# ADVANCED MACHINE LEARNING

| ID | Name |
| --- | --- |
| 20210364 | زياد ضياء عبدالله |
| 20210195 | اواب محمود الطرابيلي |
| 20211053 | يس ضياء الدين احمد |
| 20210470 | شهد محمد عبد السلام |
| 20210275 | حبيبه انور شافعي |
| 20210239 | تسنيم محسن محمود |

# CLASSIFICATION MODELS

## Dataset information

**Dataset name:** MNIST Handwritten Digits

**Number of classes and their:** 10 classes

**Labels of classes:** 0, 1, 2, 3, 4, 5, 6, 7, 8, 9

**Total number of samples in dataset:** 70,000

**The size of each sample:** 28*28 pixels

**Number of samples used in training:** 48,000 samples (60% from dataset)

**Number of samples used in validation:** 12,000 samples (20% from dataset)

**Number of samples used in testing:** 10,000 samples (20% from dataset)

# NEURAL NETWORK (ANN) MODEL

## Preprocessing:

**Normalization:** Each pixel value in the training and test images is converted to ‘float32’ format and normalized to the range [0,1] by dividing by 255.

**Reshape:** Each image 28*28 pixels is flattened into a 1D array of length 784 for both training (x_train) and test (x_test) datasets.

## Model Setup:

**Model Architecture:** The model is constructed using a ‘Sequential’ , It consists of:

Input layer ( Dense with 784 units).

Two hidden layers (Dense with 128 and 64 units ) using ReLU activation.

Output layer (Dense with 10 units) using softmax activation.

**Model Compilation:** The model is compiled with the Adam optimizer (learning_rate=0.001) , categorical cross-entropy loss and accuracy metric.

**Training Process:**

The training data is split into training and validation sets.

The model is trained using ‘model.fit()’ for 50 epochs, with a batch size of 32.

Early stopping is applied with a patience of 8 epochs based on validation accuracy ensuring the model stops training if the validation accuracy does not improve by at least ‘0.01’ over 8 consecutive epochs.

**Accuracy:**

Train accuracy:

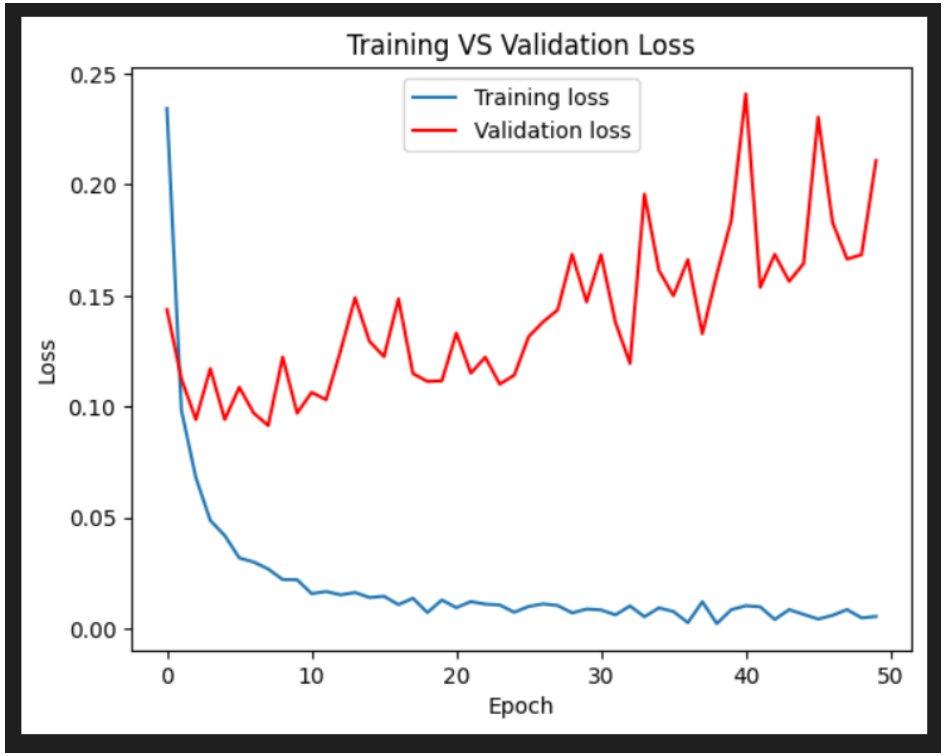
0.99

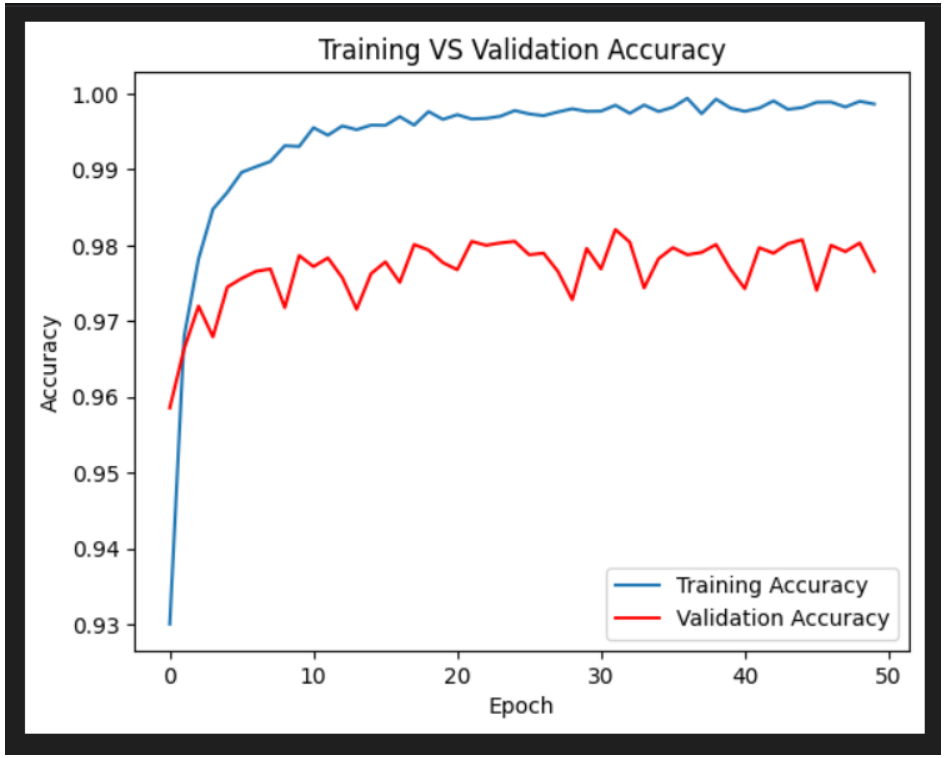
Validation accuracy:

0.97

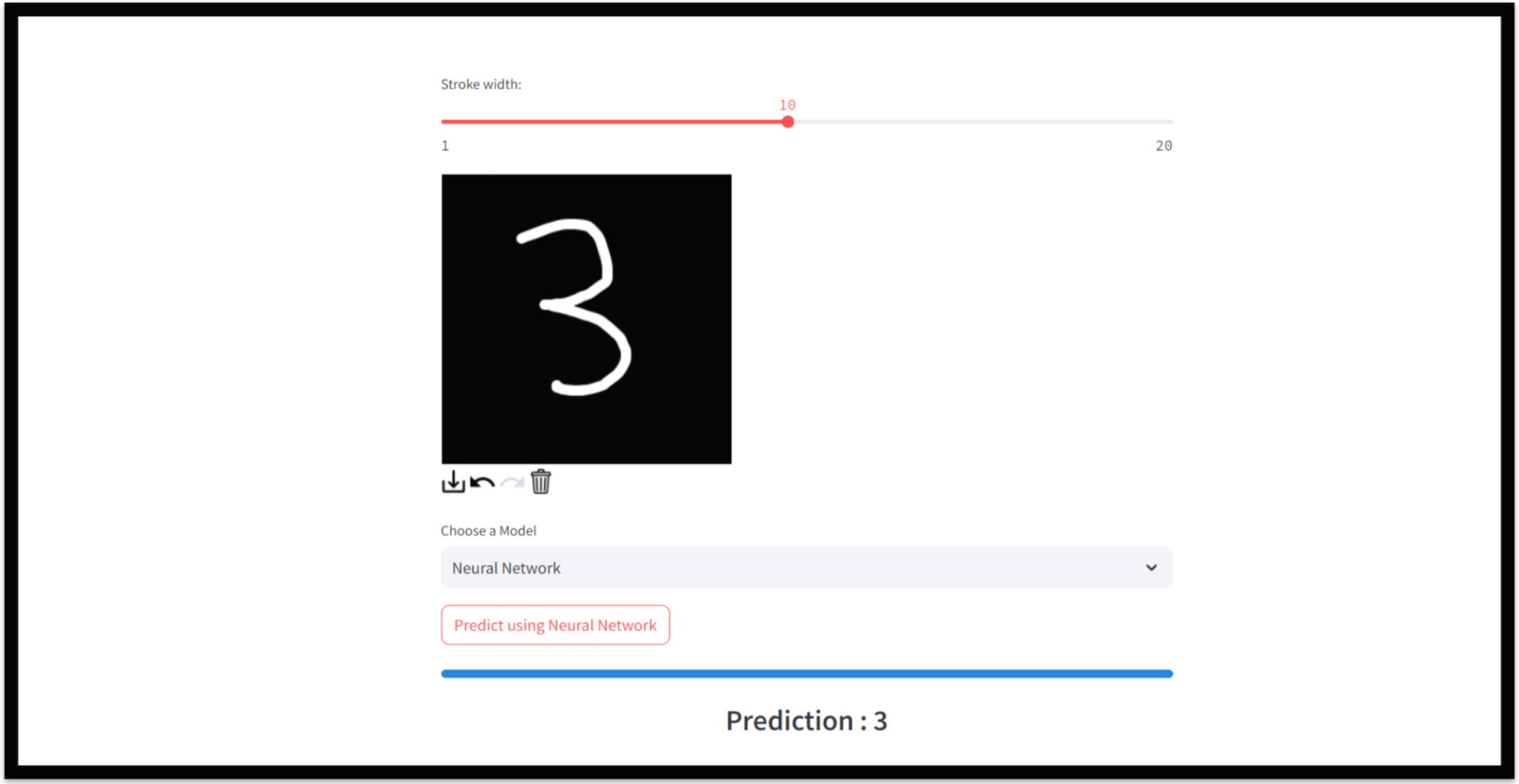
Test accuracy:

0.97



# GUI:


Stroke width:
10
1
20
Choose a Model
Neural Network
Predict using Neural Network
Prediction : 3

# DECISION TREE MODEL

## Preprocessing:

**Normalization:** Each pixel value in the the input data is converted to ‘float32’ format and normalized to the range [0,1] by dividing by 255.

**Reshape:** Each image 28*28 pixels is flattened into a 1D array of length 784 for the input data .

## Model Setup:

Initializes a Decision Tree Classifier with specified hyperparameters:

max_depth=30 ,min_samples_split=5 ,min_samples_leaf=3 , random_state=42.

