Project Report

On

ABALONE AGE PREDICTION

Submitted in partial fulfilment of the requirements for the award of

BACHELOR OF TECHNOLOGY

in

COMPUTER SCIENCE & ENGINEERING

(Artificial Intelligence & Machine Learning)

By

Ms. AIMAN RAZIA (22WH1A6614)

Ms. LAKSHMI KOSURI INDU (22WH1A6619)

Ms. JAHNVI KAKKAR (22WH1A6640)

Ms. MUSKAAN (22WH1A6650)

Under the esteemed guidance of

Ms. A Naga Kalyani

Assistant Professor, CSE(AI&ML)

Department of Computer Science & Engineering

(Artificial Intelligence & Machine Learning)

BVRIT HYDERABAD COLLEGE OF ENGINEERING FOR WOMEN

AUTONOMOUS)

(Approved by AICTE, New Delhi and Affiliated to JNTUH, Hyderabad)

Accredited by NBA and NAAC with A Grade

Bachupally, Hyderabad – 500090

2024-25

Department of Computer Science & Engineering

(Artificial Intelligence & Machine Learning)

BVRIT HYDERABAD COLLEGE OF ENGINEERING FOR WOMEN

(Approved by AICTE, New Delhi and Affiliated to JNTUH, Hyderabad)

Accredited by NBA and NAAC with A Grade

Bachupally, Hyderabad – 500090

2023-24

CERTIFICATE

This is to certify that the major project entitled **"Abalone Age Prediction"** is a Bonafide work carried out **by Ms. AIMAN RAZIA (22WH1A6614) Ms. LAKSHMI KOSURI INDU (22wh1a6619), JAHNVI KAKKAR (22wh1a6640), Ms. MUSKAAN (22WH1A6650)** ,in partial Fulfillment for the award of B. Tech degree in **Computer Science & Engineering (AI&ML), BVRIT HYDERABAD College of Engineering for Women, Bachupally, Hyderabad,** affiliated to Jawaharlal Nehru Technological University Hyderabad, Hyderabad under my guidance and supervision. The results embodied in the project work have not been submitted to any other University or Institute for the award of any degree or diploma.

Supervisor
Ms. A Naga Kalyani
Assistant Professor
Dept of CSE(AI&ML)

Head of the Department
Dr. B. Lakshmi Praveena
HOD & Professor

DECLARATION

We hereby declare that the work presented in this project entitled **"Abalone Age Prediction"** submitted towards completion of Project work in III Year of B.Tech of CSE(AI&ML) at **BVRIT HYDERABAD College of Engineering for Women,** Hyderabad is an authentic record of our original work carried out under the guidance of **Ms. A Naga Kalyani, Assistant Professor, Department of CSE(AI&ML).**

Sign with Date:
Aiman Razia
(22WH1A6614)

Sign with Date:
Lakshmi Kosuri Indu
(22WH1A6619)

Sign with Date:
Jahnvi Kakkar
(22WH1A6640)

Sign with Date:
Muskaan
(22WH1A6650)

ACKNOWLEDGEMENT

We would like to express our sincere thanks to **Dr. K V N Sunitha, Principal, BVRIT HYDERABAD College of Engineering for Women**, for her support by providing the working facilities in the college.

Our sincere thanks and gratitude to **Dr. B Lakshmi Praveena, Head of the Department, Department of CSE(AI&ML), BVRIT HYDERABAD College of Engineering for Women,** for all timely support and valuable suggestions during the period of our project.

We are extremely thankful to our Internal Guide, **Ms. A Naga Kalyani, Assistant Professor, CSE(AI&ML), BVRIT HYDERABAD College of Engineering for Women,** for her constant guidance and encouragement throughout the project.

Finally, we would like to thank our Major Project Coordinator, all Faculty and Staff of CSE(AI&ML) department who helped us directly or indirectly. Last but not least, we wish to acknowledge our **Parents** and **Friends** for giving moral strength and constant encouragement.

