan dapat dilihat pada gambar 2.



Gambar 2. Gambar Stasiun Pengamatan Peneluran Penyu di Pulau Mangkai [2]

#### c. Data Preprocessing

*Data Preprocessing* salah satu langkah awal yang harus dilakukan untuk mengubah data mentah menjadi format yang lebih mudah dipahami dan siap diproses untuk tahapan analisis data berikutnya. Proses ini penting karena data mentah sering kali tidak lengkap dan memiliki format yang tidak teratur. *Preprocessing* data terdiri dari perawatan data, transformasi data, dan pengurangan data. Proses ini tidak hanya meningkatkan kualitas dan keakuratan data, tetapi juga untuk memastikan bahwa data atau model dalam *Machine Learning* dapat berkerja dengan baik. *Preprocessing* data sangat penting untuk dilakukan karena kesalahan data, *redudansi* (duplikat data), nilai yang hilang, dan data yang tidak konsisten dapat mengurangi keakuratan hasil pengeolahan [17]. Pada tahapan ini dilakukan penambahan kolom yang nilainya berupa jumlah hari yang diperlukan telur penyu untuk menetas sebagai targetnya.

Peralatan yang digunakan dalam tahapan *preprocessing* data pada penelitian ini adalah Lapotp Hp dengan spesifikasi media penyimpanan *Solid State Drive* (SSD) berfungsi sebagai media penyimpanan

data sekunder berupa *dataset* pengamatan pendaratan penyu sekaligus sebagai media menjalankan program python. Core i7 gen 6, *Random Acces Memory* (RAM) 16 GB yang sudah di install python 3 dan berfungsi sebagai media pengoperasian *Jupiter Notebook*.

#### d. Exploratory Data Analysis

EDA salah satu tahapan penting dalam proses analisis data. EDA adalah proses analisis data yang bertujuan untuk memberikan pemahaman dan analisis rinci tentang data. Dengan kata lain EDA adalah proses uji investigasi awal yang bertujuan untuk mengidentifikasi pola data. EDA digunakan untuk membedakan kumpulan data dari *dataset* lain. Pada tahapan ini dengan menggunakan EDA, data yang akan diproses lebih lanjut sebanyak 8 (delapan) bulan dari 12 bulan *dataset*. Selanjutnya mengekstrak kolom tanggal bertelur menjadi 3 (tiga) bagian yang berisi tahun, bulan dan tanggal bertelur.

#### e. Visualisasi Data

Visualisasi adalah tahapan dimana proses menampilkan informasi dalam bentuk yang mudah dipahami menggunakan elemen visual grafik. Alat visualisasi data meningkatkan dan mengotomatiskan proses komunikasi visual untuk mendapatkan akurasi dan detail

#### f. Pembagian Data

Pembagian data adalah pembagian fitur dan target. Dalam hal ini fitur mencakup variabel kedalaman sarang, jumlah telur, jenis penyu dan targetnya adalah variable lama menetas. *Dataset* sebanyak 462 row dibagi menjadi 80% data *training* sejumlah 368 data fitur *training* dan 7 sebagai data *training* target, dan 20% data target dengan 92 fitur *test* dan 7 target *test*.

#### g. Modeling

Penerapan *Machine Learning* dengan *Algoritma Multiple Linear Regression* dan *Decission Tree* untuk data *training* dan data *test* pada *dataset* yang ada.

Penelitian yang dilakukan menggunakan *Machine Learning* dengan menerapkan *Agoritma Multiple Linear Regression* dan *Decission Tree* dalam memprediksi penetasan telur penyu. *Macnine Learning* merupakan cabang dari ilmu *Artificial Inteligent* (AI) yang fokus dalam pengolahan *dataset* valid dalam jumlah besar, dalam hal ini mesin diajarkan agar mampu belajar secara mandiri dari *dataset* yang ada [18]. *Algoritma Multiple Linear Regression* dan *Decision Tree* digunakan untuk memprediksi penetasan penyu berdasarkan variable kode sarang, tanggal bertelur, tanggal menetas, kedalaman sarang, jumlah telur, dan jenis penyu untuk menarik kesimpulan diakhir berdasarkan hubungan antar variable [8].

