

LAPORAN TUGAS AKHIR MACHINE LEARNING 1



Dosen Pengampu:

Teguh Iman Hermanto, S.Kom, M.Kom

Disusun Oleh:

Kelompok 2

- | | |
|------------------------|-------------|
| 1. Rani Evi Isnaini | : 231351128 |
| 2. Ratu Putri Danur | : 231351129 |
| 3. Saepul Zam Zam | : 231351169 |
| 4. Zalfa Khaerunnisa M | : 231351146 |

TEKNIK INFORMATIKA

STT WASTUKANCANA PURWAKARTA

2025

Laporan Tugas Akhir Pembelajaran Mesin 1

1. Identitas Aplikasi

- Nama Aplikasi : Prediksi Produksi Daging Unggas di Kota Subang
Open Data Provinsi Jawa Barat (Data Unggas Kabupaten Subang)
<https://opendata2.subang.go.id/dataset/populasi-unggas-menurut-jenis-unggas-dan-kecamatan-di-kabupaten-subang>
- Sumber Dataset : [subang](https://opendata2.subang.go.id/dataset/produksi-daging-unggas-menurut-jenis-unggas-dan-kecamatan-di-kabupaten-subang)
<https://opendata2.subang.go.id/dataset/produksi-daging-unggas-menurut-jenis-unggas-dan-kecamatan-di-kabupaten-subang>
- Link Github : https://github.com/zalfaaslm/ML_Project_Kel2
- Link Deployment : https://github.com/zalfaaslm/ML_Project_Kel2
- Link Video : <https://drive.google.com/file/d/1f9HXDtCYBxW7dFS2VqWLjMHkvFhoTTYy/view?usp=sharing>

2. Indentitas Pengembang

• Data Engineer

Nama : Zalfa Khaerunnisa Marwasalam
NIM : 231351169

• ML Analyst

Nama : Rani Evi Isnaini
NIM : 231351128

• AI App Developer

Nama : Ratu Putri Danur
NIM : 23131129

- **Project Doc**

Nama : Saepul Zam Zam

NIM : 231351146

A. BUSINESS UNDERSTANDING

1. Project Domain

Proyek ini berada dalam domain Agrikultur (Peternakan) dan Pengambilan Keputusan Pemerintah Daerah (Local Government Decision Making) menggunakan analisis Data Sektoral Peternakan. Fokusnya adalah pada pemanfaatan data populasi unggas untuk memprediksi hasil produksi daging, yang merupakan indikator penting dalam stabilitas pangan dan ekonomi lokal.

2. Problem Statements

Pemerintah daerah menghadapi kesulitan dalam memprediksi volume ketersediaan (pasokan) daging unggas di tingkat kecamatan. Kesulitan ini disebabkan oleh data populasi yang harus diolah secara manual dan tidak adanya sistem otomatis yang dapat menghasilkan prediksi nilai kontinu (Kg) yang akurat.

3. Goals

- Membangun model Regresi Machine Learning untuk memprediksi nilai kontinu (berat/jumlah) Total Produksi Daging Unggas di Kabupaten Subang.
- Mengidentifikasi jenis unggas dan populasi yang paling berkorelasi kuat dengan total produksi daging.
- Mengembangkan aplikasi interaktif berbasis Streamlit yang memungkinkan pengguna (staf dinas/analisis) untuk memasukkan data populasi dan mendapatkan hasil prediksi produksi secara real-time.

4. Solution Statements

Solusi yang dikembangkan berupa:

- Model **Regresi Linier** atau model regresi lainnya yang dilatih menggunakan data historis populasi unggas.
- Fitur input mencakup:

- Populasi Ayam Kampung
- Populasi Ayam Pedaging
- Populasi Ayam Petelor
- Populasi Itik
- Output berupa prediksi total produksi daging (kg) berdasarkan populasi unggas pada tahun dan kecamatan tertentu.
- Model diintegrasikan dengan antarmuka Streamlit untuk memudahkan staf dinas/analisis menggunakan aplikasi.

B. DATA UNDERSTANDING

1. Dataset Description

Dataset yang digunakan:

Data dikumpulkan dari dua sumber public dataset Open Data Provinsi Jawa Barat. Setelah digabungkan (merged) dan di-pivot, dataset final memiliki 60 baris dan 7 kolom (merekpresentasikan 30 kecamatan dikali 2 tahun: 2023 dan 2024).