Aiman Razia
(22WH1A6614)

Lakshmi Kosuri Indu
(22WH1A6619)

Jahnvi Kakkar
(22WH1A6640)

Muskaan
(22WH1A6650)

PROBLEM STATEMENT

Abalones, a type of marine mollusk, are vital for ecological balance and economic activities such as aquaculture and seafood production. However, determining their age accurately remains a challenge, typically involving labor-intensive and invasive methods like counting growth rings on their shells.

This project seeks to address the challenge of predicting abalone age with high accuracy by employing a machine learning-based approach. The primary goals include:

1. Develop a regression model based on historical data of abalone physical measurements to predict their age.

2. Create a classification framework to identify abalones above or below critical age thresholds for better resource management and breeding programs.

3. Evaluate the effectiveness of the models through error metrics such as RMSE, accuracy, and confusion matrix, ensuring their applicability in real-world scenarios.

The ultimate objective is to enhance age prediction accuracy, support sustainable harvesting practices, and promote effective conservation strategies for abalones.

.

ABSTRACT

The **Abalone Age Prediction Project** aims to develop a machine-learning model capable of accurately predicting the age of abalones. This model leverages a Random Forest Regression approach, utilizing various physical measurements of abalones, such as shell length, diameter, and weight, to estimate their age. Accurate age prediction is essential for ecological studies, sustainable harvesting, and effective conservation strategies.

Historical data on abalone measurements is analyzed and processed to train the regression model, ensuring it can provide reliable age estimates. Additionally, the project incorporates a classification model to identify abalones that fall within specific age thresholds, which is vital for resource management and breeding programs.

To ensure the model's robustness and reliability, performance is evaluated using several metrics, including Root Mean Squared Error (RMSE), accuracy, and a confusion matrix. These evaluations help fine-tune the model, enhancing its predictive capabilities and ensuring it meets the project's goals.

By providing precise age predictions, the Abalone Age Prediction Project supports efforts in sustainable aquaculture, helping to maintain balanced abalone populations and promote long-term ecological health.

