A logo with black text

Description automatically generated

CS770 Machine Learning

Assignment2: BMI Classification using Machine learning.

03/28/2024

Submitted by,

Rajeet Chaudhary

Abstract

Introduction

Methods

Results and Discussion

Conclusions

This data frame contains the following columns:

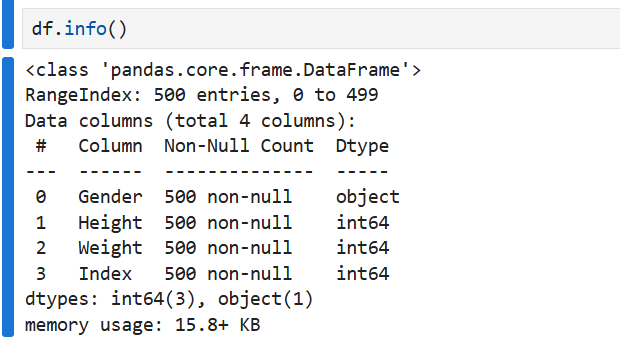
* Gender : Male / Female
* Height : Number (cm)
* Weight : Number (Kg)
* Index :
  + 0 - Extremely Weak
  + 1 – Weak
  + 2 – Normal
  + 3 – Overweight
  + 4 – Obesity
  + 5 - Extreme Obesity

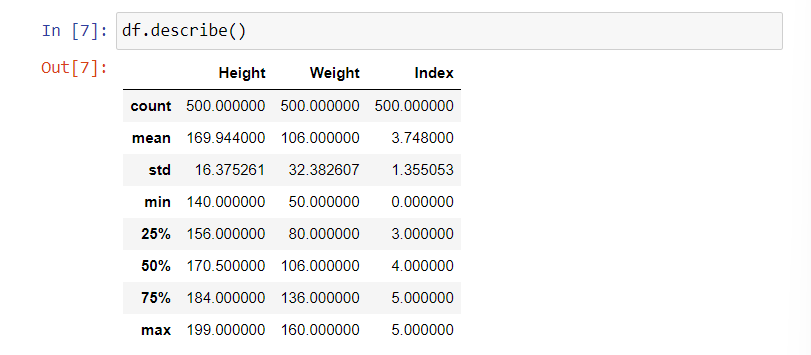
It has 500,4 dimensions

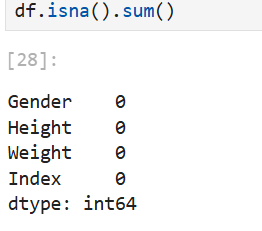
Here Index is the target value and the Gender, Height and Weight are features.

1. **Data Exploration and Cleanup**

It doesn’t contain any missing values.





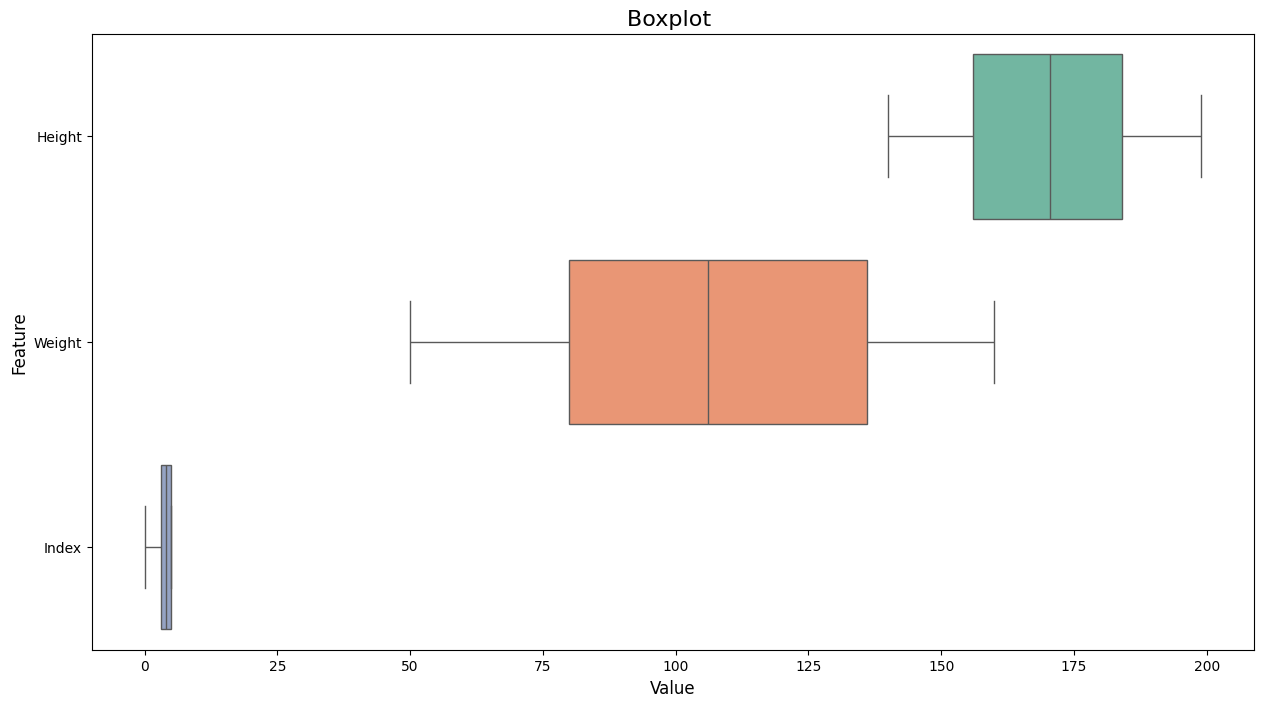


Checking duplicate is not necessary as many can have same height, weight, gender and index.

Now, checking outliers we got to know that:  
Column 'Height' has 0 outliers

Column 'Weight' has 0 outliers

Column 'Index' has 0 outliers

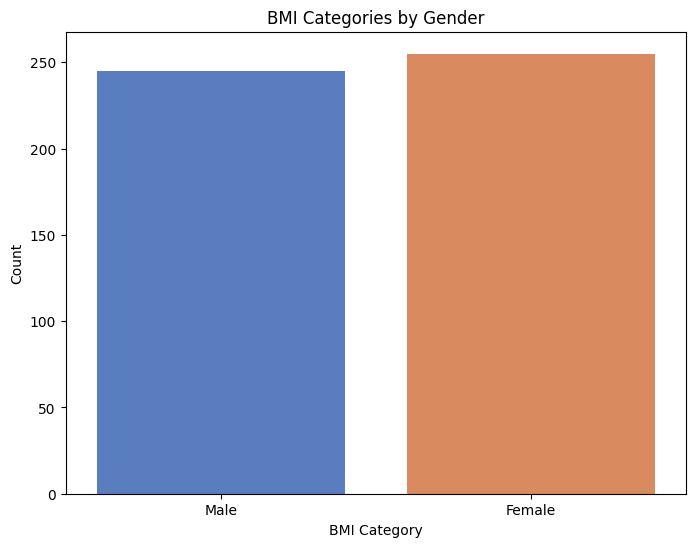


After data is explored, we found no anomalities or missing values that needs to be handled we can move further with Exploratory Data Analysis (EDA)

