CS770 Machine Learning

Assignment2: BMI Classification Based on Gender, Height, and Weight Using Machine Learning

10/30/2024

Submitted by:
Buddha Thapa Magar
E923Q669

Contents

List of Tables

List of Figures

1 Introduction

Classification is a supervised machine learning model. It is used for the prediction of categorical data points. E.g.: Distinguishing the Gender of the person, Distinguishing the BMI index of the person, etc. Initially, the classification model will be trained using the predefined labeled training datasets with input features and corresponding target labels. After that, the classification model will learn patterns from the given labeled training datasets to associate them with specific output categories. After training, the model evaluation will be performed to measure the performance parameters e.g. Accuracy, Precision, Recall, F1-Score, etc. After that model is ready and given new input, it will assign it to one of the learned categories.

There are many classification algorithms available. 'Logistic Regression', 'K-Nearest Neighbor', and 'Support Vector Machine' are among the most popular methods for this task, each with its strengths and limitations.

1.1 Logistic Regression

Logistic Regression is a popular supervised learning algorithm used for classification tasks. Despite its name, it is used primarily for binary and multiclass classification, not regression. Logistic Regression predicts the probability that an instance belongs to a particular class and converts it into a class label using the logistic (sigmoid) function.

Logistic Regression is one of the popular supervised learning classification algorithms. It is mainly used for binary classification and multiclass classification. Logistic regression uses a sigmoid function to predict the probability that an instance belongs to a particular class or not and based on the probability calculated, it assigns the instance to the class.

$$sigmoid(z) = \frac{1}{1+e^{-z}} \tag{1}$$

1.2 K-Nearest Neighbors

K-Nearest Neighbors (KNN) is a simple but powerful supervised machine-learning algorithm. It can be used for both classification and regression tasks. For a given data instance, KNN will identify the K closest data points from the training datasets, and based on distance metrics like Euclidean, or Manhattan distance, it will assign the data instance to the most common class among its K neighbors. Choosing the right value of K is critical. If we choose a small value then it can cause overfitting and if we choose a large K then it can cause the model to underfit. It is a non-parametric lazy learning algorithm. It classifies instances based on their proximity to training examples in the feature space.

1.3 Support Vector Machine

Support Vector Machine (SVM) is a supervised machine learning algorithm mainly used for classification tasks, but it can be used in regression as well. This algorithm mainly focuses on finding the optimal hyperplane that separates data categorical classes into feature space.

2 Classification Performance Matrices

There are many parameters to measure the performance of the classification machine learning models. Some of the commonly used performance matrices are listed below:

- Accuracy
- Precision
- Recall
- F1-Score

2.1 Accuracy

Accuracy is defined by the ratio of the total correctly classified data instances to the total instances.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

$$\begin{aligned} &\text{where,} \\ &\text{TP} = \text{True Positives}, \\ &\text{TN} = \text{True Negatives}, \\ &\text{FP} = \text{False Positives}, \\ &\text{FN} = \text{False Negatives} \end{aligned}$$

2.2 Precision

Precision is defined as the ratio of true positive classified instances to the total predicted positives instances.

$$\text{Accuracy} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

where,

$$\text{TP} = \text{True Positives},$$

$$\text{FP} = \text{False Positives}$$

2.3 Recall

Recall is defined as the ratio of number of true positive classified instances to the total number of actual positives instances.

$$\text{Accuracy} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

where,

$$\text{TP} = \text{True Positives},$$

$$\text{FN} = \text{False Negatives}$$

2.4 F1-Score

Recall is defined as the harmonic mean of the precision and recall.

$$\text{F1-Score} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$$

3 Data Exploration and Cleanup

We are using the BMI dataset to conduct classification machine learning models. There are a total of 500 records. It has 3 independent features and one target variable. Feature lists are given below.

- Gender: It is categorical data. It has 2 distinct classes. They are ”Male” and ”Female”.
- Height: It is numerical data.
- Weight: It is numerical data.

The Target Variable name is:

- Index: It is the BMI index. It has 6 distinct classes and they are 0, 1, 2, 3, 4, and 5.

Upon analyzing the dataset, There are no missing records in the dataset. A descriptive analysis of the dataset is mentioned in the table below:

	Height	Weight	Index
count	500.000000	500.000000	500.000000
mean	169.944000	106.000000	3.748000
std	16.375261	32.382607	1.355053
min	140.000000	50.000000	0.000000
255075max	199.000000	160.000000	5.000000

Table 1: Descriptive analysis of BMI Dataset

Outlier detection is performed using interquartile range calculation in "Height" and "Weight" features. We are not performing outlier detection in the "Gender" and "Index" features as these features contain categorical records.

