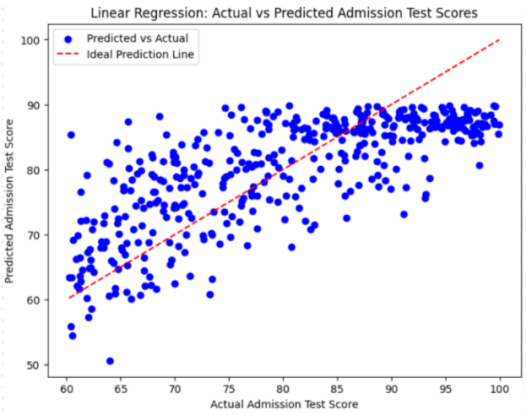
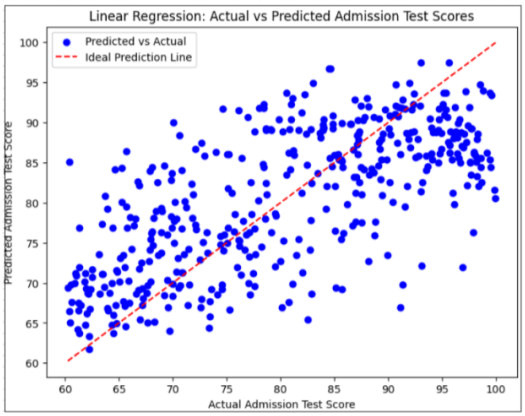
Machine Learning Project

1) Numerical Dataset
   - **Overview:** The goal of this project is to apply Linear Regression on a student dataset to predict the Admission Test Score based on various factors. The dataset includes several data quality issues such as missing values, nulls, duplicates, and outliers. Therefore, thorough data processing and cleansing steps were applied before running the regression model
   - **Name of Dataset:** Students Admission Records
   - **Columns:**
     - Name
     - Age
     - Gender,
     - Admission Test Score,
     - Hight School Percentage,
     - City
     - Admission Status
   - **Dataset Details:** The dataset initially contained 158 rows. To increase the sample size for training the regression model, an additional 2,000 rows were generated and appended to the dataset, making the total number of rows 2,158.
   - **Applied Algorithm for data processing:**
     1. Handling missing data:
        a. Mean for numerical columns
        b. Mode for categorical columns
     2. Handling Outliers:
        a. Using the IQR method as a function for all numerical columns
     3. Removing duplicates
     4. Encoding Categorical Variables
     5. Data Augmentation: Generate 2000 rows and store into a new csv file
   - **Model Implementation:**
     - X: Age, Gender, Hight School Percentage, City, Admission Status
     - y: Admission Test Score
     - A **Multilinear Regression** model was implemented to predict the Admission Test Score based on the independent variables.

```
Mean Squared Error: 61.57
Mean Absolute Error: 6.50
Linear Regression R^2: 0.55
```




```
Mean Squared Error: 72.20
Mean Absolute Error: 6.83
KNN Regression R^2: 0.47
```

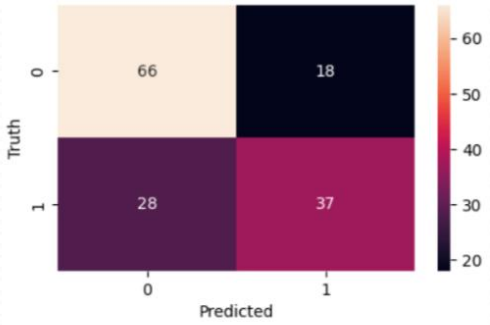


2) Image Dataset
   (We made 2 different logistic regressions, one for species (dogs or cats), other for predicting breed type of pet)
   - **Database name:** Oxford-IIIT Pet Dataset
   -
   - **Number of class:** 37 categories of animals
   - **Labels:** Image, CLASS-ID, SPECIES, BREED ID
     i. ID: 1:37 Class ids
     ii. SPECIES: 1:Cat 2:Dog
     iii. BREED ID: 1-25:Cat 1:12:Dog
   - **Total Samples:** 7,349 images
   - **Image Size:** 224x224 pixels for processing.