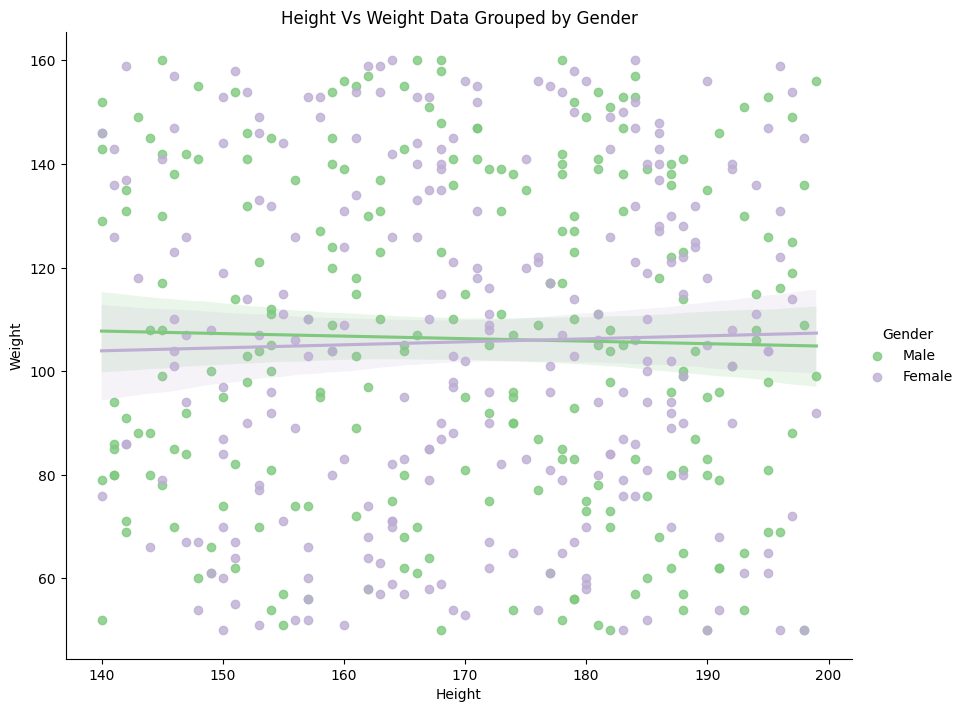
1. **Exploratory Data Analysis (EDA)**

We need to analyze the distribution of BMI categories and explore the relationship between gender, height, weight, and BMI category

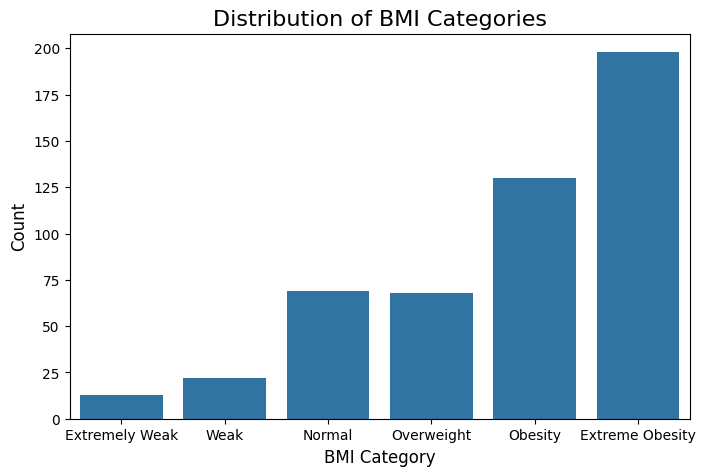
i. Count plot showing the distribution of BMI categories by gender



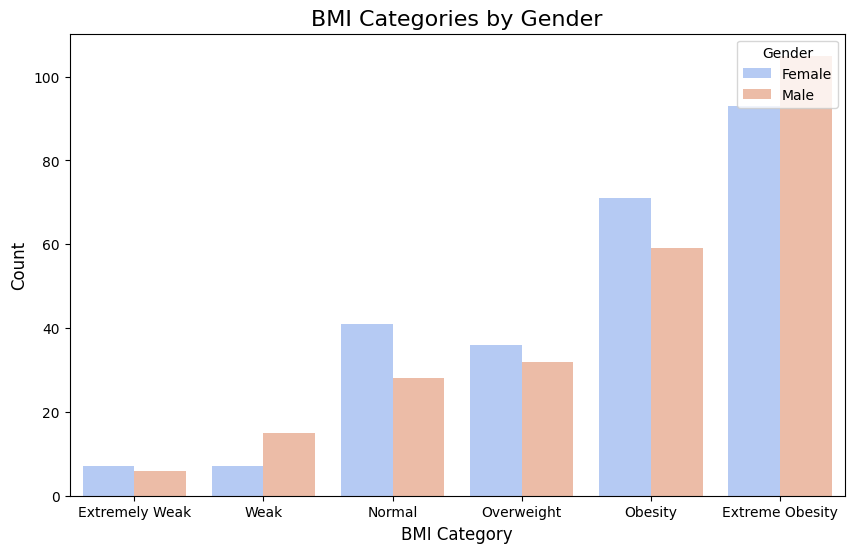
ii. Height Vs Weight Data Grouped by Gender



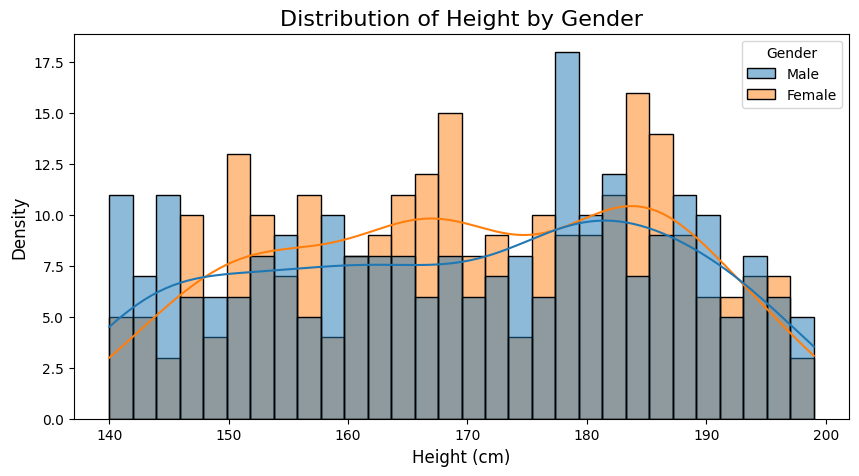
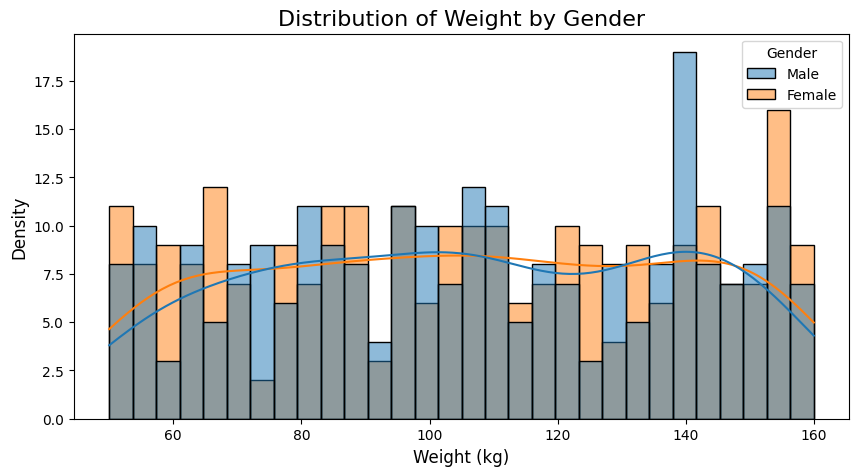
iii. BMI category distribution