Pemodelan dengan *Multiple Linear Regression* dan *Decission Tree* dilakukan dengan penambahan fitur *K-Fold Cross Validation* dengan  $k=10$ . *K-Fold Cross Validation* adalah metoda validasi *Machine Learning*

yang digunakan untuk meningkatkan kinerja model, dan mencegah overfitting, serta memastikan bahwa model memiliki kemampuan generalisasi yang baik. Selanjutnya *dataset* di lakukan *training* dengan menggunakan model *Multiple Linear Regression* yang ditambahkan *K-Fold Cross Validation* dengan  $k=10$  untuk melihat RMSE. *Root Mean Squared Error* (RMSE) adalah salah satu metode untuk mengevaluasi model regresi linier berganda untuk menghitung besaran tingkat kesalahan dari hasil prediksi [19].

#### h. Visualisasi Modeling

Visualisasi merupakan tahapan penampilan hasil yang diperoleh dari proses penerapan modeling pada *dataset*.

### 3. Hasil dan Pembahasan

Data yang akan diolah lebih lanjut dalam bentuk file *.csv* dengan jumlah fitur awal sebelum *preprocessing* sebanyak 14 fitur yakni tanggal awal, kode sarang, tanggal bertelur, tanggal menetas, hidup, mati, gagal menetas, tanggal rilis, jam/pukul, kedalaman sarang, tanggal bertelur, tanggal menetas, jumlah telur dan jenis penyu, serta dengan 462 row set data. Data ditarik kedalam jupyter *notbook* dalam bentuk format file *.csv*.

Data setelah dilakukan *preprocessing* dalam penelitian ini sebanyak 462 *record*. *Dataset* kemudian dilakukan penghapusan kolom yang tidak digunakan, hal ini dilakukan untuk memilih fitur yang dianggap penting untuk dilakukan pengolahan lebih lanjut, selanjutnya data ditampilkan dalam bentuk informasi umum data. Hasil penghapusan kolom dan ditampilkan dalam informasi umum *dataset* dapat dilihat apda gambar dibawah ini:

```
Informasi umum data set
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 462 entries, 0 to 461

Data columns (total 6 columns):
# column          Non-Null Count  Dtype
0 Kode_Sarang      462 non-null   object
1 Tanggal_Bertelur 462 non-null   object
2 Tanggal_Menetas  462 non-null   object
3 kedalaman_sarang 462 non-null   int64
4 jumlah_telur     462 non-null   int64
5 yenis_penyu      462 non-null   object

Dtypes: int64(2), object (4)
Memory usage: 21.8+ KB
```

Dari data set sebelumnya dengan 14 kolom atau fitur dihapus menjadi 6 fitur atau kolom, selanjutnya data 6 fitur atau kolom ini yang akan dilakukan pemrosesan dengan menggunakan algoritma lebih lanjut. Data kemudian di gambarkan dengan statistis deskriptif data set. Secara deskriptif *dataset* dapat dilihat pada gambar dibawah ini :

Tabel 1. Gambaran Umum Data Set Kolom Bernilai Angka

Ket	count	mean	std	min	25 %	50 %	75 %	max
-----	-------	------	-----	-----	------	------	------	-----

kdlmn_	462	63.91	21.	18	48.	60.	80	135
srmg			07		5	0		
jmlh_te	462	90.43	23.	18	76.	90.	10	180
lur			77		0	0	1.0	

Pada tabel diatas dapat dilihat bahwa pola sebaran data serta variasi dalam *dataset* yang akan diolah. Data tersebut diatas dipanggil menggunakan fungsi `df.describe().T`. `count` ditampilkan untuk melihat data *non-null*, *mean* untuk melihat rata-rata data, *std* (standar deviasiasi), *min* untuk nilai minimum, 25% untuk melihat kuartil bawah, 50% untuk melihat data kuartil tengah dan 75% kuartil atas, nilai *max* merupakan nilai maksimum pada *dataset*. Tabel 1. Ditampilkan untuk mendapatkan gambaran tentang distribusi data dan sebaran data numerik pada objek penelitian.