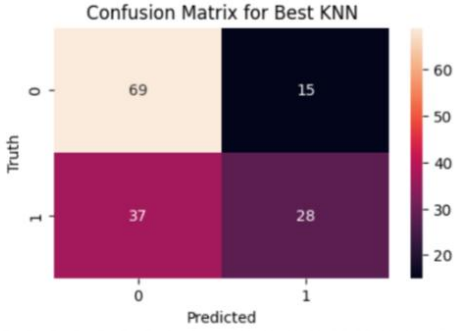
Species Dataset: dataset for dogs and cats

- **Classes:** Dogs, Cats
- **Sample Size**: resized to 32x32
- **Implementation Details:**

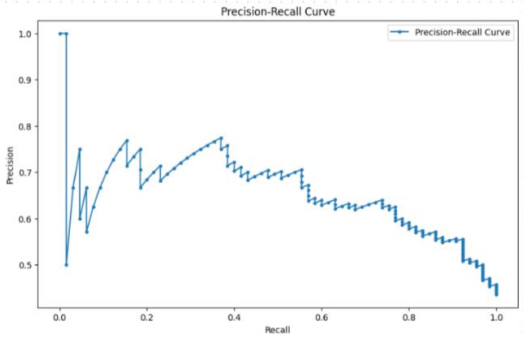
```
Accuracy for Class: 0.6912751677852349
```



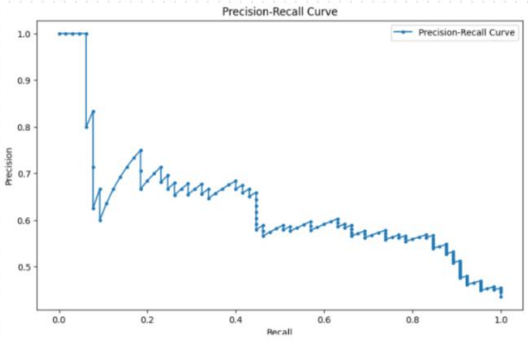

```
Accuracy for Best KNN: 0.6510067114093959
```

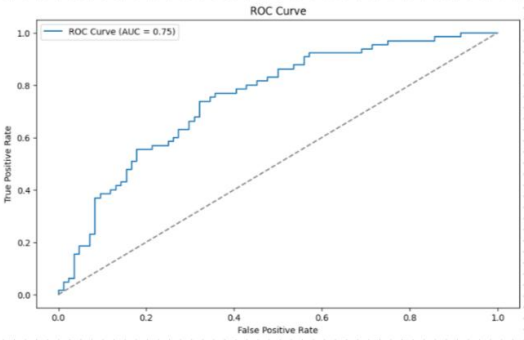



```
Training Size: 27, Training Loss: 0.0093, Test Loss: 1.2308
Training Size: 90, Training Loss: 0.0132, Test Loss: 1.3408
Training Size: 152, Training Loss: 0.0153, Test Loss: 1.1808
Training Size: 214, Training Loss: 0.0169, Test Loss: 1.2629
Training Size: 277, Training Loss: 0.0186, Test Loss: 1.2599
```

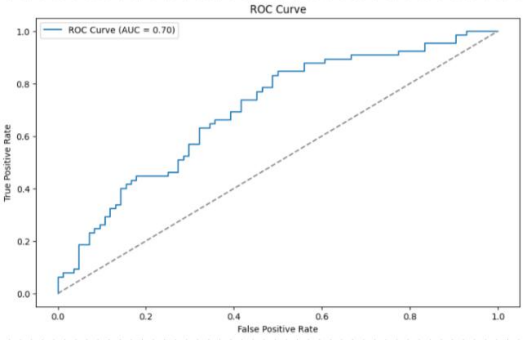



```
Training Size: 27, Training Loss: 0.0000, Test Loss: 0.6754
Training Size: 90, Training Loss: 0.0000, Test Loss: 0.8690
Training Size: 152, Training Loss: 0.0000, Test Loss: 0.7474
Training Size: 214, Training Loss: 0.0000, Test Loss: 0.6426
Training Size: 277, Training Loss: 0.0000, Test Loss: 0.7268
```