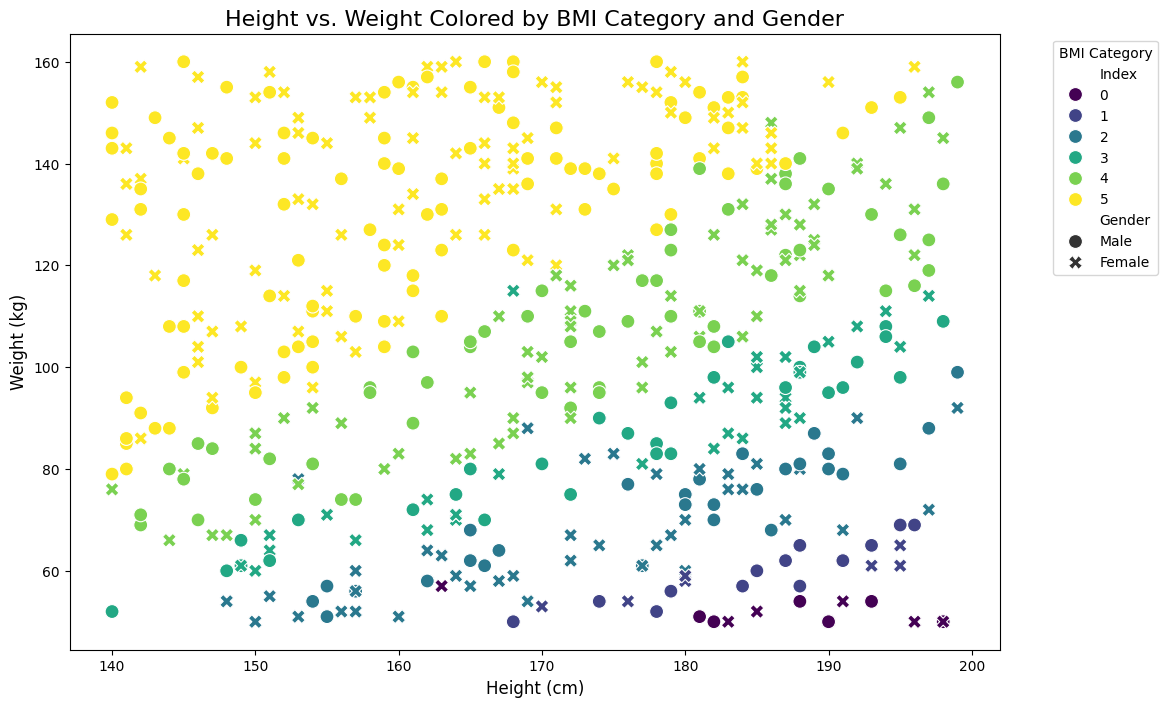
iv. BMI category distribution w.r.t gender



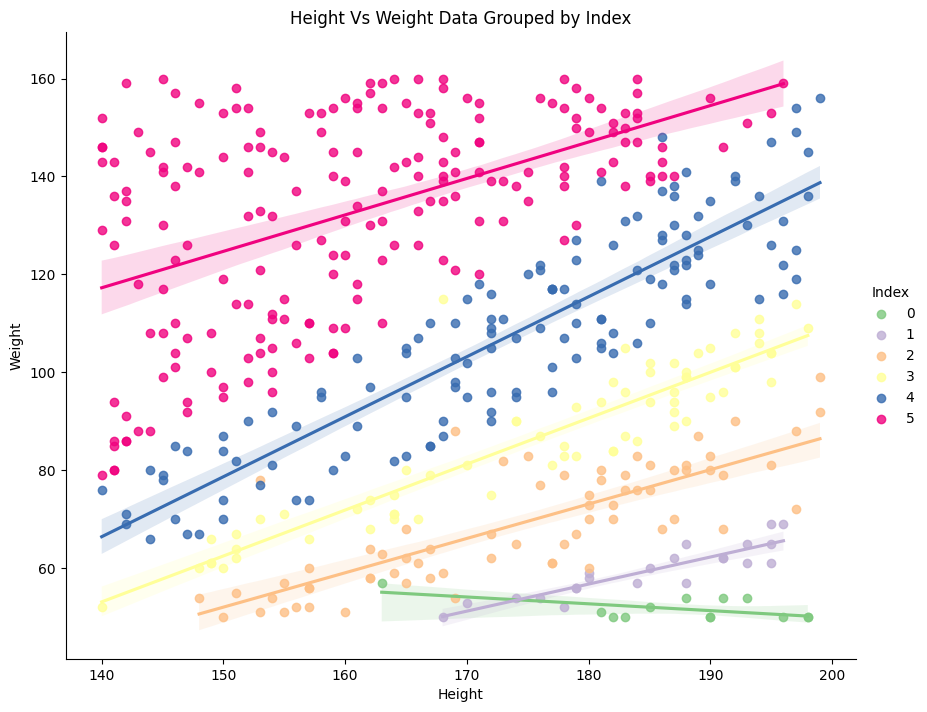
v. Distribution of Height and Weight by Gender



vi. Variation of Height and Weight w.r.t Index and Gender



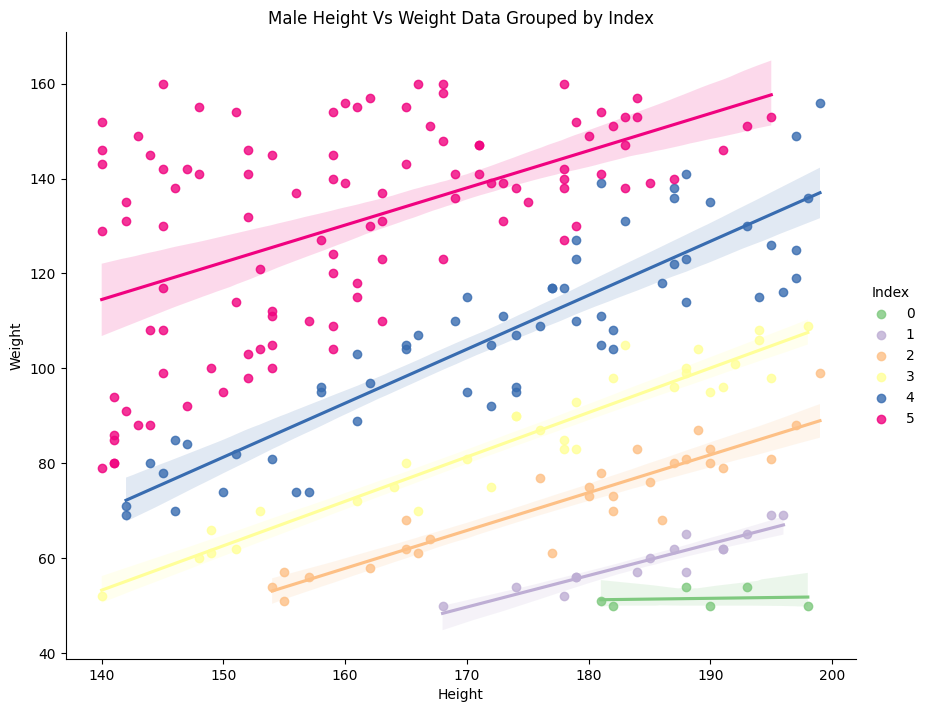
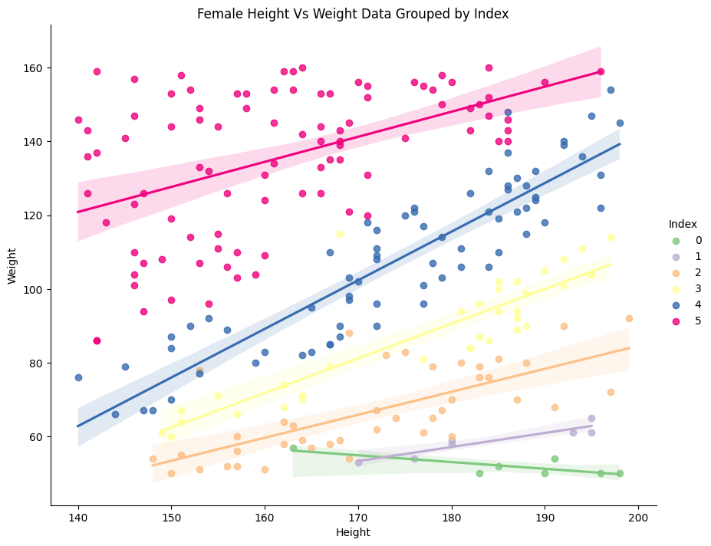
vii. Trend in Index based on relationship between Height and Weight



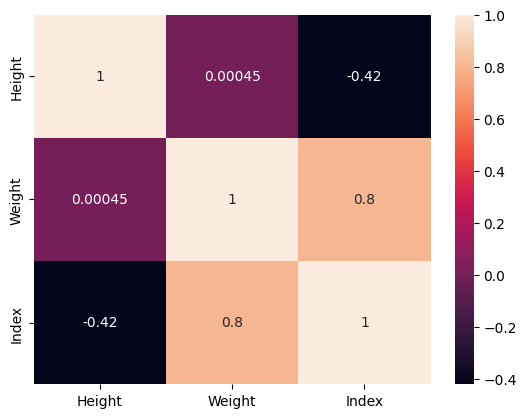
We can make out distinct bands in the data based on the index value.