Tabel 2. Gambaran Data Bertipe Kategorik

Keterangan	count	unique	Top	freq
Kode_sarang	462	292	R85	3
tanggal_bertelur	462	145	14 mey 22	13
Tanggal_menetas	462	135	31 jul 22	14
Jenis_penyu	462	2	Hijau	391

Tabel 2. diatas menjelaskan tentang rangkuman statistik deskriptif untuk kolom yang berisi data tipe objek. *Count* untuk melihat data yang ada, *unique* untuk melihat nilai yang unik, *top* untuk melihat nilai yang sering muncul, dan *freq* adalah nilai yang paling sering muncul. Dengan tabel 2. ini peneliti dapat memahami karakteristik dan variasi dalam data kategorikal. Data ini dipanggil menggunakan fungsi `df.describe(include='object').T` pada *python*.

Pada data *preprocessing* dilakukan penambahan fitur atau kolom dengan nama “lama menetas”, dan kolom ini dijadikan fitur atau kolom target untuk dilakukan proses analisis selanjutnya. Fitur atau kolom diperoleh dari pengurangan fitur tanggal menetas dikurangi dengan fitur tanggal bertelur, selanjutnya di konfersi ke jumlah hari. Secara detail dapat dilihat pada tabel dibawah ini :

Tabel 3. *Dataset* dengan Penambahan Fitur Lama Menetas

Kode saran	Tgl bertelur	Tgl menetas	Kdl man saran	Jmlh telur	Jenis penyu	Lama menetas
R136	110722	200822	91	135	Hijau	40
R259	070722	230822	45	105	Hijau	47
R193	110822	230922	78	67	Hijau	43
R14	020422	230522	30	98	Hijau	51
R77	010522	150622	57	96	Hijau	45

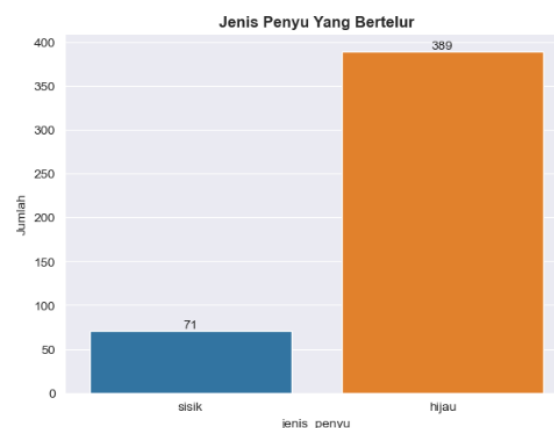
Fitur lama menetas diperoleh dari fungsi `df['lama_menetas'] = (df['tanggal_menetas'] - df['tanggal_bertelur']).dt.days` pada *python*. Fitur ini untuk melihat nilai lama atau jumlah hari yang diperlukan untuk menetas. Fitur ini dijadikan target dalam pengolahan data selanjutnya menggunakan model algoritma.

Selanjutnya data dilakukan *Exploratory Data Analysis* (EDA), hal ini dilakukan untuk menampilkan tanggal minium data dan tanggal maksimum *dataset* tanggal bertelur penyus dengan melakukan proses `df.tanggal_bertelur.min()`, `df.tanggal_bertelur.max()` sehingga diperoleh hasil sebagai berikut : `Timestamp('2022-03-11 00:00:00')`, `Timestamp('2022-10-26 00:00:00')`. Artinya *dataset* jumlah penyus bertelur di mulai pada bulan Maret 2022 sampai bulan Oktober 2023, secara visualisasi dapat di lihat pada gambar dibawah ini:



Gambar 3. Data Hasil EDA

Dari gambar 3. diatas terlihat bahwa intensitas penyus bertelur tinggi secara berturut-turut terjadi pada bulan April sampai Agustus 2022, dengan jumlah penyus bertelur sebanyak 96 ekor di bulan April, 110 ekor di bulan Mei, 60 ekor di bulan Juni 82 ekor di bulan Juli dan 70 ekor di bulan Agustus. Data ini senada dengan musim bertelur penyus yakni kisaran bulan April hingga September [3]. Selanjutnya gambaran perbandingan berdasarkan jenis penyus yang bertelur selama tahun 2022 dapat dilihat pada gambar dibawah ini:



Gambar 4. Jumlah Penyus Bertelur Berdasarkan Jenis Penyus

Pada gambar 4. diatas dapat dilihat bahwa selama bulan Maret 2022 sampai dengan Oktober 2022 jumlah penyus bertelur berdasarkan jenisnya adalah penyus sisik (*Eretmochelys imbricata*) berjumlah 71 ekor dan penyus hijau (*Chelonia mydas*) sebanyak 389 ekor. Kelimpahan populasi bertelur penyus hijau diperkirakan

rata-rata 4500-5000 ekor per tahun [3]. Berdasarkan laporan monitoring yang dilakukan Loka Kawasan Konservasi Perairan Nasioinal Pekanbaru yang dilakukan di Pulau Mangkai spesies penyu yang dominan mendarat adalah jenis penyu sisik dan penyu hijau [2].