DATASET

Abalone Age Predictiion– Kaggle

https://www.kaggle.com/code/shubh247/abalone-age-prediction

SOURCE CODE

```
# Import necessary libraries

import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

from sklearn.model_selection import train_test_split

from sklearn.linear_model import LinearRegression, Ridge

from sklearn.svm import SVR

from sklearn.ensemble import RandomForestRegressor,
GradientBoostingRegressor

from sklearn.neighbors import KNeighborsRegressor

from sklearn.metrics import (

    mean_squared_error, mean_absolute_error, r2_score

)

from sklearn.preprocessing import StandardScaler

# Load the dataset

data = pd.read_csv('../input/abalone.csv')

# Display basic information
```

```
print(data.describe())

print(data.columns)

# Add a new 'age' column and drop 'Rings'

data['age'] = data['Rings'] + 1.5

data = data.drop('Rings', axis=1)

# Correlation Heatmap

plt.figure(figsize=(10, 6))

sns.heatmap(data.corr(), annot=True, cmap='coolwarm')

plt.title("Correlation Heatmap")

plt.show()

# Distribution of age

plt.figure(figsize=(8, 5))

sns.histplot(data['age'], kde=True, color='blue')

plt.title("Distribution of Abalone Age")

plt.xlabel("Age")

plt.ylabel("Frequency")

plt.show()

# Pairplot to analyze feature relationships

sns.pairplot(data)

plt.show()

# Countplot for categorical feature 'Sex'

sns.countplot(x='Sex', data=data, palette='Set3')
```

```
plt.title("Count of Sex Categories")

plt.show()

# Handle categorical data using one-hot encoding

data = pd.get_dummies(data)

# Select features and target

X = data.drop(['age'], axis=1)

y = data['age']

# Split the data into train and test sets

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

# Initialize a scaler and scale the features

scaler = StandardScaler()

X_train = scaler.fit_transform(X_train)

X_test = scaler.transform(X_test)

# Train and evaluate models

# Linear Regression

lr = LinearRegression()

lr.fit(X_train, y_train)

lr_pred = lr.predict(X_test)

print("Linear Regression:")

print(f"RMSE: {np.sqrt(mean_squared_error(y_test, lr_pred))}")

print(f"R-squared: {r2_score(y_test, lr_pred)}")
```

```
# Ridge Regression
ridge = Ridge(alpha=0.001, solver='sag', random_state=42)
ridge.fit(X_train, y_train)
ridge_pred = ridge.predict(X_test)
print("\nRidge Regression:")
print(f"RMSE: {np.sqrt(mean_squared_error(y_test, ridge_pred))}")
print(f"R-squared: {r2_score(y_test, ridge_pred)}")
# Support Vector Regression
svr = SVR(kernel='linear')
svr.fit(X_train, y_train)
svr_pred = svr.predict(X_test)
print("\nSupport Vector Regression:")
print(f"RMSE: {np.sqrt(mean_squared_error(y_test, svr_pred))}")
print(f"R-squared: {r2_score(y_test, svr_pred)}")
# Random Forest Regressor
rf = RandomForestRegressor(n_estimators=100, random_state=42)
rf.fit(X_train, y_train)
rf_pred = rf.predict(X_test)
print("\nRandom Forest Regression:")
print(f"RMSE: {np.sqrt(mean_squared_error(y_test, rf_pred))}")
print(f"R-squared: {r2_score(y_test, rf_pred)}")
# Gradient Boosting Regressor
```

```
gbr = GradientBoostingRegressor(random_state=42)

gbr.fit(X_train, y_train)

gbr_pred = gbr.predict(X_test)

print("\nGradient Boosting Regression:")

print(f"RMSE: {np.sqrt(mean_squared_error(y_test, gbr_pred))}")

print(f"R-squared: {r2_score(y_test, gbr_pred)}")

# K-Nearest Neighbors Regressor

knn = KNeighborsRegressor(n_neighbors=4)

knn.fit(X_train, y_train)

knn_pred = knn.predict(X_test)

print("\nK-Nearest Neighbors Regression:")

print(f"RMSE: {np.sqrt(mean_squared_error(y_test, knn_pred))}")

print(f"R-squared: {r2_score(y_test, knn_pred)}# Plot True vs Predicted values for the best model

plt.figure(figsize=(8, 5))

sns.scatterplot(x=y_test, y=rf_pred, alpha=0.7)

plt.plot([min(y_test), max(y_test)], [min(y_test), max(y_test)], color='red', linestyle='--')

plt.title("True vs Predicted Age (Random Forest)")

plt.xlabel("True Age")

plt.ylabel("Predicted Age")

plt.show()
```