So, there is a general positive correlation between height and weight when categorized by index value.Now, let us see if there are any discrepencies in the relation when looking at each gender separately.

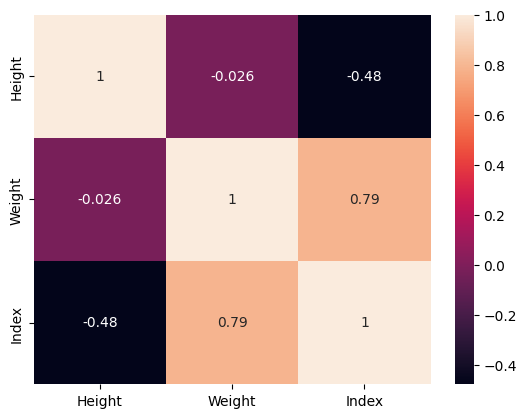
viii. Male and Female Trend in Index based on relationship between Height and Weight



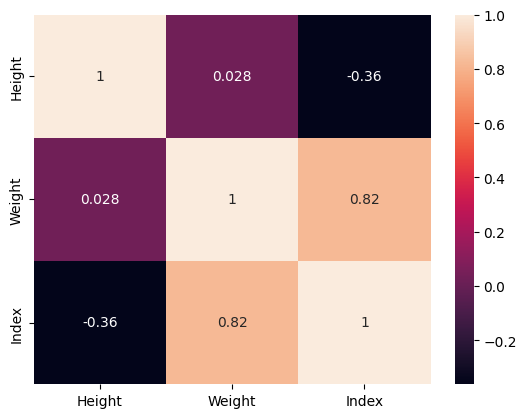
ix. Heatmap



x. Heatmap for Male



xi. Heatmap for female

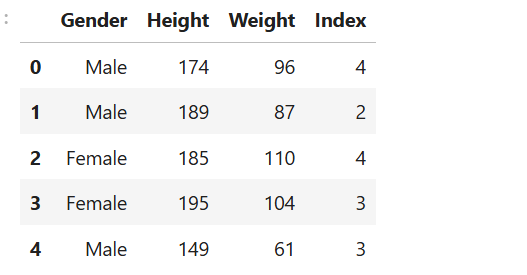
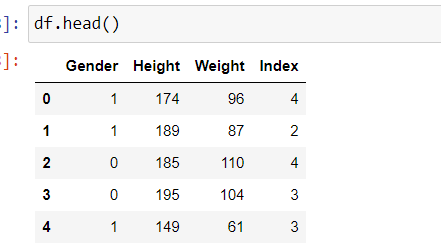


1. **Data Preprocessing**

Also, it can be seen that Gender has categorical value i.e., Male and Female. we need to encode the data for Gender to make it useable using LabelEncoder

Male denoted with 0

Female denoted with 1



We can see in above df.head () that gender are encoded with 0, 1. Now let us move further with feature scaling, building model and improving the model accuracy.

Input features are:

1. **Gender**
2. **Height**
3. **Weight**

Target label is **Index**

We also need to check whether the target label is imbalanced or not.

Index

5 198

4 130

2 69

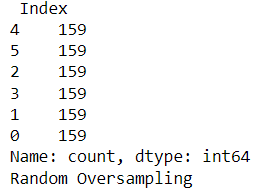
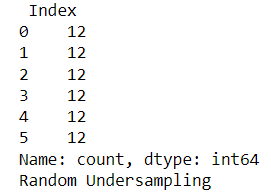
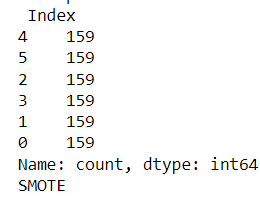
3 68

1 22

0 13

Since, it is not balanced we can use oversampling, undersampling or applying SMOTE. To see which one yield better results, let us use logistic regression and compare the classification reports for all the cases.

Let us first see the resampled index in each of the 3 sampling techniques.

|  | **Model** | **Accuracy** | **F1-Score (Weighted)** | **Precision (Weighted)** | **Recall (Weighted)** |
| --- | --- | --- | --- | --- | --- |
| **0** | Random Oversampling | 0.77 | 0.778316 | 0.806181 | 0.77 |
| **1** | Random Undersampling | 0.74 | 0.736892 | 0.757690 | 0.74 |
| **2** | SMOTE | 0.77 | 0.776847 | 0.795667 | 0.77 |

SMOTE provides a balanced approach with metrics close to Random Oversampling, but slightly lower recall. It prevents overfitting (compared to Random Oversampling) by creating synthetic data instead of duplicating instances, which might explain its stable performance.

As a result, we go further with SMOTE

Also, we need to standardise height and weight during building models to yield better results

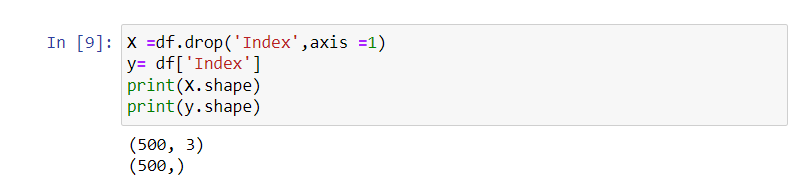
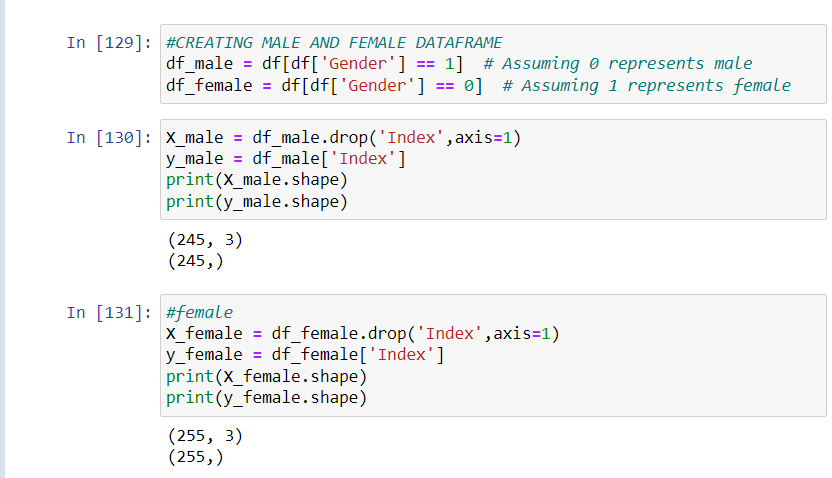
Now, let us standardize the height and weight for better model performance

We will be creating 3 models of logistic regression with

1. Generalised Dataset which we already have
2. Male Dataset, i.e, only the one with Gender Male
3. Female Dataset, i.e, only the one with Gender Female

**SMOTE will be used as a Sampling method StandardScaler will be used for Feature Scaling of height and weight**

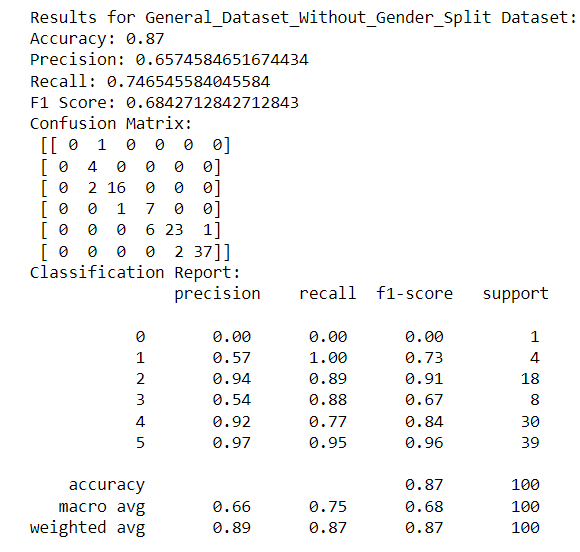
Firstly, let us divide our dataset into training and testing set for all the above 3 cases