Sebelum dilakukan pemodelan terhadap *dataset*, perlu dilakukan uji normalitas data. Hal ini untuk melihat apakah data sudah terdistribusi normal, sehingga dapat dilanjutkan ketahapan selanjutnya. Maka dari itu perlu dilakukan perlakuan untuk melihat data terdistribusi normal atau tidak, seperti gambar dibawah ini :

Tabel 4. Kodngan Pengujian Normalitas Data

```

Kodingan untuk melihat data numerik terdistribusi normal atau tidak

columns_name=['kedalaman_sarang', 'jumlah_telur']

results = []

for col in columns_name:
    stat, pvalue = st.normaltest(df[col])
    alpha = .05 #alpha 5%

    if pvalue < alpha:
        results.append({"Data":col, "p-value":pvalue, "Distribusi normal":"Tidak"})
    else:
        results.append({"Data":col, "p-value":pvalue, "Distribusi normal":"Ya"})
    
```

Dari kodingan phyton diatas diperoleh hasil bahwa p-value: 0.0002646702306004279 - Data "kedalaman\_sarang" tidak terdistribusi normal dan p-value: 4.5266328614822695e-10 - Data "jumlah\_telur" tidak terdistribusi normal. Secara lengkap dapat dilihat pada tabel dibawah ini:

Tabel 5. Hasil uji Normalitas Untuk Variabel / Kolom Kedalman\_Sarang Dan Jumlah\_Telur

No	Data	p-Value	Distribusi normal
0	Kedalaman_Sarang	2.646702e-04	Tidak
1	Jumlah_Telur	4.526633e-10	Tidak

Dikarenakan *dataset* yang akan diolah tidak bedistribusi normal maka diperlukan teknik penskalaan data dengan menggunakan *robustscaler* dan tidak menghapus *oulier* karena data tersebut sangat penting [19]. *Robustscaler* akan menghasilkan distribusi data yang lebih merata untuk data distribusi tidak normal, *robustscaler* memberikan hasil yang lebih konsisten dan tidak terpengaruh oleh nilai ekstrim serta dapat meningkatkan performa model yang akan digunakan. Penggunaan *robustscaler* dapat dilihat pada kodingan berikut:

```

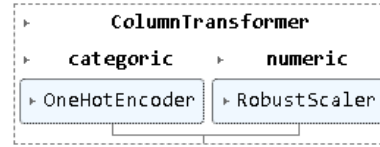
Kodingan robustscaler
Numeric_cols = features.describe().
columns.to_list()

Categoric_cols = features.describe(include=
'object').columns.to_list()
    
```

```

Preprocessor = ColumnTransformer
(transformers =[('catagoric',
OneHotEncoder(), categoric_cols), ('numeric',
RobustScaler(), numeric_cols)
    
```

Dari kodingan tersebut ditampilkan seperti gambar dibawah ini :



Gambar 5. Scema Penskalaan Menggunakan OneHoteEncoder dan RobustScaler

Dari gambar 5. tersebut diatas, dari hasil penskalaan yang dilakukan memperoleh hasil sebagai berikut:

Tabel 5. Hasil dari Penskalaan Dari Data Yang Tidak Terdistribusi Normal

No	0	1	2	3	4	5	6
0	0.0	1.0	-0.938	2.376	0.0	-1.0	-0.385
1	1.0	0.0	-0.594	0.000	0.0	-1.0	-0.154
2	1.0	0.0	-0.156	-0.713	0.0	-1.0	-0.231
3	1.0	0.0	0.000	0.475	0.0	-1.0	0.538
4	1.0	0.0	0.000	0.079	0.0	-1.0	0.615

Dari tabel 5. diatas terlihat bahwa berdasarkan hasil penskalaan menggunakan *robustscaler* sebaran data mempunyai jarak antar data tidak terlalu jauh. Interval datanya mulai dari -1 hingga 1 setiap datanya, sehingga data sudah terdistribusi normal.

