

Project Description Document

numerical dataset

- ***a- General Information on dataset:***

Name: Exercise and Calories Dataset

Number of classes: none

Total number of samples: 13691

Number of samples used in:

- *Training / Validation:* 10952
- *Validation:* 2738
- *Testing:* 2739

b-Implementation details:

- At feature extraction phase
 - number of features : 7
 - feature names : ['Age', 'Height', 'weight', 'Duration', 'Heart_Rate', 'Body_Temp', 'C_male']
 - feature dimension :(13691, 7)

▷ ∨

```
1 features = data.drop(columns='Calories' , axis=1)
2
3 # Number of features
4 num_features = features.shape[1]
5
6 # Feature names
7 feature_names = features.columns.tolist()
8
9 # Dimension of features
10 feature_dimension = features.shape
11
12 # Print the results
13 print("Number of Features:", num_features)
14 print("Feature Names:", feature_names)
15 print("Feature Dimension:", feature_dimension)
16
```

[17] ✓ 0.1s

```
... Number of Features: 7
Feature Names: ['Age', 'Height', 'Weight', 'Duration', 'Heart_Rate', 'Body_Temp', 'C_male']
Feature Dimension: (13691, 7)
```

- Cross-validation:

- used in Linear regression and KNN
- the number of fold=5
- ratio of training/validation: 80% training and 20% validation

- Hyperparameters used in KNN:

- KNeighborsRegressor(n_neighbors=3)

C-Results details:

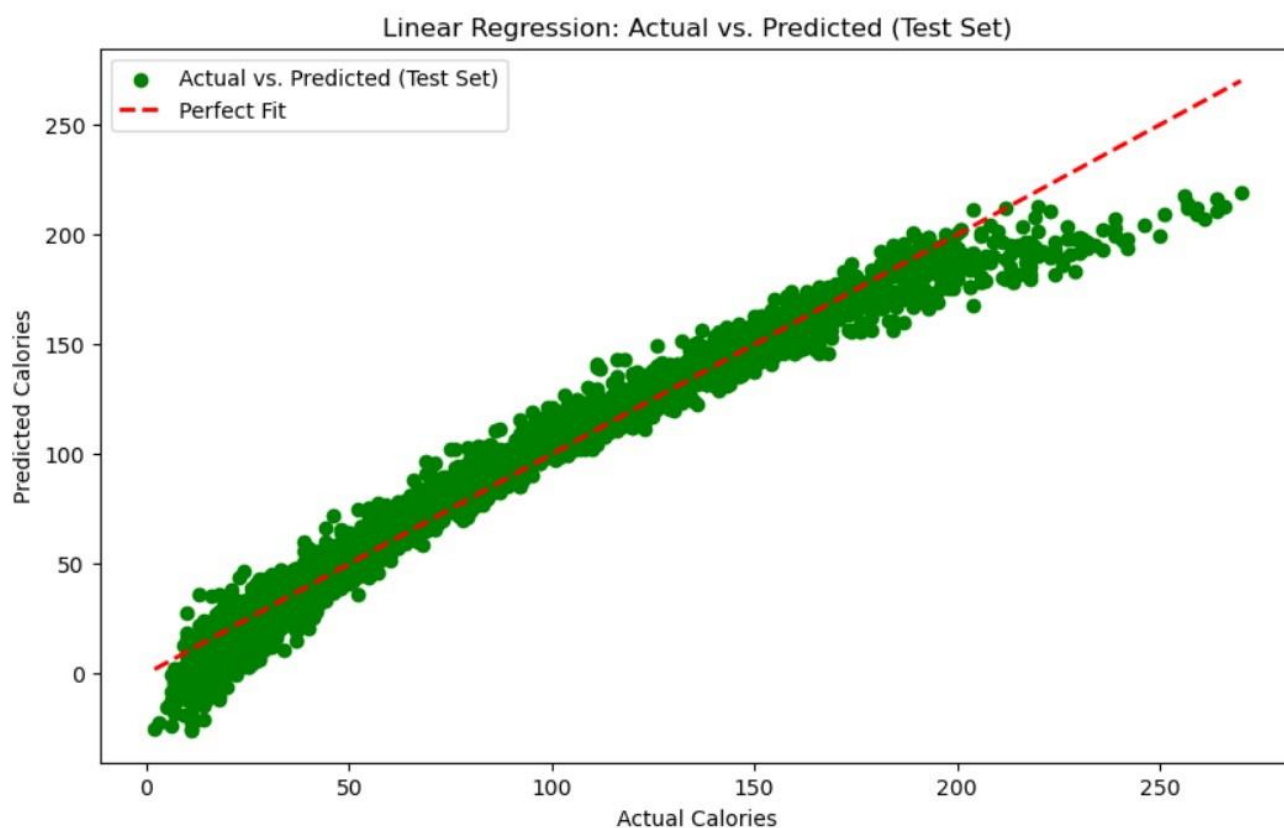
Linear regression:

R2 Score on Test Set: 0.9667

Model Score on Training Set: 0.9668

Mean Absolute Error on Test Set: 7.831

Mean Squared Error on Test Set: 115.202



KNN:

R2 Score on Test Set: 0.989

Model Score on Training Set: 0.994

Mean Absolute Error on Test Set: 4.483

Mean Squared Error on Test Set: 35.602

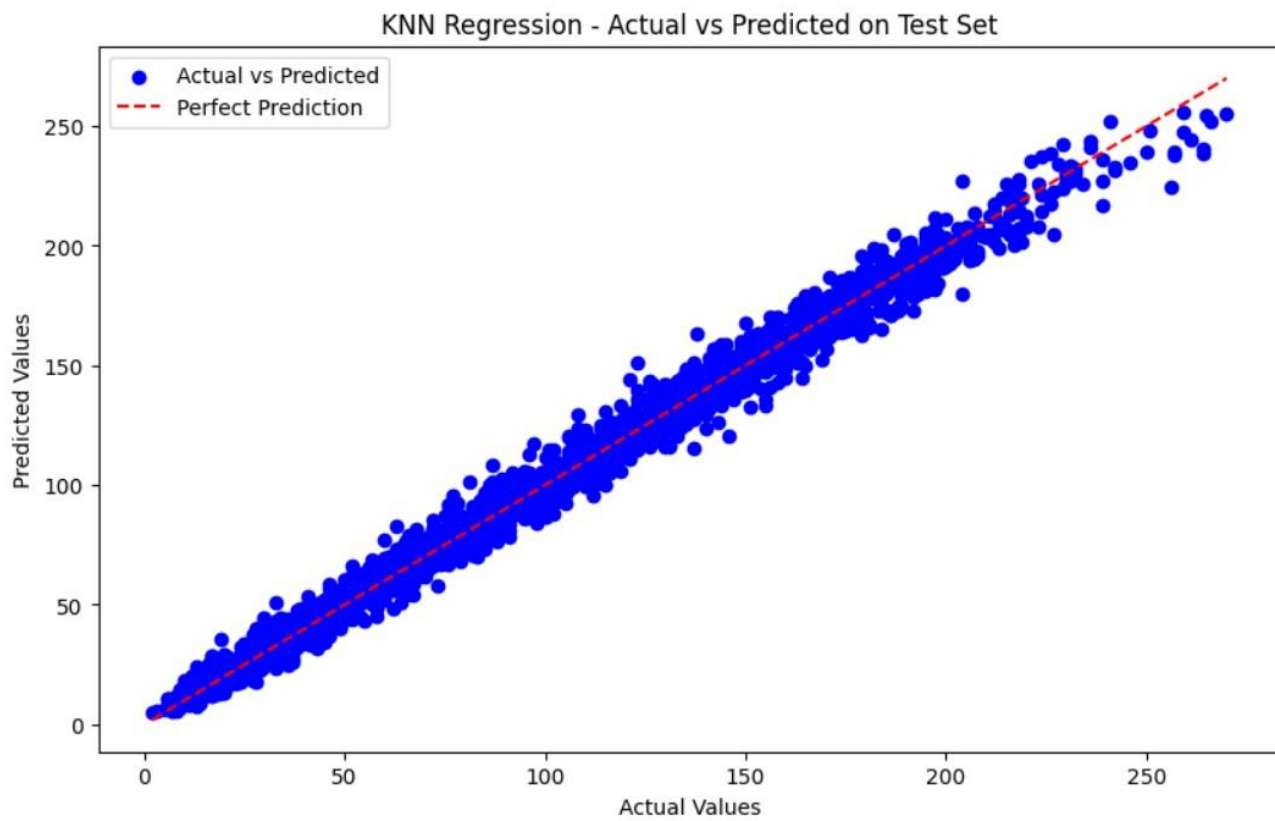


Image dataset

a- General Information on dataset:

Name: **Fashion-MNIST**

Number of classes: **10**

0 T-shirt/top
1 Trouser
2 Pullover
3 Dress
4 Coat
5 Sandal
6 Shirt
7 Sneaker
8 Bag
9 Ankle boot

Total number of samples: **70000**

Size of each mage: **28*28**

Number of samples used in:

- *Training / Validation:* 60,000
- *Testing:* 10,000

b- Implementation details:

Hyperparameters used in :

- Logistic Regression:
LogisticRegression(max_iter=200)
- K means:
n_clusters = 5

C-Results details:

Logistic regression:

accuracy: 0.9488

test predictions: [2 0 1 ... 4 1 2]