Kolom	Deskripsi Singkat	Tipe
bps_nama_kecamatan	Nama Kecamatan di Kabupaten Subang (30 nilai unik)	<i>Object/String</i>
tahun	Tahun data diambil (2023 dan 2024)	<i>Integer</i>
ayam_kampung	Populasi Ayam Kampung (EKOR)	<i>Float</i>
ayam_pedaging	Populasi Ayam Pedaging (EKOR)	<i>Float</i>
ayam_petelor	Populasi Ayam Petelor (EKOR)	<i>Float</i>
itik	Populasi Itik/Itik Manila (EKOR)	<i>Float</i>
total_produksi_daging	TARGET (Y): Total produksi daging (KG) dari semua jenis unggas di kecamatan dan tahun tersebut.	<i>Float</i>

Sumber DataSet:

Open Data Subang – Jawa Barat

1. Populasi Unggas Menurut Jenis Unggas dan Kecamatan di Kabupaten Subang

(<https://opendata2.subang.go.id/dataset/populasi-unggas-menurut-jenis-unggas-dan-kecamatan-di-kabupaten-subang>)

2. Produksi Daging Unggas Menurut Jenis Unggas dan Kecamatan di Kabupaten Subang

(<https://opendata2.subang.go.id/dataset/produksi-daging-unggas-menurut-jenis-unggas-dan-kecamatan-di-kabupaten-subang>)

2. Library Requirements

Library yang digunakan dalam proyek:

- Pandas
- Numpy
- Matplotlib
- Seaborn
- scikit-learn
- streamlit

3. Exploratory Data Analysis

EDA dilakukan untuk memahami distribusi dan hubungan data setelah preprocessing karena data mentah (2 file terpisah dengan masalah tipe data) tidak dapat divisualisasikan untuk analisis korelasi.

Grafik	Hasil Observasi Kunci	Implikasi terhadap Pemodelan
Correlation Matrix	Menunjukkan korelasi positif yang sangat kuat (diperkirakan > 0.90) antara ayam_pedaging dan total_produksi_daging.	Variabel ayam_pedaging akan menjadi predictor utama (bobot/koeffisien terbesar) dalam model Regresi.
Scatter Plot	Menunjukkan hubungan yang linier antara Populasi Ayam Pedaging dan Total Produksi Daging.	Menguatkan justifikasi penggunaan model Regresi Linier karena

		asumsi linieritas terpenuhi.
Box Plot	Mengidentifikasi adanya Outlier di variabel total_produksi_daging atau populasi tertentu.	Outlier tersebut harus diperiksa. Jika outlier merupakan data valid (misalnya peternakan besar), outlier harus dipertahankan. Jika tidak, perlu diatasi di fase Preprocessing Lanjutan.
Line Plot & Pie Chart	Menunjukkan tren kenaikan (atau penurunan) Populasi/Produksi antar tahun dan komposisi dominan (misalnya Ayam Pedaging mungkin mendominasi populasi).	Memberikan konteks bisnis tentang status peternakan Subang.

C. DATA PREPARATION

1. Dataset Information

- Sumber data awal: 2 file CSV (Populasi dan Produksi Daging Unggas).
- Total baris awal: ~232 baris untuk tiap dataset sebelum digabungkan.
- Kunci penggabungan: 'bps_nama_kecamatan', 'jenis_unggas', dan 'tahun'.

2. Dataset Transformation

Transformasi yang dilakukan untuk mengatasi masalah data:

Jenis Pembersihan/Transformasi	Logika Problem-Solving
Pembersihan Tipe Data	Mengatasi masalah titik pemisah ribuan ('44.475') dengan menghapus titik tersebut lalu mengkonversi string menjadi tipe data

	numerik (float) menggunakan <code>pd.to_numeric()</code> .
Pivoting Data	Mengubah data dari <i>Long Format</i> ke <i>Wide Format</i> menggunakan <code>pivot_table()</code> agar setiap jenis unggas (X) menjadi kolom fitur yang independen.
Agregasi Target (Y)	Menggunakan <code>groupby().sum()</code> untuk menjumlahkan semua jenis produksi daging per Kecamatan-Tahun, menciptakan kolom Target <code>total_produksi_daging</code> yang akurat.
Merging Data	Menggabungkan <i>DataFrame</i> Populasi (X) yang sudah di- <i>pivot</i> dengan <i>DataFrame</i> Target (Y) yang sudah diagregasi menjadi satu <i>DataFrame</i> lengkap.

Dataset disimpan sebagai:

```
df_final_data.to_csv('dataset_final.csv', indeks=False)
```

3. Determine Variabels

berdasarkan pengecekan pada dataset, kolom seperti: 'id', 'kode_provinsi', 'nama_provinsi', 'bps_kode_kabupaten_kota', 'bps_nama_kabupaten_kota', 'bps_kode_kecamatan', 'kemendagri_kode_kecamatan', 'kemendagri_nama_kecamatan', tidak digunakan dalam analisis maupun pemodelan, karena:

- Tidak memiliki pengaruh langsung terhadap prediksi,
- Berpotensi menurunkan performa model.

Oleh sebab itu, fitur yang relevan untuk analisis dan modeling adalah:

Fitur (X):

- 'ayam_kampung'
- 'ayam_pedaging'
- 'ayam_petelor'

- 'itik'

Target (Y):

- 'total_produksi_daging'

D. MODELING

1. Pemilihan Algoritma

Algoritma yang digunakan pada proyek ini adalah **Linear Regression (Regresi Linear Berganda)** yang merupakan Supervised Learning.