Selanjutnya data dibagi dimana 80% dari data tersebut sebagai data *training* dengan rincian 368 data sebagai data *training* 20% data *testing* dengan rincian sebanyak 92 data. Variabel *Independent* atau variabel bebas nya adalah 1. kedalaman sarang, 2. jumlah telur, 3. Jenis penyu, dan variable *dependent* atau variable targetnya adalah lama menetas.

Dari model yang telah dijalankan dengan menggunakan algoritma *Multiple Linear Regression* dan *Algoritma Decision Tree* dapat dilihat pada tabel 6. dibawah ini:

Tabel 6. Hasil Uji Normalitas Untuk Variabel / Kolom Kedalman\_Sarang Dan Jumlah\_Telur

No	Rincian	Algoritma Multiple Linear Regression	Algoritma Decision Tree
1	RMSE data Train	3,96387	4,29728
3	RMSE data Test	4,95446	4,82765

Dari model algoritma *Multiple Linear Regression* yang telah dilakukan diperoleh nilai RMSE sebesar 3,96387. Nilai RMSE mendekati nilai Nol (0) artinya model sudah berkerja dengan baik pada data *training*. Selanjutnya dilakukan implementasi model pada 92 data testing, hasil dari pengujian model pada data testing diperoleh nilai RMSE sebesar 4,95446. Dengan demikian untuk model *Multiple Linear Regression* dapat berkerja dengan baik pada *dataset* monitoring

penyu dan dapat diinterpretasikan sebagai model algoritma ini memiliki tingkat kesalahan cukup rendah, artinya prediksinya secara keseluruhan cukup dekat dengan nilai aktual.

Pemodelan *Algoritma Decision Tree* pada dataset diperoleh nilai RMSE pada data *training* sebesar 4,29728. Hal ini menunjukkan bahwa rata-rata perbedaan antara nilai target aktual dan nilai prediksi yang dihasilkan oleh data *training* secara keseluruhan mampu memprediksi nilai target dengan baik. Sedangkan untuk data *test* diperoleh nilai RMSE sebesar 4,82765. Secara relatif nilai ini sedikit lebih tinggi dibandingkan dengan nilai RMSE data *training*. Hal ini mengindikasikan bahwa model *Decision Tree* memiliki kinerja sedikit lebih rendah pada data *test*.

Nilai RMSE antara *Algoritma Multiple Linear Regression* dan *Decision Tree* tidak terpaut jauh, dan masih diatas nol. Tingginya nilai *error* dapat disebabkan banyak faktor. Sehingga diperlukan penelitian lebih lanjut untuk menganalisa tingginya nilai RMSE dari kedua model algoritma tersebut. Hasil dari pemodelan *Algoritma Multiple Linear Regression* dan *Decision Tree* antara nilai aktual dan nilai prediksi dapat divisualisasikan pada tabel dibawah ini.

Tabel 7. Data Perbandingan Nilai Sebenarnya Dan Nilai Prediksi Regresi Linier Berganda Dan *Decision Tree* Untuk Data Test

No	Nilai Aktual	Nilai Prediksi <i>Decision Tree</i>	Nilai Prediksi Regresi Linier Berganda
0	47	43.0	45.3
1	46	45.0	44.4
2	44	44.0	44.6
3	46	46.0	46.4
4	46	46.0	46.2
5	48	48.0	47
6	44	45.0	44.2
7	42	43.0	45.4
8	45	46.0	44.8
9	45	46.0	43.5

Dari tabel 7. Diatas Perbandingan nilai sebenarnya atau aktual dan nilai prediksi pemodelan Regresi Linier Berganda serta algoritma *Decision Tree* tidak banyak mengalami perbedaan. Hal ini dapat dikarenakan *Missing value* pada proses training sehingga pemodelan *Algoritma Multiple Linier Regresion* tidak menghasilkan kinerja yang baik, hal ini terlihat dari nilai *R-squer score* yang dihasilkan dialgoritma regresi linier berganda dengan nilai 0.07154 sedangkan pada algoritma *Decision Tree* nilai *R-square score* 1.000 walaupun nilai RMSE sudah mendekati angka 0-1 untuk kedua model secara visualisasi perbandingan antara nilai sebenarnya dan nilai prediksi dapat dilihat pada gambar dibawah ini:



Gambar 6. Grafik Perbandingan Nilai Sebenarnya Dan Nilai Prediksi Menggunakan *Multiple Linear Regression*

Dari gambar 6. diatas terlihat antara nilai sebenarnya dengan nilai prediksi menggunakan *Multiple Linear Regression*, garis biru merupakan nilai sebenarnya dan garis oren adalah nilai prediksi. Secara umum nilai prediksi mendekati nilai sebenarnya, hal ini terlihat garis oren merupakan nilai prediksi mendektai garis biru atau nilai sebenarnya. Namun ada beberapa data yang hasil prediksinya jauh dari nilai sebenarnya, terlihat pada data yang di tandai kotak merah pada gambar 6, hal ini bisa saja disebabkan faktor kesalahan dalam pengumpulan data dilapangan dan juga perlu dilakukan fitur *engineering* pada model *Multiple Linear Regression*.



Gambar 7. Grafik perbandingan Nilai Sebenarnya Dan Nilai Prediksi Menggunakan *Decision Tree*

Dari gambar 7. diatas terlihat perbandingan nilai sebenarnya dan nilai prediksi hasil dari algoritma *Decision Tree*. Garis biru merupakan nilai sebenarnya dan garis orange merupakan nilai prediksi. Algoritma model *Decision Tree* bisa dikatakan lebih persisi dalam memprediksi, hal ini terlihat pada banyaknya garis oren (nilai prediksi) yang berhimpit dengan garis biru (nilai aktual). Dari gambar terlihat tanda kotak merah, ada penyimpangan hasil prediksi menggunakan algoritma *Decision Tree*. Diperlukan penelitian lanjutan dan mendalam unuk memastikan terjadinya penyimpangan nilai prediksi menggunakan algoritma *Decision Tree*.

Pengujian pada data aktual untuk model prediksi pada *Algoritma Multiple Linear Regression* dan *Algoritma Decision Tree* dengan kedalaman sarang = 47, jumlah telur = 173 butir, jenis penyu = sisik, tanggal bertelur = 27 April 2022. Dengan menggunakan pengujian data tersebut diatas diperoleh hasil seperti dibawah ini:

```

Uji model menggunakan data baru

Kedalaman_srng = 47
jumlah_telur = 173
jenis_penyu = 'sisik'
tahun = 2022
bulan = 4
hari = 27

New_dt = pd.DataFrame({
    'kedalaman_srng': [kedalaman_srng],
    'jumlah_telur': [jumlah_telur],
    'jenis_penyu': [jenis_penyu],

```

```

        'tahun': [tahun]
        'bulan': [bulan]
        'hari': [hari]
    })
new_dt
dt_model.predict(new_dt) [0]
lr_model.predict(new_dt) [0]

```

dari kodingan diatas diperoleh hasil sebagai berikut pada tabel dibawah ini:

Tabel 8. Data baru Hasil Inputan Untuk Pengujian Model

Kedalaman_srng	Jumlah_telur	Jenis_penyu	tahun	bulan	hari
47	173	Sisik	2022	4	48

Hasil yang diperoleh dari prediksi *Algoritma Multiple Linear Regression* memprediksi telur akan menetas selama 47,99 hari dan dibulatkan menjadi 48 hari, sementara *Algoritma Decission Tree* memprediksi telur akan menetas selama 48 hari. Dari kedua prediksi antara *Algoritma Multiple Linear Regression* dan *Algoritma Decission Tree*, *Algoritma Decission Tree* menghasilkan prediksi lama penetasan sesuai jumlah hari aktual.