## Exploring Different Max Depths and Evaluating Model Performance:

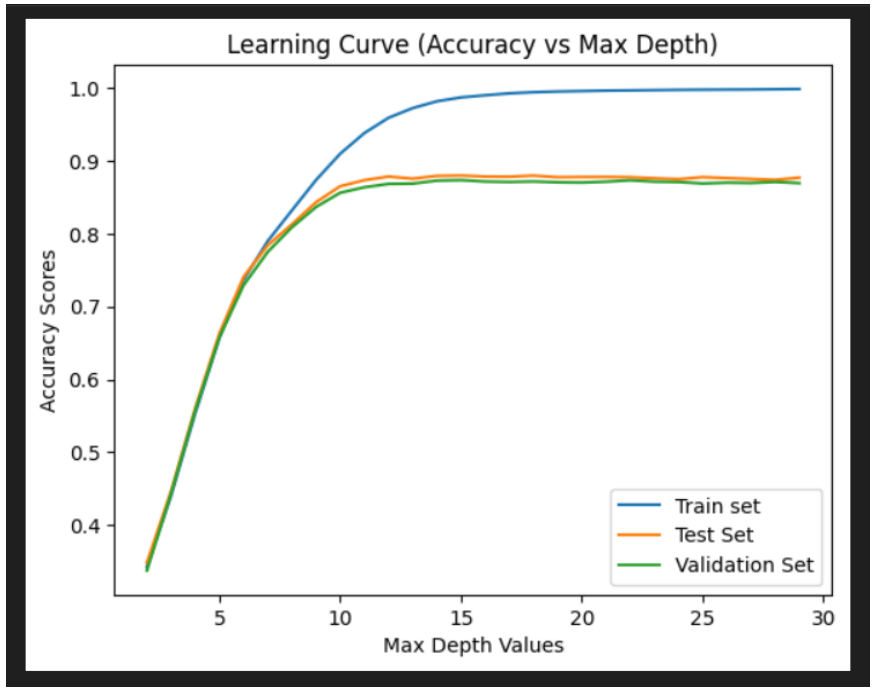
Defines a range of ‘max_depth’ values 2 to 29 to explore different depths for the decision tree model.

Defines empty lists that will be used to store accuracy scores for the training, test, and validation sets respectively, corresponding to each ‘max_depth’ value explored.

Iterates over each ‘max_depth’ in the ‘explored_depths’.

The analysis indicates that the best performance for your decision tree classifier was achieved with a ‘max_depth’ of 15.

Retraining the Pruned Model with Best Depth.



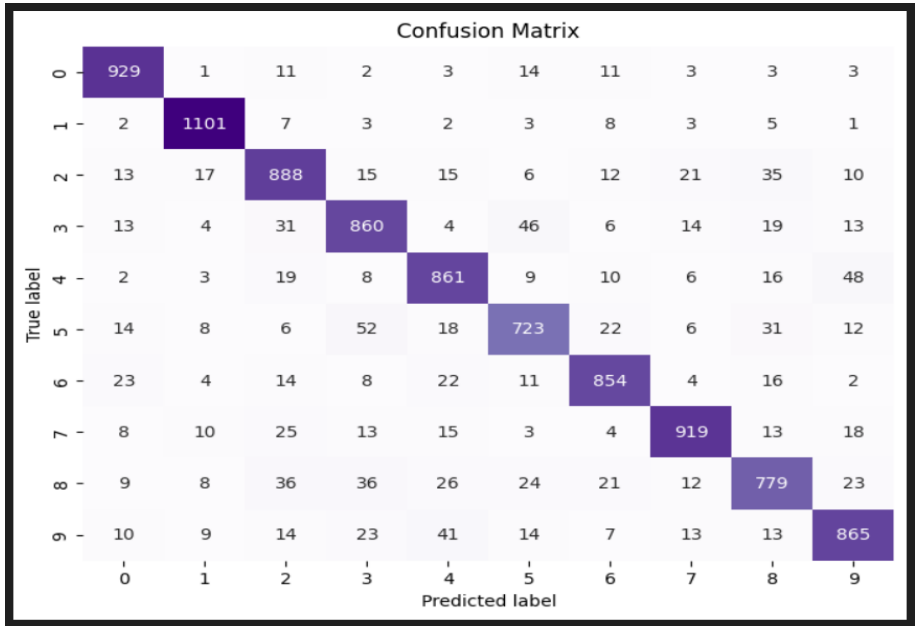

**Calculating F1-Scores:** The ‘av_f1’ value (0.8808) represents the overall F1-score of the decision tree classifier on the test dataset, providing a consolidated measure of its classification performance across all classes.

**Identifying the Hardest Class:** The class 8 is determined as the class where the model struggled the most, based on the lowest F1-score among all classes.