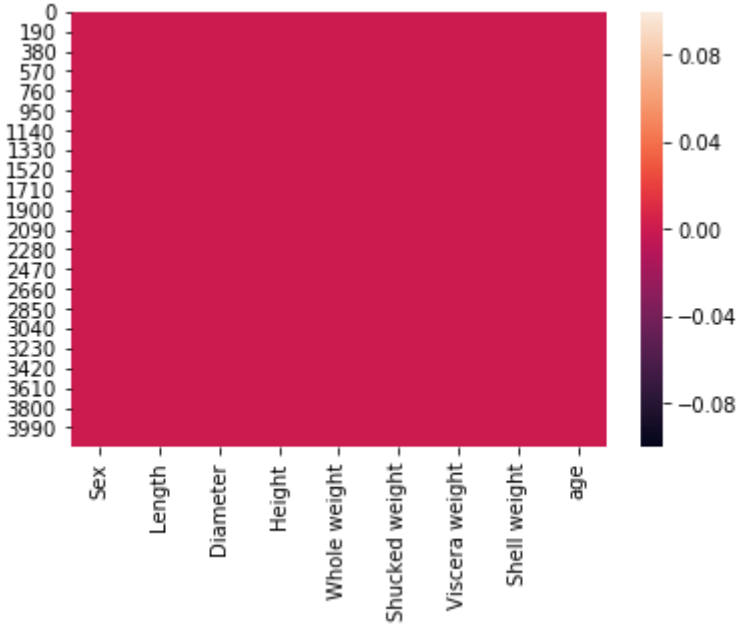
OUTPUT

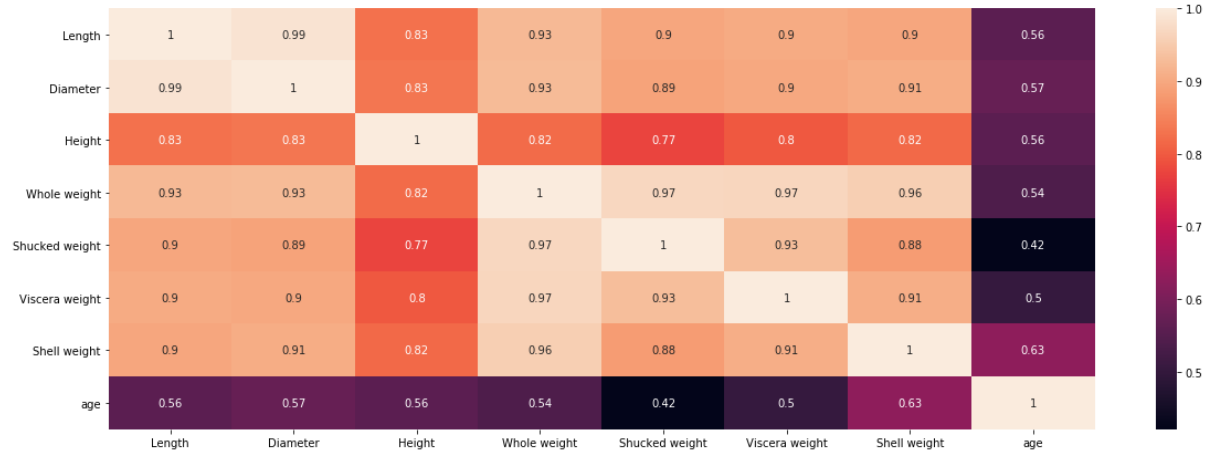
Displaying the Dataset

	Sex	Length	Diameter	...	Viscera weight	Shell weight	Rings
0	M	0.455	0.365	...	0.1010	0.150	15
1	M	0.350	0.265	...	0.0485	0.070	7
2	F	0.530	0.420	...	0.1415	0.210	9
3	M	0.440	0.365	...	0.1140	0.155	10
4	I	0.330	0.255	...	0.0395	0.055	7