#### 4. Kesimpulan

Berdasarkan keseluruhan rangkaian penelitian yang telah dilakukan pada implementasi *Machine Learning* dalam memprediksi penetasan penyu dapat disimpulkan bahwa dari *dataset* pendaratan penyu yang diteliti, terdiri dari 462 row dan 14 kolom. Pada tahap *Preprocessing dataset* menjadi 461 row data dan 6 kolom yang dilakukan pemrosesan *Algoritma Multiple Linear Regression* dan *Decission Tree*. Variable *dataset* yang diolah adalah tanggal bertelur, tanggal menetas, kedalaman sarang, jumlah telur, jenis penyu dan lama menetas. Pemodelan *Algoritma Multiple Linear Regression* diperoleh nilai RMSE 3,96387 pada data training dan 4,95446 pada data testing, sementara pada *Algoritma Decission Tree* nilai RMSE pada data training 4,29728 dan 4,82765 pada data testing. Implementasi *Algoritma Multiple Linear Regression* dan *Decission Tree* dalam mengkomparasi prediksi penetasan penyu (*Chelonioidea Sp*) sudah berjalan dengan baik, namun nilai RMSE nya masih diatas nilai nol. Hal ini bisa disebabkan *dataset* yang masih sedikit yakni hanya 461 dan tidak dilakukannya fitur *engineering* pada model. Untuk itu perlu dilakukan penelitian lanjutan dalam memprediksi penetasan penyu dengan pengimplementasian *Machine Learning* dengan data set yang lebih besar dan fitur atau variable yang lebih banyak selanjutnya dilakukan fitur *engineering* pada model. Pengujian pada data actual untuk model prediksi pada *Algoritma Multiple Linear Regression* dan *Algoritma Decission Tree* dengan kedalaman sarang = 47, jumlah telur = 173 butir, jenis penyu = sisik, tanggal bertelur = 27 April 2022. *Algoritma Multiple Linear Regression* memprediksi

telur akan menetas selama 47,99 dibulatkan menjadi 48 hari, sementara *Algoritma Decission Tree* memprediksi telur akan menetas selama 48 hari. Nilai prediksi yang dihasilkan algoritma *Decission tree* sama dengan nilai aktual atau nilai sebenarnya.

#### Ucapan Terimakasih

Ucapan terimakasih penulis sampaikan kepada Pimpinan dan seluruh pegawai LKKPN Pekanbaru, Direktorat Jenderal Pengelolaan Kelautan dan Ruang Laut, Kementerian Kelautan dan Perikanan yang telah memfasilitasi terlaksananya penelitian ini.