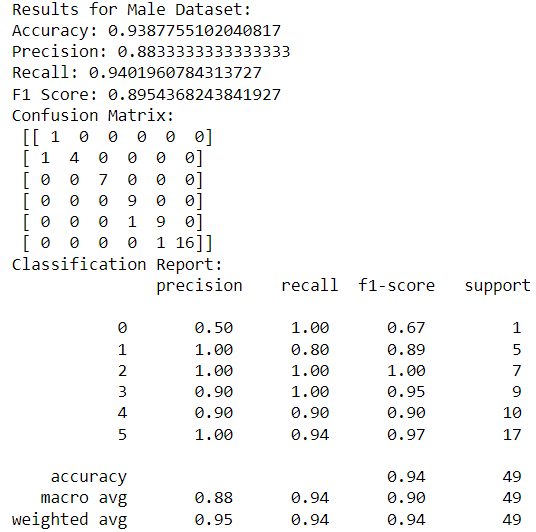
1. **LOGISTIC REGRESSION**

Classification Reports For Logistic Regression

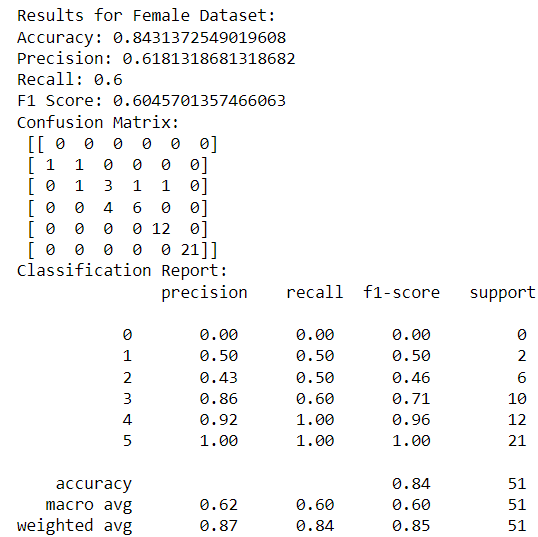
1. General\_Dataset\_Without\_Gender\_Split Dataset



1. Male Dataset



1. Female Dataset



Comparison of performance between 3 cases:

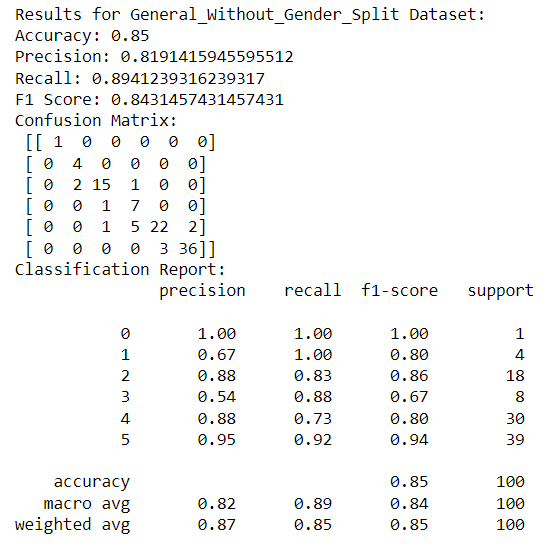
|  | **Dataset** | **Accuracy** | **Precision** | **Recall** | **F1 Score** |
| --- | --- | --- | --- | --- | --- |
| **0** | General\_Dataset\_Without\_Gender\_Split | 0.870000 | 0.657458 | 0.746546 | 0.684271 |
| **1** | Male | 0.938776 | 0.883333 | 0.940196 | 0.895437 |
| **2** | Female | 0.843137 | 0.618132 | 0.600000 | 0.604570 |

1. **K-Nearest Neighbors (KNN)**

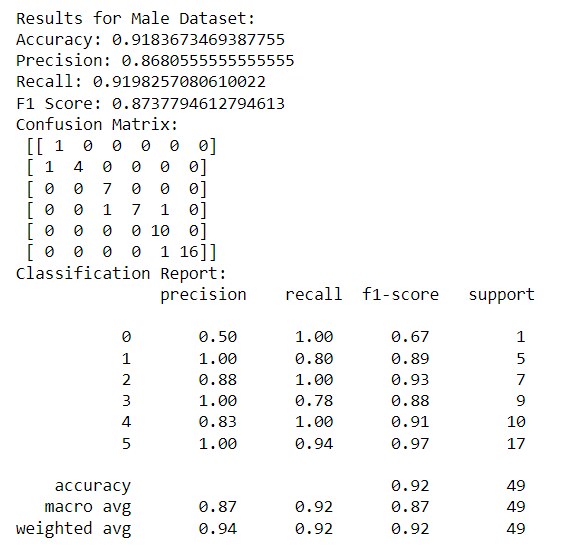
Here, we initially start with no. of neigbors, k=5.

Then similarly we will find the classification report for each of the 3 cases. Lastly we will compare all three of them.

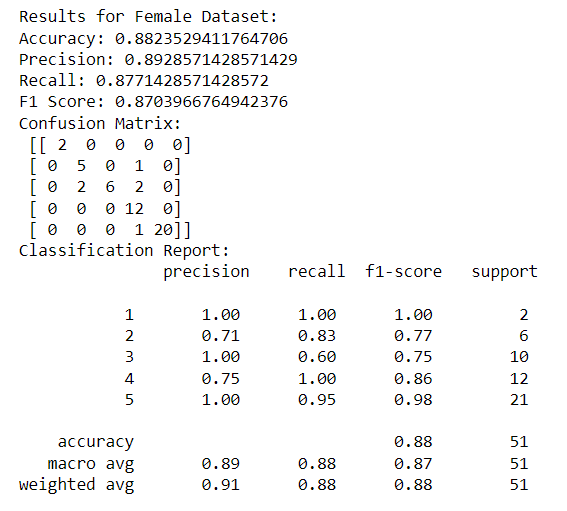
i. General\_Dataset\_Without\_Gender\_Split Dataset