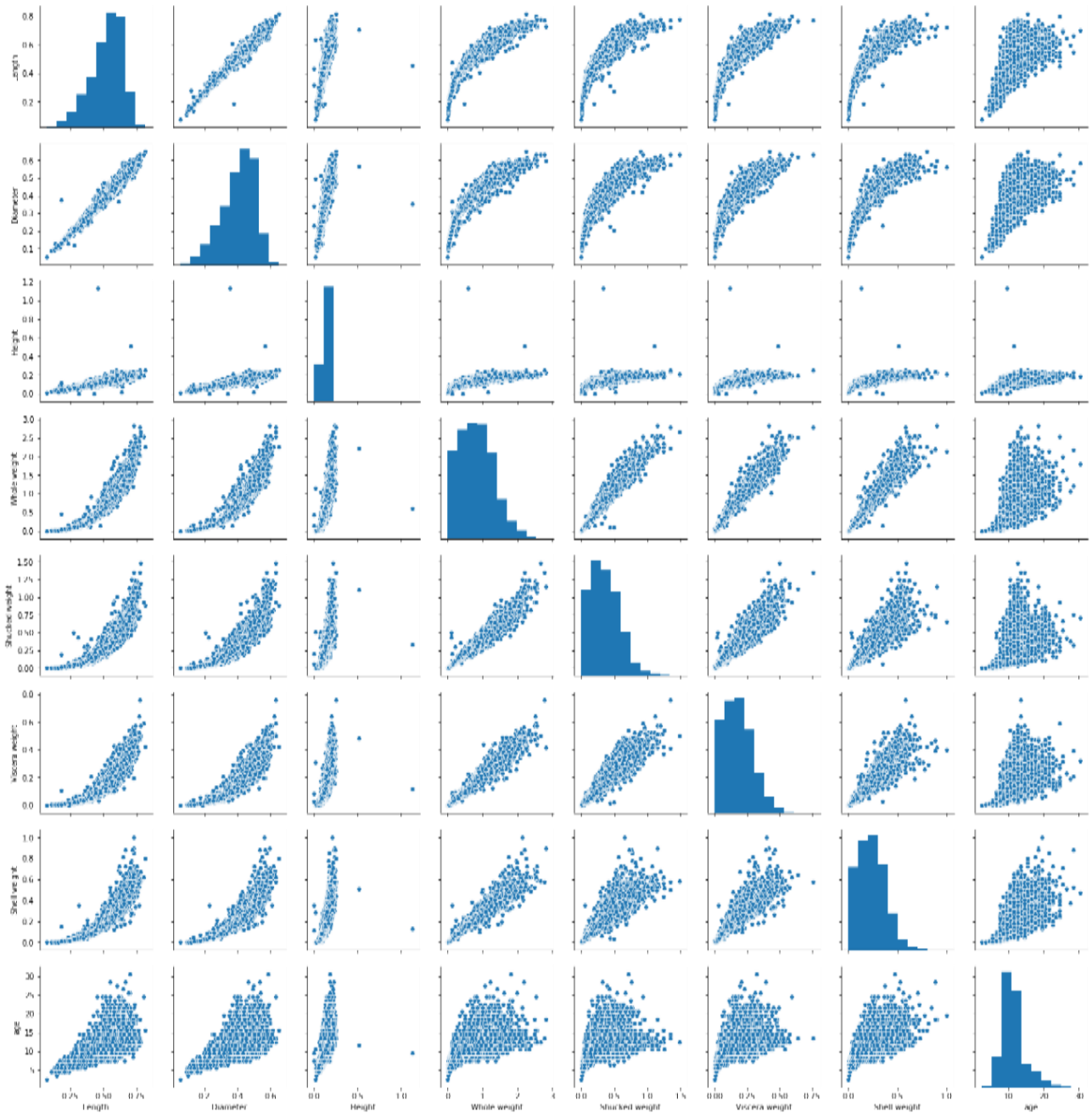
[5 rows x 9 column]