#### Daftar Rujukan

- [1] K. K. dan Perikanan, "Direktorat Pengelolaan Kelautan dan Ruang Laut." [Online]. Available: [https://kkp.go.id/djprl/artikel/12083-profil-pulau-mangkai-kepulauan-anambas-kepulauan-riau#:~:text=Kondisi Geografis,°35'00"BT](https://kkp.go.id/djprl/artikel/12083-profil-pulau-mangkai-kepulauan-anambas-kepulauan-riau#:~:text=Kondisi Geografis,°35'00).
- [2] L. P. LKKPN Pekanbaru, "Laporan Monitoring Penyu Kawasan Konservasi Pieh dan Kawasan Konservasi Anambas Tahun 2022," 2022.
- [3] D. K. dan P. R. J. Direktorat Konservasi dan Taman Nasional Laut, Direktorat Jenderal Kelautan, Pesisir dan Pulau-Pulau Kecil, *Pedoman Teknis Pengelolaan Konservasi Penyu*. 2009, 2009.
- [4] IUCN, "Red List Of Threatened Species," *Biodiversity*, vol. 8, pp. 17–26, 2007, [Online]. Available: <https://doi.org/10.1080/148888386.2007.9712825>
- [5] I. P. O. Saduarsa, I. G. N. P. Dirgayusa, and N. L. P. R. Puspitha, "Pengembangan Ekowisata Di Pulau Penyu Tanjung Benoa Bali Berdasarkan Prinsip-Prinsip Ekowisata," *J. Mar. Aquat. Sci.*, vol. 8, no. 1, p. 78, 2023, doi: 10.24843/jmas.2022.v08.i01.p09.
- [6] H. Damanhuri, A. Putra, and R. A. Troa, "Karakteristik bio-fisik pantai peneluran penyu di Pulau Laut-Sekatung Kabupaten Natuna – Provinsi Kepulauan Riau," *Pros. Simp. Nas. Magister*, vol. 3, no. 2, pp. 1–15, 2019.
- [7] H. Mukhtar, Y. Rizki, F. A. Wenando, and M. Abdul Al Aziz, "Prediksi Kunjungan Wisatawan Ke Indonesia Dengan Reduksi Noise Pada Mesin Pencari Menggunakan Metode Hilbert Huang Transform," *J. Fasilkom*, vol. 12, no. 3, pp. 152–159, 2022, doi: 10.37859/jf.v12i3.4332.
- [8] A. Hanif, H. Damanhuri, S. Suparno, and M. U. Rusli, "Tingkat Penetasan Penyu Hijau di Pulau Pandan Kawasan Konservasi Pulau Pieh, Sumatera Barat," *J. Akuatiklestari*, vol. 6, no. 1, pp. 1–9, 2022, doi: 10.31629/akuatiklestari.v6i1.4696.
- [9] A. Fauzi, F. F. Wati, I. Sulistyowati, M. F. Akbar, E. Rahmawati, and R. K. Sari, "Penerapan Metode Machine Learning Dalam Memprediksi Keberhasilan Panggilan Telemarketing Menjual Produk Bank," *Indones. J. Softw. Eng.*, vol. 6, no. 2, pp. 213–222, 2020, doi: 10.31294/ijse.v6i2.8977.
- [10] S. Ningsih and H. H. Dukalang, "Penerapan Metode Suksesif Interval pada Analisis Regresi Linier Berganda," *Jambura J. Math.*, vol. 1, no. 1, pp. 43–53, 2019, doi: 10.34312/jjom.v1i1.1742.
- [11] K. aswin N. ramadhan Wiga Maulana Baihaqi, Melia Dianingrum, "Regresi linier sederhana untuk memprediksi kunjungan pasien di rumah sakit berdasarkan jenis layanan dan umur pasien," *Simetris*, vol. 10, no. 2, pp. 671–680, 2019.
- [12] N. Luh, P. Wulandari, A. Kartika, Y. Sarja, G. Ayu, and D. Saryanti, "Prediksi Jumlah Pelanggan Dan Persediaan Barang Menggunakan Metode Regresi Linier Berganda Pada Bali Orchid," *Eksplora Inform.*, vol. L2, pp. 1–12, 2015.
- [13] I. Sungkawa, "Penerapan Regresi Linier Ganda untuk Mengukur Efisiensi Pola Penggunaan Air Tanah System Rice Intensification (SRI) di Kabupaten Bandung, Subang, dan Karawang," *ComTech Comput. Math. Eng. Appl.*, vol. 6, no.

- 2, p. 259, 2015, doi: 10.21512/comtech.v6i2.2270.
- [14] Dwita Elisa Sinaga, Agus Perdana Windarto, and Rizki Alfidillah Nasution, "Analisis Data Mining Algoritma Decision Tree Pada Prediksi Persediaan Obat (Studi Kasus : Apotek Franch Farma)," *KLIK Kaji. Ilm. Inform. dan Komput.*, vol. 2, no. 4, pp. 123–131, 2022, doi: 10.30865/klik.v2i4.328.
- [15] M. A. Abdillah, A. Setyanto, and S. Sudarmawan, "Implementasi Decision Tree Algoritma C4.5 Untuk Memprediksi Kesuksesan Pendidikan Karakter," *Respati*, vol. 15, no. 2, p. 59, 2020, doi: 10.35842/jtir.v15i2.349.
- [16] F. A. Artanto, I. Rosyadi, S. E. Rahmawati, and H. T. B. J. Pangestu, "Decision Tree Dalam Analisis Keputusan Pembelian Program Pada Perkumpulan Penggiat Programmer Indonesia," *J. Fasilkom*, vol. 12, no. 3, pp. 141–144, 2022, doi: 10.37859/jf.v12i3.3948.
- [17] K. Maharana, S. Mondal, and B. Nemade, "A review: Data pre-processing and data augmentation techniques," *Glob. Transitions Proc.*, vol. 3, no. 1, pp. 91–99, 2022, doi: 10.1016/j.gltp.2022.04.020.
- [18] I. Cholissodin and A. A. Soebroto, "AI , MACHINE LEARNING & DEEP LEARNING ( Teori & Implementasi )," no. July 2019, 2021.
- [19] M. B. B. and E. Plutzer, "Jurnal Manajemen Diversitas," vol. 1, p. 6, 2021.