ii. Male Dataset



iii. Female Dataset

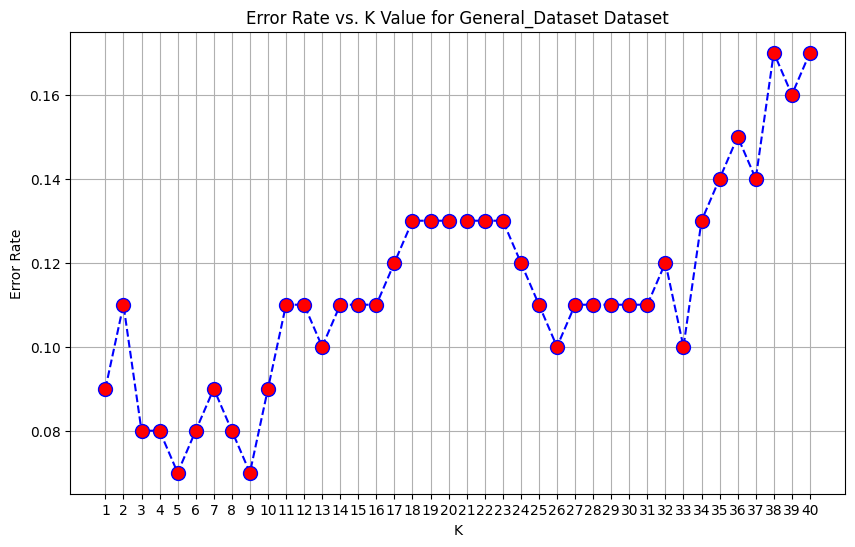


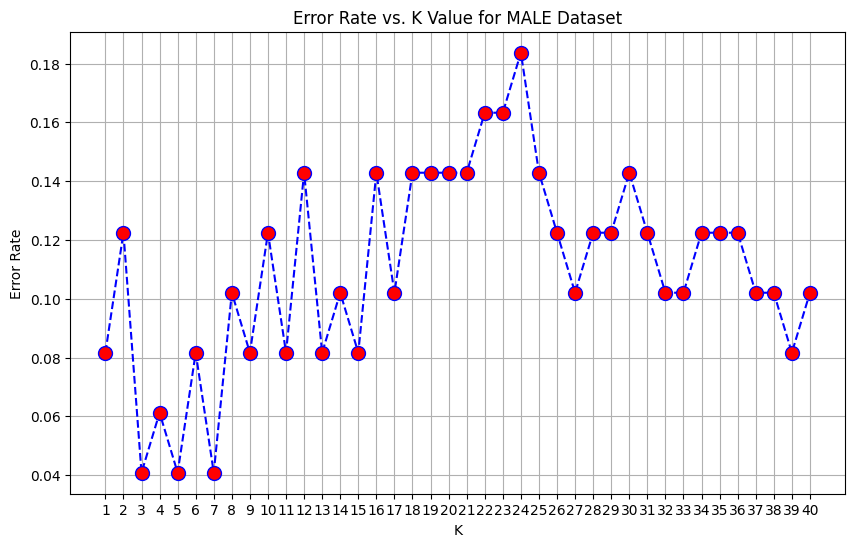
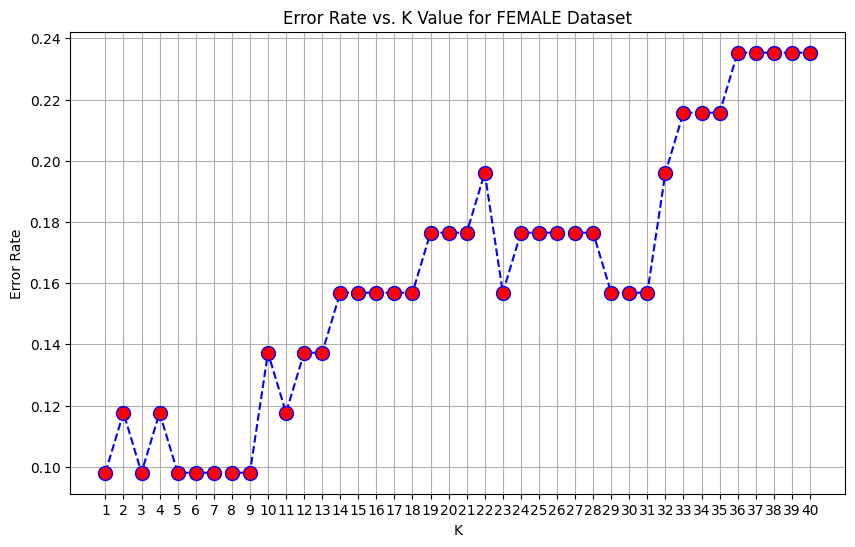
Comparison across each of the 3 models

| **Dataset** | **Accuracy** | **Precision** | **Recall** | **F1 Score** |
| --- | --- | --- | --- | --- |
| General\_Without\_Gender\_Split | 0.850000 | 0.819142 | 0.894124 | 0.843146 |
| Male | 0.918367 | 0.868056 | 0.919826 | 0.873779 |
| Female | 0.882353 | 0.892857 | 0.877143 | 0.870397 |

Now let us find the best k value with the lowest error rate and use that k to generate best classification report with improved accuracy, precision, recall, F1-score and support

**Plot of error rate vs K for each of 3 datsets**



* For General the lowest error rate with k = 5
* For Male the lowest error rate with k = 3
* For Female the lowest error rate with k = 1

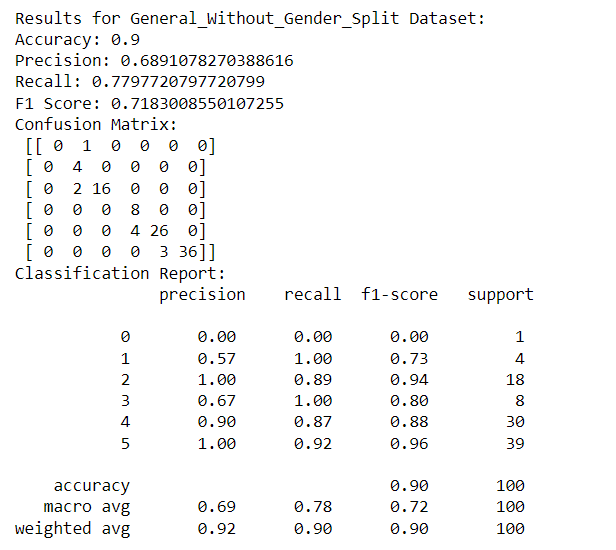
Again, constructing the model and running with each of the different datasets and then,

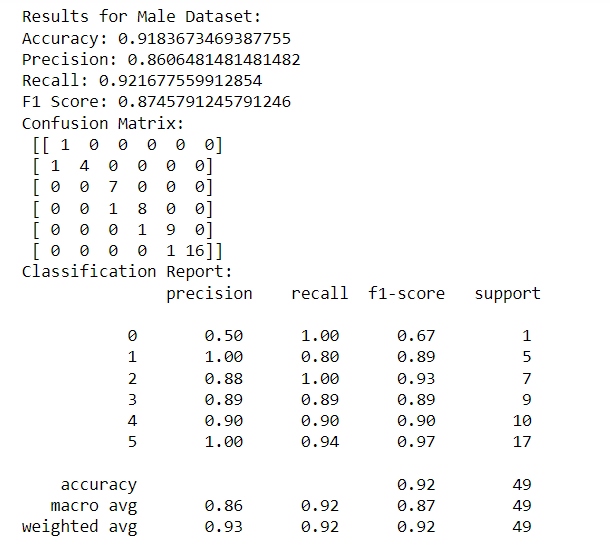
Comparison of Results For KNN with best k value:

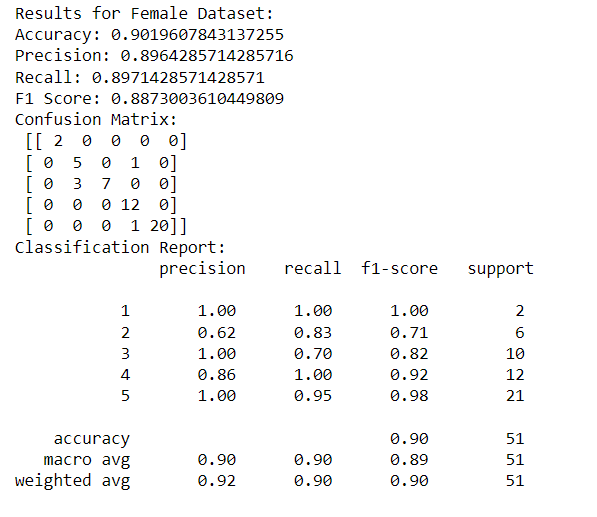
|  | **Dataset** | **Accuracy** | **Precision** | **Recall** | **F1 Score** |
| --- | --- | --- | --- | --- | --- |
| **0** | General\_Without\_Gender\_Split | 0.850000 | 0.819142 | 0.894124 | 0.843146 |
| **1** | Male | 0.938776 | 0.888889 | 0.938344 | 0.895920 |
| **2** | Female | 0.901961 | 0.898810 | 0.890000 | 0.891453 |

* The accuracy for the general dataset did not change after tuning 𝑘, indicating that the overall model performance remained stable with k = 5
* For male, the accuracy improved from 0.918367 to 0.938776 with the best k = 3. The precision and recall also improved, which suggests that the model is better at identifying male instances with fewer false positives and negatives.
* The accuracy improved from 0.882353 to 0.901961 with the best k=1. Precision and recall improved, showing that the model is also more effective for females with this lower k

**Support Vector Machine (SVM)**







Comparison of Results For SVM:

Out[187]:

|  | **Dataset** | **Accuracy** | **Precision** | **Recall** | **F1 Score** |
| --- | --- | --- | --- | --- | --- |
| **0** | General\_Without\_Gender\_Split | 0.900000 | 0.689108 | 0.779772 | 0.718301 |
| **1** | Male | 0.918367 | 0.860648 | 0.921678 | 0.874579 |
| **2** | Female | 0.901961 | 0.896429 | 0.897143 | 0.887300 |