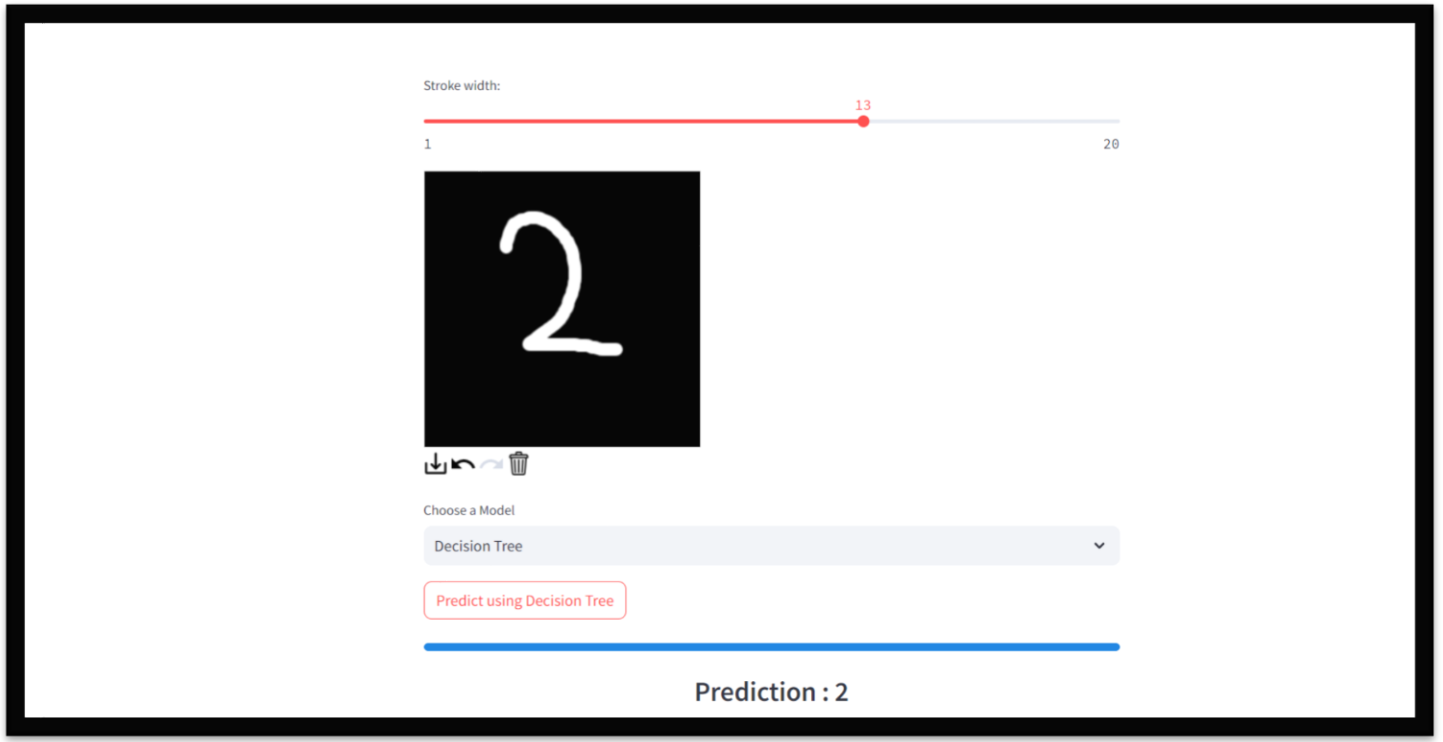
# Accuracy:

Validation accuracy: 0.87

Test accuracy: 0.8808




# GUI:



# REGRESSION MODEL

## Dataset information

**Dataset name:** CO2 Emissions_Canada

**Number of samples used in training:** 4300 (60% from dataset)

**Number of samples used in validation:** 1076 (20% from dataset)

**Number of samples used in testing:** 1344 sample (20% from dataset)

# SUPPORT VECTOR MACHINE MODEL

**Data preprocessing:**

Drop unnecessary columns and remove outliers.