Formula used:

$$lowerlimit = Q_1 - 1.5 * InterQuartileRange$$
$$upperlimit = Q_3 + 1.5 * InterQuartileRange$$

The "Height" and "Weight" fields do not contain any outlier records.

4 Exploratory Data Analysis (EDA)

In this analysis, we conducted Exploratory Data Analysis (EDA) on a dataset comprising Gender, Height, Weight, and BMI index to uncover patterns and relationships among these variables.

We are performing a pair plot for the dataset. This is done to identify the relationship between features and target variables. In this figure, Gender is kept as a hue parameter to distinguish all the relationships on the basis of gender as well.

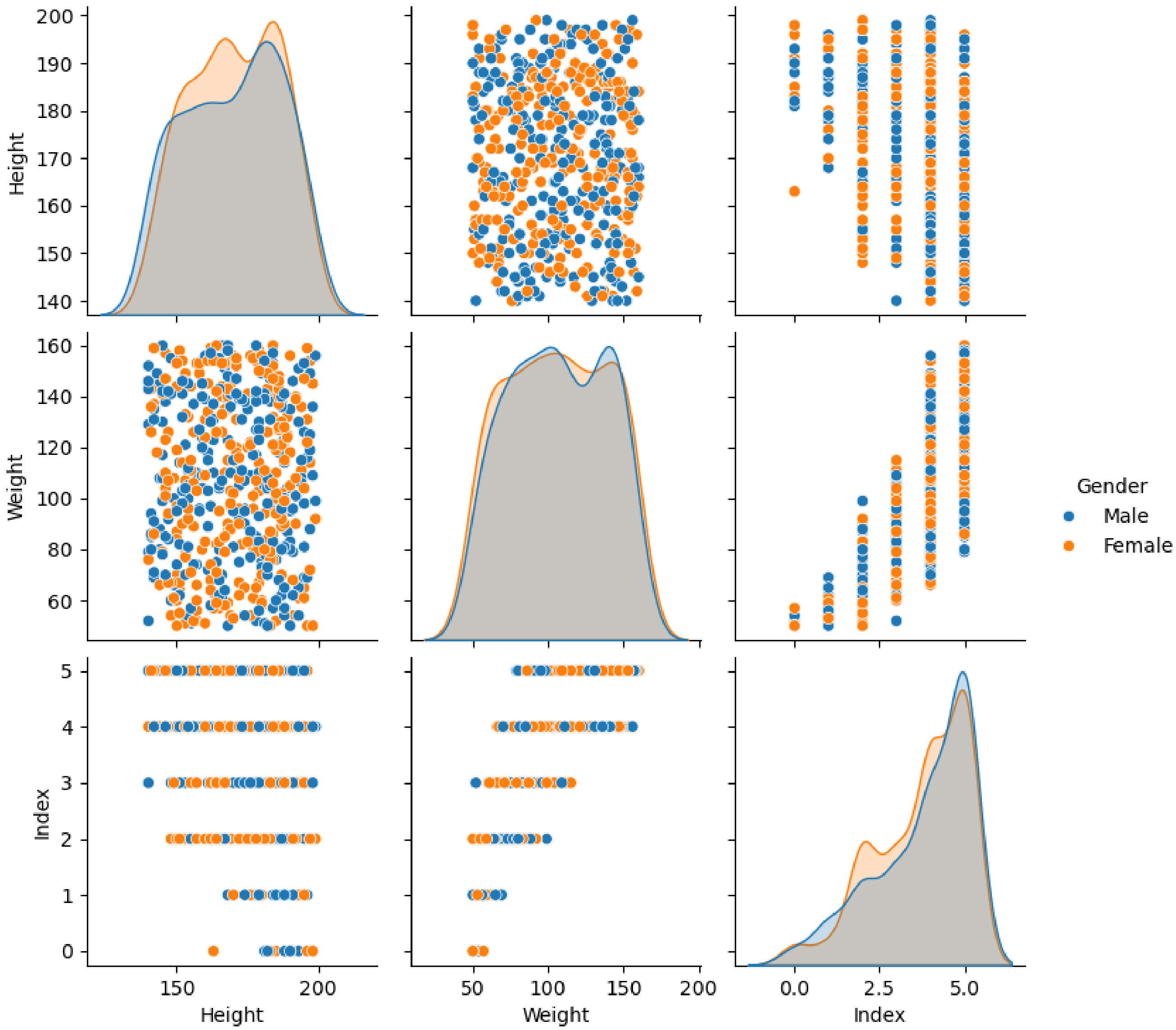

Figure 1: pair plot BMI dataset

While counting the records based on gender, male and female frequencies are almost similar. This means there is no imbalance in the data set based on gender.