Fitting 5 folds for each of 72 candidates, totalling 360 fits

Results for General\_Without\_Gender\_Split Dataset:

Best Parameters: {'C': 100, 'degree': 2, 'gamma': 'scale', 'kernel': 'rbf'}

Best Cross-validated Score: 0.9466

Test Set Accuracy: 0.9600

Fitting 5 folds for each of 72 candidates, totalling 360 fits

Results for Male Dataset:

Best Parameters: {'C': 100, 'degree': 2, 'gamma': 'scale', 'kernel': 'linear'}

Best Cross-validated Score: 0.9830

Test Set Accuracy: 1.0000

Fitting 5 folds for each of 72 candidates, totalling 360 fits

Results for Female Dataset:

Best Parameters: {'C': 100, 'degree': 2, 'gamma': 'scale', 'kernel': 'rbf'}

Best Cross-validated Score: 0.9376

Test Set Accuracy: 0.9216

Comparison of GridSearchCV Results with hyperparameter tuning:

Out[191]:

|  | **Dataset** | **Best Parameters** | **Cross-validated Score** | **Test Set Accuracy** |
| --- | --- | --- | --- | --- |
| **0** | General\_Without\_Gender\_Split | {'C': 100, 'degree': 2, 'gamma': 'scale', 'ker... | 0.946558 | 0.960000 |
| **1** | Male | {'C': 100, 'degree': 2, 'gamma': 'scale', 'ker... | 0.983001 | 1.000000 |
| **2** | Female | {'C': 100, 'degree': 2, 'gamma': 'scale', 'ker... | 0.937557 | 0.921569 |

score support

0 0.00 0.00 0.00 1

1 0.75 0.75 0.75 4

2 0.89 0.89 0.89 18

3 0.78 0.88 0.82 8

4 0.97 0.93 0.95 30

5 0.95 0.97 0.96 39

accuracy 0.92 100

macro avg 0.72 0.74 0.73 100

weighted avg 0.91 0.92 0.92 100

Results for Male Dataset:

Accuracy: 0.803921568627451

Precision: 0.6409090909090909

Recall: 0.6266666666666667

F1 Score: 0.5813833225597931

Confusion Matrix:

[[ 0 2 0 0 0]

[ 0 5 0 1 0]

[ 0 4 3 3 0]

[ 0 0 0 12 0]

[ 0 0 0 0 21]]

Classification Report:

precision recall f1-score support

1 0.00 0.00 0.00 2

2 0.45 0.83 0.59 6

3 1.00 0.30 0.46 10

4 0.75 1.00 0.86 12

5 1.00 1.00 1.00 21

accuracy 0.80 51

macro avg 0.64 0.63 0.58 51

weighted avg 0.84 0.80 0.77 51

Results for Female Dataset:

Accuracy: 0.7959183673469388

Precision: 0.5496031746031746

Recall: 0.5925925925925926

F1 Score: 0.5474101921470342

Confusion Matrix:

[[ 0 1 0 0 0 0]

[ 0 0 5 0 0 0]

[ 0 0 7 0 0 0]

[ 0 0 0 5 4 0]

[ 0 0 0 0 10 0]

[ 0 0 0 0 0 17]]

Classification Report:

precision recall f1-score support

0 0.00 0.00 0.00 1

1 0.00 0.00 0.00 5

2 0.58 1.00 0.74 7

3 1.00 0.56 0.71 9

4 0.71 1.00 0.83 10

5 1.00 1.00 1.00 17

accuracy 0.80 49

macro avg 0.55 0.59 0.55 49

weighted avg 0.76 0.80 0.75 49

|  | **Dataset** | **Accuracy** | **Precision** | **Recall** | **F1 Score** |
| --- | --- | --- | --- | --- | --- |
| **0** | General\_Without\_Gender\_Split | 0.920000 | 0.722031 | 0.736930 | 0.728933 |
| **1** | Male | 0.803922 | 0.640909 | 0.626667 | 0.581383 |
| **2** | Female | 0.795918 | 0.549603 | 0.592593 | 0.547410 |

KNN  
First, I found the classification reports for generalized(combined male and female),male and female dataset with KNN algorithm with initially no. of neighbors =5.

The results were:

Results for General\_Without\_Gender\_Split Dataset:

Accuracy: 0.89

Precision: 0.697198797934092

Recall: 0.772934472934473

F1 Score: 0.7250404492243266

Confusion Matrix:

[[ 0 1 0 0 0 0]

[ 0 4 0 0 0 0]

[ 0 1 16 1 0 0]

[ 0 0 0 8 0 0]

[ 0 0 1 2 24 3]

[ 0 0 0 0 2 37]]

Classification Report:

precision recall f1-score support

0 0.00 0.00 0.00 1

1 0.67 1.00 0.80 4

2 0.94 0.89 0.91 18

3 0.73 1.00 0.84 8

4 0.92 0.80 0.86 30

5 0.93 0.95 0.94 39

accuracy 0.89 100

macro avg 0.70 0.77 0.73 100

weighted avg 0.89 0.89 0.89 100

Results for Male Dataset:

Accuracy: 0.8431372549019608

Precision: 0.6599999999999999

Recall: 0.6771428571428573

F1 Score: 0.64789972899729

Confusion Matrix:

[[ 0 2 0 0 0]

[ 0 5 0 1 0]

[ 0 3 6 1 0]

[ 0 0 0 12 0]

[ 0 0 0 1 20]]

Classification Report:

precision recall f1-score support

1 0.00 0.00 0.00 2

2 0.50 0.83 0.62 6

3 1.00 0.60 0.75 10

4 0.80 1.00 0.89 12

5 1.00 0.95 0.98 21

accuracy 0.84 51

macro avg 0.66 0.68 0.65 51

weighted avg 0.85 0.84 0.83 51

Results for Female Dataset:

Accuracy: 0.9387755102040817

Precision: 0.9513888888888888

Recall: 0.9444444444444443

F1 Score: 0.9404040404040405

Confusion Matrix:

[[ 1 0 0 0 0 0]

[ 0 5 0 0 0 0]

[ 0 0 7 0 0 0]

[ 0 0 1 6 2 0]

[ 0 0 0 0 10 0]

[ 0 0 0 0 0 17]]

Classification Report:

precision recall f1-score support

0 1.00 1.00 1.00 1

1 1.00 1.00 1.00 5

2 0.88 1.00 0.93 7

3 1.00 0.67 0.80 9

4 0.83 1.00 0.91 10

5 1.00 1.00 1.00 17

accuracy 0.94 49

macro avg 0.95 0.94 0.94 49

weighted avg 0.95 0.94 0.94 49

Comparison of Results:

|  | **Dataset** | **Accuracy** | **Precision** | **Recall** | **F1 Score** |
| --- | --- | --- | --- | --- | --- |
| **0** | General\_Without\_Gender\_Split | 0.890000 | 0.697199 | 0.772934 | 0.725040 |
| **1** | Male | 0.843137 | 0.660000 | 0.677143 | 0.647900 |
| **2** | Female | 0.938776 | 0.951389 | 0.944444 | 0.940404 |

Now, again let us check the error rate vs k for different datasets and update k so that accuracy can be improved further.