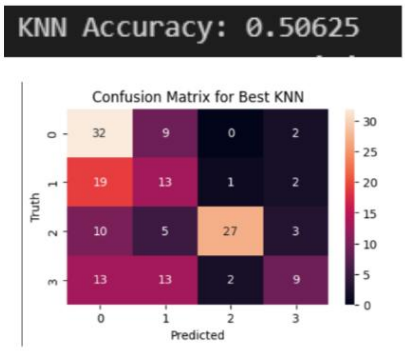

```
AUC: 0.7503663003663004
```

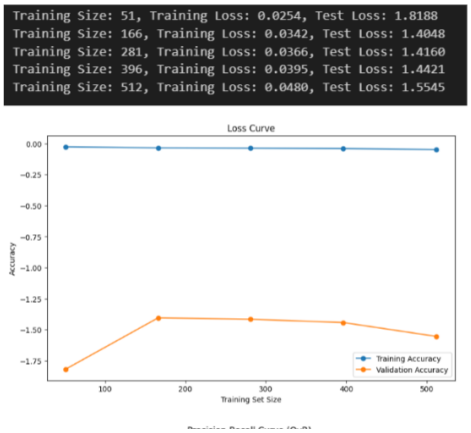


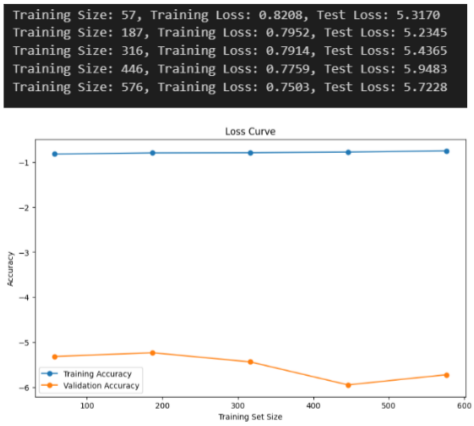

```
AUC: 0.6990842490842492
```

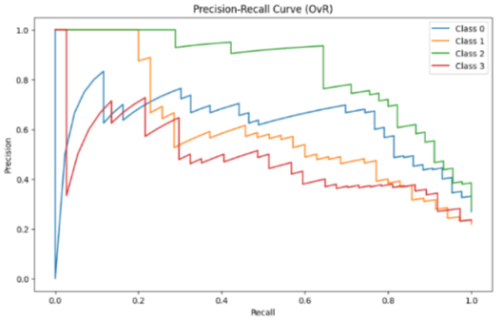
Breed Dataset:

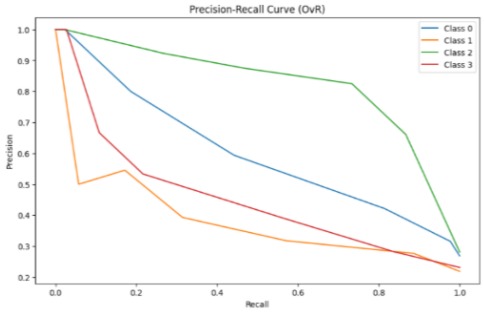
- Classes: 5 classes=> basset hound, beagle, Bengal, Birman, Bombay
- Sample Size: resized to 64x64
- Implementation Details

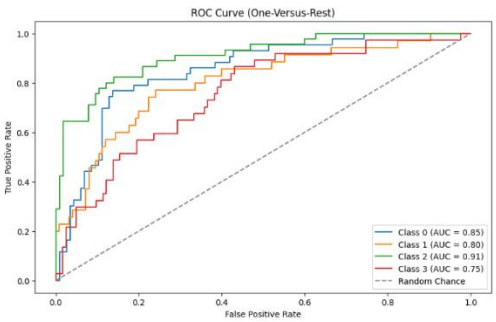




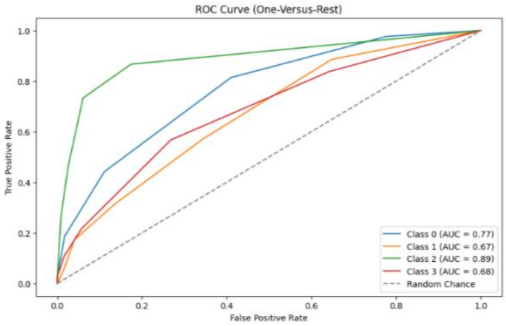









```
AUC Scores for Each Class:
Class 0: AUC = 0.85
Class 1: AUC = 0.80
Class 2: AUC = 0.91
Class 3: AUC = 0.75
```




```
AUC Scores for Each Class:
Class 0: AUC = 0.77
Class 1: AUC = 0.67
Class 2: AUC = 0.89
Class 3: AUC = 0.68
```