Alasan pemilihan algoritma ini karena sesuai untuk *continuous prediction* (nilai numerik / Kg produksi daging), termasuk model yang *interpretative* sehingga mudah dijelaskan pada laporan akademik, proses training cepat dan tidak membutuhkan data sangat besar, serta cocok untuk *baseline* model pada dataset yang sederhana.

2. Pembagian Dataset (Train – Test Split)

Dataset hasil proses Data Preparation dipisahkan menjadi:

- 80% data latih (training)
- 20% data uji (testing)

```
X_train, X_test, y_train, y_test = train_test_split(X, y,
                                                    test_size=0.2, random_state=42)
```

Tujuan pemisahan ini adalah agar model dapat mempelajari pola dari data latih dan data uji digunakan untuk mengevaluasi performa secara objektif.

3. Pelatihan Model

Model dilatih menggunakan perintah:

```
model = LinearRegression()
model.fit(X_train, y_train)
```

Hasil dari proses ini:

- Model mempelajari hubungan linier antara jumlah populasi unggas dengan produksi daging.
- Koefisien dan intercept terbentuk sebagai dasar prediksi.

4. Interpretasi Parameter Model

a) Intercept

Nilai intercept menunjukkan prediksi produksi ketika seluruh fitur (populasi unggas) bernilai 0. Secara makna, ini merupakan nilai dasar prediksi model.

2

b) Coeficient

Model menghasilkan 4 koefisien:

ayam_kampung, ayam_petelor, ayam_pedanging, itik

jika koefisien positif maka bertambahnya populasi menaikkan produksi, sebaliknya jika koefisien negative maka populasi tidak berkaitan langsung atau data tidak stabil.

E. EVALUATION

Evaluasi menggunakan 4 metrik utama:

1. Mean Absolute Error (MAE)

Yaitu rata-rata selisih antara prediksi dan nilai sebenarnya. Dari project Prediksi Produksi Daging unggas:

```
mae = mean_absolute_error(y_test, y_pred)
```

Mean Absolute Error (MAE) : 16458301.759269664

Nilai MAE sebesar 16.4 juta kg menunjukkan bahwa rata-rata kesalahan prediksi model adalah sekitar 16 juta kilogram dari nilai aktual.

2. Mean Squared Error (MSE)

MSE menghitung kesalahan kuadrat, dan sensitif terhadap outlier.

```
mse = mean_squared_error(y_test, y_pred)
```

Mean Squared Error (MSE) : 1622578903353677.8

MSE menghasilkan nilai yang sangat besar karena skala data yang besar dan proses kuadrat pada error.

3. Root Mean Squared Error (RMSE)

RMSE Adalah akar kuadrat dari MSE

```
rmse = mse ** 0.5
```

Root Mean Squared Error (RMSE) : 40281247.53968871

RMSE sebesar 40 juta kg memberikan gambaran bahwa rata-rata kesalahan prediksi berada pada kisaran 40 juta kg.

4. R-Squared (R^2)

Mengukur seberapa besar model dapat menjelaskan variasi data

```
r2 = r2_score(y_test, y_pred)
```

```
R-square (R2) : 0.18430948655780544
```

Sementara itu, nilai R^2 sebesar **0.184** menunjukkan bahwa model hanya mampu menjelaskan **18% variasi data**

F. DEPLOYMENT

Tahap deployment dilakukan dengan membangun aplikasi prediksi berbasis web menggunakan Streamlit, sehingga model dapat digunakan secara interaktif oleh pengguna. Proses deployment dimulai dengan menyimpan model regresi dalam format `.pkl` menggunakan library pickle, sehingga model dapat di-load kembali tanpa perlu dilatih ulang.

Aplikasi Streamlit dibangun dengan tiga fitur utama:

- Prediksi, yaitu fitur dimana pengguna dapat memasukkan jumlah populasi setiap jenis unggas melalui slider dan memperoleh hasil prediksi produksi.
- EDA (Exploratory Data Analysis), yang menampilkan grafik line plot, box plot, scatter plot, pie chart, dan correlation matrix untuk memberikan pemahaman visual terhadap pola data.
- Dataset, yaitu tab untuk menampilkan data yang digunakan.

Tampilan aplikasi dibuat lebih interaktif dengan penambahan CSS khusus, struktur card, dan penggunaan sidebar untuk input parameter. Aplikasi kemudian dapat dijalankan secara lokal menggunakan perintah `streamlit run app.py`. Deployment online dapat dilakukan melalui platform seperti Streamlit io atau *GitHub Pages*. Tahap ini memastikan bahwa model machine learning tidak hanya berhenti sebagai konsep, tetapi dapat

digunakan langsung oleh user dalam bentuk aplikasi yang mudah dioperasikan.