With the plot we can see the the lowest error rate is for

K=3, for generalized dataset

K=1, for male dataset

K=3, for female dataset

The results were:

Results for General\_Without\_Gender\_Split Dataset:

Accuracy: 0.9

Precision: 0.7036575221367186

Recall: 0.7772079772079773

F1 Score: 0.7298530416951469

Confusion Matrix:

[[ 0 1 0 0 0 0]

[ 0 4 0 0 0 0]

[ 0 1 16 1 0 0]

[ 0 0 0 8 0 0]

[ 0 0 1 2 24 3]

[ 0 0 0 0 1 38]]

Classification Report:

precision recall f1-score support

0 0.00 0.00 0.00 1

1 0.67 1.00 0.80 4

2 0.94 0.89 0.91 18

3 0.73 1.00 0.84 8

4 0.96 0.80 0.87 30

5 0.93 0.97 0.95 39

accuracy 0.90 100

macro avg 0.70 0.78 0.73 100

weighted avg 0.90 0.90 0.90 100

Results for Male Dataset:

Accuracy: 0.8823529411764706

Precision: 0.875

Recall: 0.7899999999999999

F1 Score: 0.811965811965812

Confusion Matrix:

[[ 1 1 0 0 0]

[ 0 5 0 1 0]

[ 0 1 7 2 0]

[ 0 0 1 11 0]

[ 0 0 0 0 21]]

Classification Report:

precision recall f1-score support

1 1.00 0.50 0.67 2

2 0.71 0.83 0.77 6

3 0.88 0.70 0.78 10

4 0.79 0.92 0.85 12

5 1.00 1.00 1.00 21

accuracy 0.88 51

macro avg 0.88 0.79 0.81 51

weighted avg 0.89 0.88 0.88 51

Results for Female Dataset:

Accuracy: 0.9387755102040817

Precision: 0.9513888888888888

Recall: 0.9531590413943355

F1 Score: 0.9478535353535354

Confusion Matrix:

[[ 1 0 0 0 0 0]

[ 0 5 0 0 0 0]

[ 0 0 7 0 0 0]

[ 0 0 1 7 1 0]

[ 0 0 0 0 10 0]

[ 0 0 0 0 1 16]]

Classification Report:

precision recall f1-score support

0 1.00 1.00 1.00 1

1 1.00 1.00 1.00 5

2 0.88 1.00 0.93 7

3 1.00 0.78 0.88 9

4 0.83 1.00 0.91 10

5 1.00 0.94 0.97 17

accuracy 0.94 49

macro avg 0.95 0.95 0.95 49

weighted avg 0.95 0.94 0.94 49

Comparison of Results:

|  | **Dataset** | **Accuracy** | **Precision** | **Recall** | **F1 Score** |
| --- | --- | --- | --- | --- | --- |
| **0** | General\_Without\_Gender\_Split | 0.900000 | 0.703658 | 0.777208 | 0.729853 |
| **1** | Male | 0.882353 | 0.875000 | 0.790000 | 0.811966 |
| **2** | Female | 0.938776 | 0.951389 | 0.953159 | 0.947854 |

From the previous table, we can see that the performance of KNN model has improved.

For example in the male dataset, the accuracy has increased the highest. While for other, recall, precision, and F1-score has slightly improved.

SVM

The results were:

Results for General\_Without\_Gender\_Split Dataset:

Accuracy: 0.89

Precision: 0.6715399610136452

Recall: 0.7423789173789174

F1 Score: 0.6966057757132808

Confusion Matrix:

[[ 0 1 0 0 0 0]

[ 0 4 0 0 0 0]

[ 0 2 16 0 0 0]

[ 0 0 2 6 0 0]

[ 0 0 0 3 26 1]

[ 0 0 0 0 2 37]]

Classification Report:

precision recall f1-score support

0 0.00 0.00 0.00 1

1 0.57 1.00 0.73 4

2 0.89 0.89 0.89 18

3 0.67 0.75 0.71 8

4 0.93 0.87 0.90 30

5 0.97 0.95 0.96 39

accuracy 0.89 100

macro avg 0.67 0.74 0.70 100

weighted avg 0.89 0.89 0.89 100

Results for Male Dataset:

Accuracy: 0.8235294117647058

Precision: 0.6547619047619048

Recall: 0.6466666666666667

F1 Score: 0.61001221001221

Confusion Matrix:

[[ 0 2 0 0 0]

[ 0 5 0 1 0]

[ 0 5 4 1 0]

[ 0 0 0 12 0]

[ 0 0 0 0 21]]

Classification Report:

precision recall f1-score support

1 0.00 0.00 0.00 2

2 0.42 0.83 0.56 6

3 1.00 0.40 0.57 10

4 0.86 1.00 0.92 12

5 1.00 1.00 1.00 21

accuracy 0.82 51

macro avg 0.65 0.65 0.61 51

weighted avg 0.86 0.82 0.81 51

Results for Female Dataset:

Accuracy: 0.9183673469387755

Precision: 0.7555555555555555

Recall: 0.7743386243386242

F1 Score: 0.7623788492209544

Confusion Matrix:

[[ 0 1 0 0 0 0]

[ 0 5 0 0 0 0]

[ 0 0 6 1 0 0]

[ 0 0 0 8 1 0]

[ 0 0 0 1 9 0]

[ 0 0 0 0 0 17]]

Classification Report:

precision recall f1-score support

0 0.00 0.00 0.00 1

1 0.83 1.00 0.91 5

2 1.00 0.86 0.92 7

3 0.80 0.89 0.84 9

4 0.90 0.90 0.90 10

5 1.00 1.00 1.00 17

accuracy 0.92 49

macro avg 0.76 0.77 0.76 49

weighted avg 0.91 0.92 0.91 49

Comparison of Results:

|  | **Dataset** | **Accuracy** | **Precision** | **Recall** | **F1 Score** |
| --- | --- | --- | --- | --- | --- |
| **0** | General\_Without\_Gender\_Split | 0.890000 | 0.671540 | 0.742379 | 0.696606 |
| **1** | Male | 0.823529 | 0.654762 | 0.646667 | 0.610012 |
| **2** | Female | 0.918367 | 0.755556 | 0.774339 | 0.762379 |

Lets to hyperparameter tining for each of different datasets to see if performance can be further improved or not:

Generalised :

Fitting 5 folds for each of 72 candidates, totalling 360 fits

Best parameters: {'C': 100, 'degree': 2, 'gamma': 'scale', 'kernel': 'linear'}

Best cross-validated score: 0.9375

Accuracy on test set: 0.94

Male:

Fitting 5 folds for each of 72 candidates, totalling 360 fits

Best parameters: {'C': 10, 'degree': 2, 'gamma': 'scale', 'kernel': 'linear'}

Best cross-validated score: 0.9024390243902438

Accuracy on test set: 0.9019607843137255

Female:

Fitting 5 folds for each of 72 candidates, totalling 360 fits

Best parameters: {'C': 100, 'degree': 2, 'gamma': 'scale', 'kernel': 'linear'}

Best cross-validated score: 0.9592307692307692

Accuracy on test set: 0.9795918367346939

We can see for each of the datasets with hyperparameter tuning, performance seems to improve.

Comparison of GridSearchCV Results:

|  | **Dataset** | **Best Parameters** | **Cross-validated Score** | **Test Set Accuracy** |
| --- | --- | --- | --- | --- |
| **0** | General\_Without\_Gender\_Split | {'C': 100, 'degree': 2, 'gamma': 'scale', 'ker... | 0.937500 | 0.940000 |
| **1** | Male | {'C': 10, 'degree': 2, 'gamma': 'scale', 'kern... | 0.902439 | 0.901961 |
| **2** | Female | {'C': 100, 'degree': 2, 'gamma': 'scale', 'ker... | 0.959231 | 0.979592 |

It can be inferred that the general model performs well achieving both high cross-validation and test accuracy.It treats male and female genders uniformly, which could hide specific gender-based nuances.

The male-specific model is slightly less accurate than the general model. This suggests that a gender-agnostic model performs comparably or slightly better for male predictions. However, this could also indicate limited male-specific patterns. The female-specific model outperforms both the general and male models, achieving the highest test accuracy. This indicates that female-specific data benefits from gender-specific modeling and contains patterns that a general model might not capture well.

 **Regularization Parameter CCC**:

* Higher C=100C = 100C=100 for the general and female datasets suggests that these models need to fit the data more closely (lower bias, more focus on correct classifications).
* A lower C=10C = 10C=10 for the male model indicates that a more generalized model (with some tolerance for misclassifications) performs better for male data.

 **Kernel Selection**:

* **Polynomial kernel**: Chosen for both the general and female datasets, suggesting these datasets benefit from capturing complex non-linear relationships.
* **RBF kernel**: Works best for the male dataset, suggesting that male data patterns might be more smoothly distributed.

This analysis suggests that gender-specific modeling may be beneficial, especially for female data, while for male data, the benefits are less pronounced.