CO-RELATION HEAT MAP

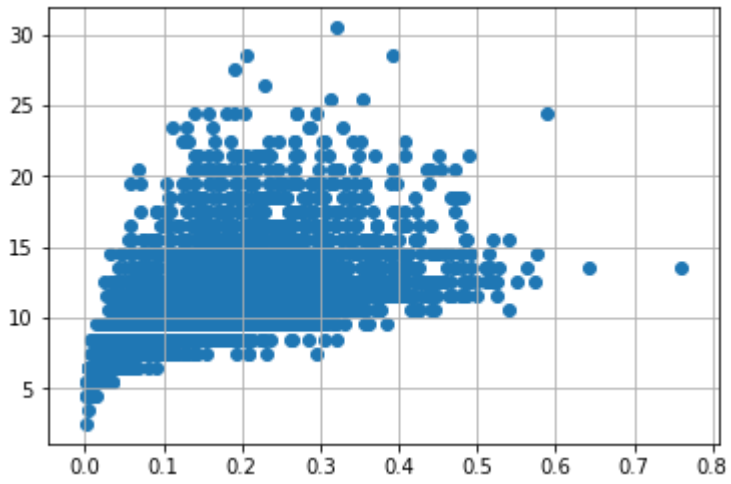


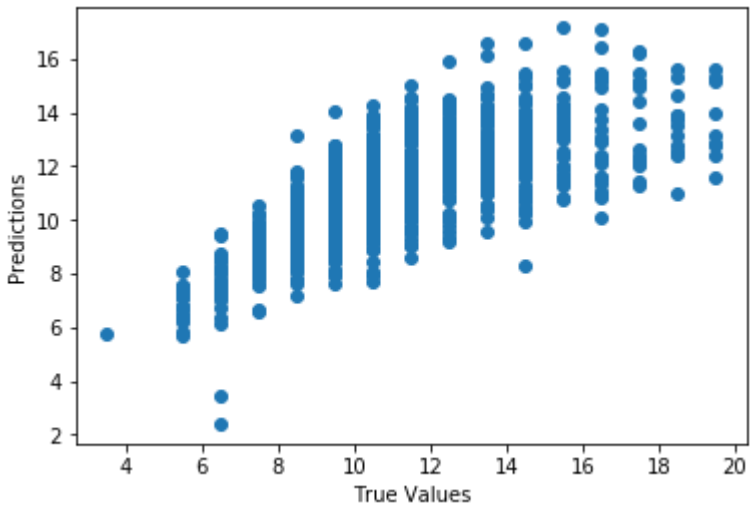


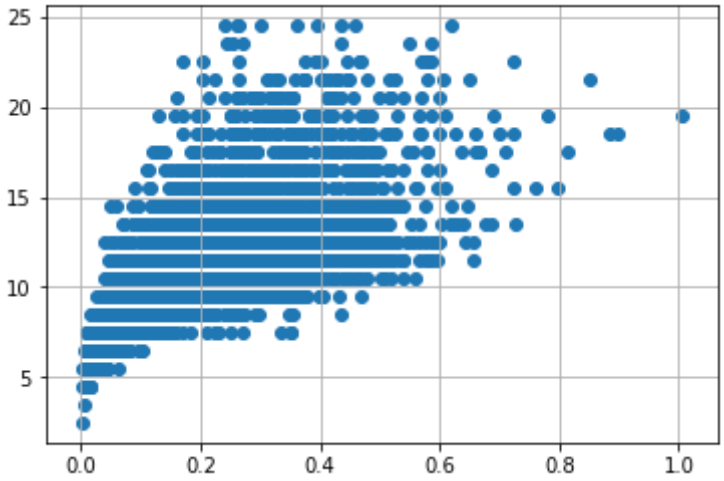
DISTRIBUTION OF TARGET VARIABLE



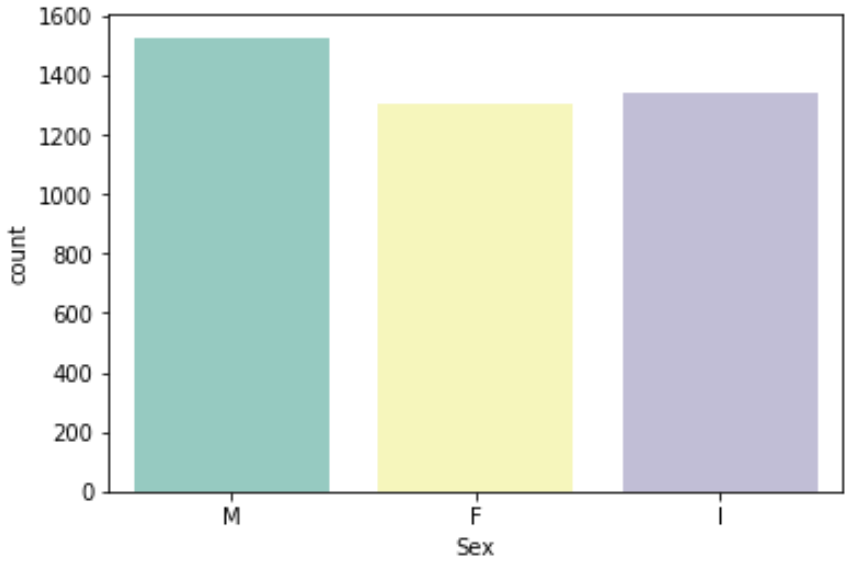
Scatter plot of two features







Frequency



Support vector Regression

SVR(C=1.0, cache_size=200, coef0=0.0, degree=3, epsilon=0.1,

gamma='auto_deprecated', kernel='linear', max_iter=-1, shrinking=True,

tol=0.00, verbose=False

Github Link

https://github.com/aimanrazia/Abalone-Age-Prediction