Dealing with the categorical data (vehicles , transmission , fuel type).

Standardize Numerical Columns.

## Model:

Initialize an SVR model with a radial basis function (RBF) kernel and set the regularization parameter ‘c’ to ‘100.0’.

Initialize an SVR model with a radial basis function (RBF) kernel and set the regularization parameter to learn the underlying patterns in the data.

Use the trained SVR model to make predictions on the test data.

## Evaluation Metrics:

Compute various evaluation metrics to assess the performance of the SVR model:

Mean Squared Error: 7.63

Mean Absolute Error:1.97

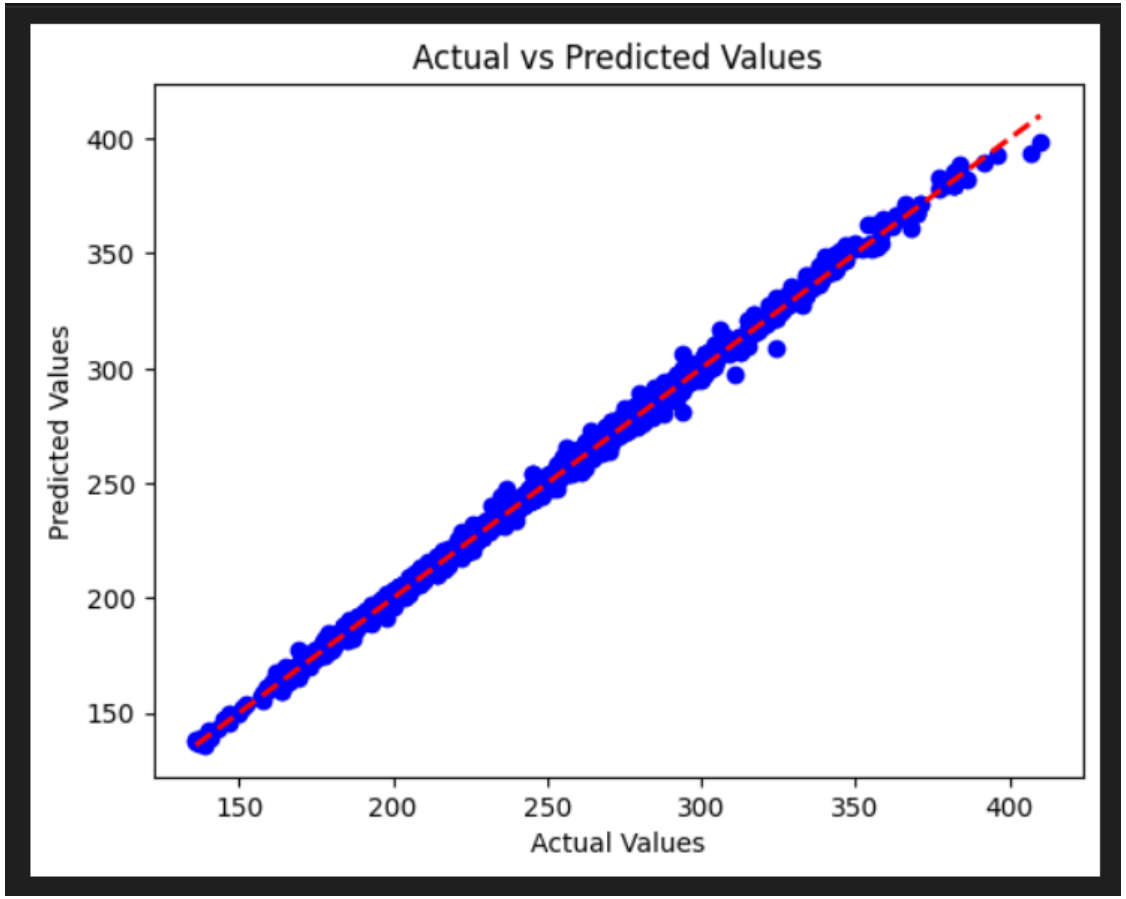
R-Squared Score:99.6%

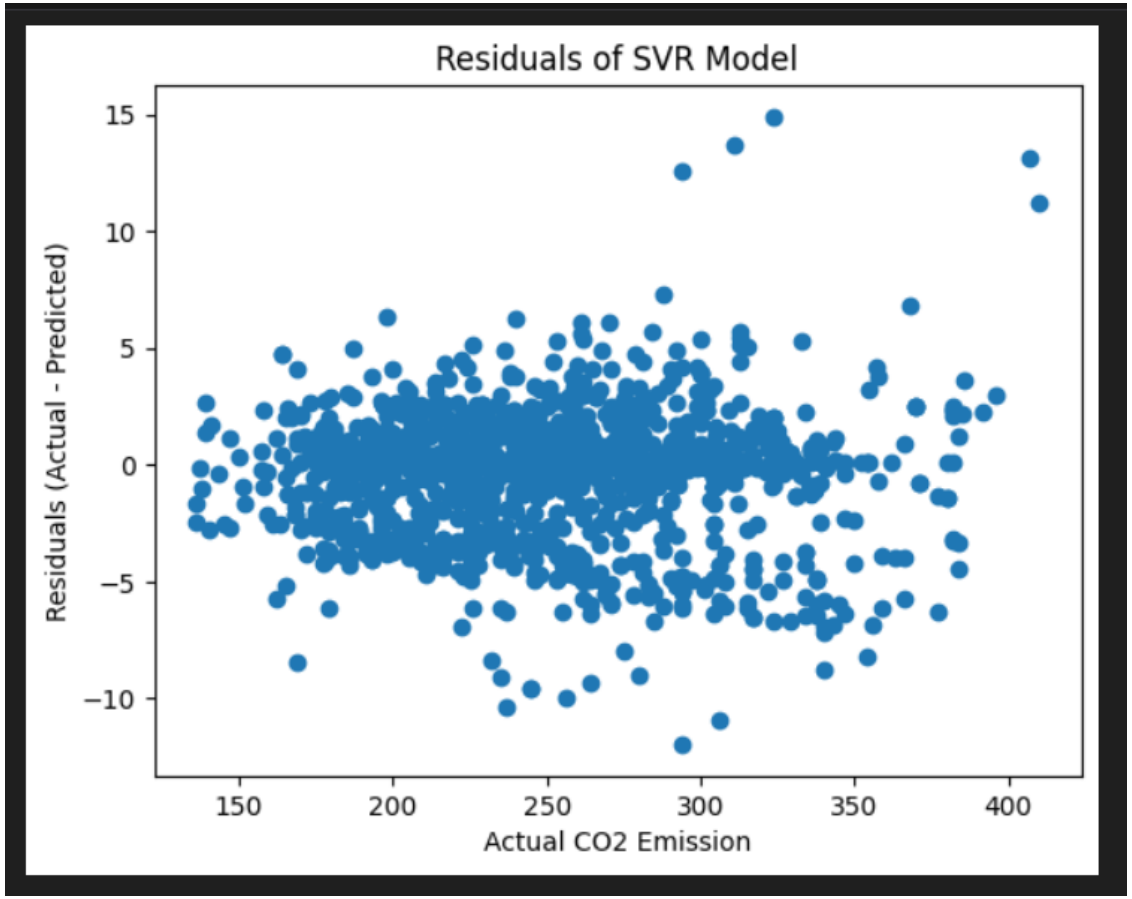
**Accuracy:**

Train accuracy: 0.99

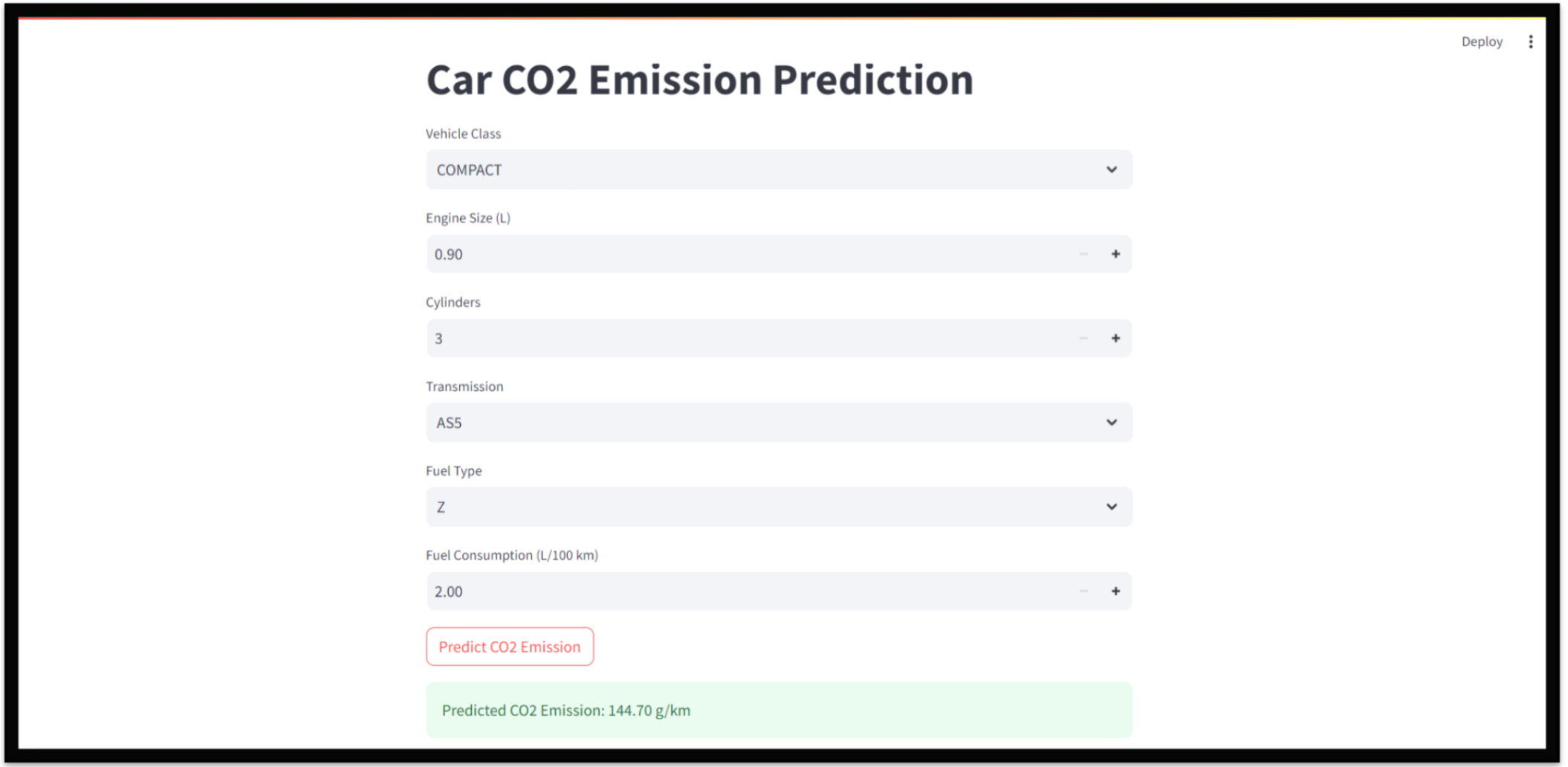
Validation accuracy: 0.99

Test accuracy: 0.99


Actual vs Predicted Values
Predicted Values
Actual Values
150
200
250
300
350
400


Residuals of SVR Model
Residuals (Actual - Predicted)
Actual CO2 Emission
15
10
5
0
−5
−10
150
200
250
300
350
400

# GUI:


Deploy
Car CO2 Emission Prediction
Vehicle Class
COMPACT
Engine Size (L)
0.90
Cylinders
3
Transmission
AS5
Fuel Type
Z
Fuel Consumption (L/100 km)
2.00
Predict CO2 Emission
Predicted CO2 Emission: 144.70 g/km