Gender	Count
Female	255
Male	245

Table 2: Gender and frequency count

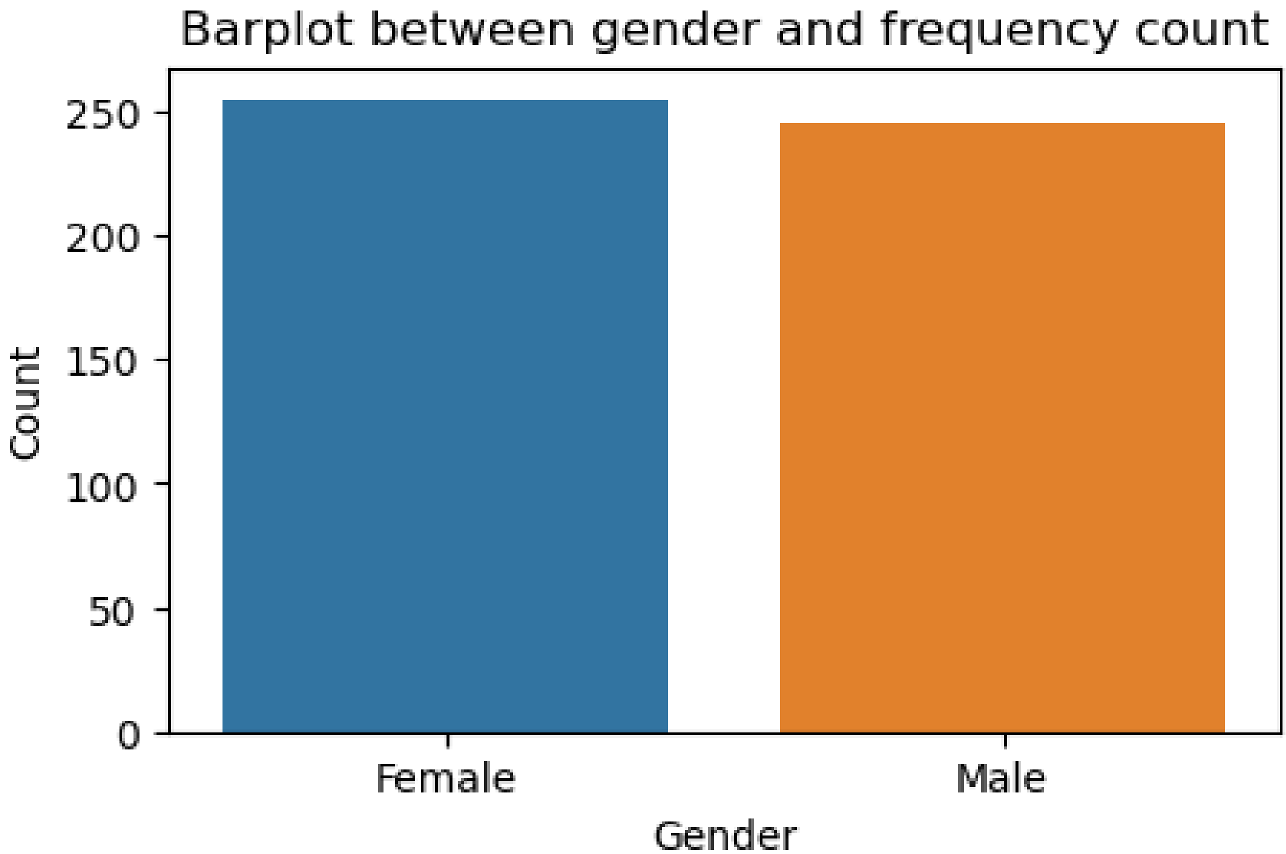

Figure 2: Barplot between gender and frequency count

4.1 Analyisis of Height feature

Analysis of height is performed. The distribution of height based on gender seems identical as we see in the first box plot. They have similar lower bound, upper bound, Q1, Q2, and Q3. In the second figure, What we can extract is that people with having BMI index have higher median height for both males and females.

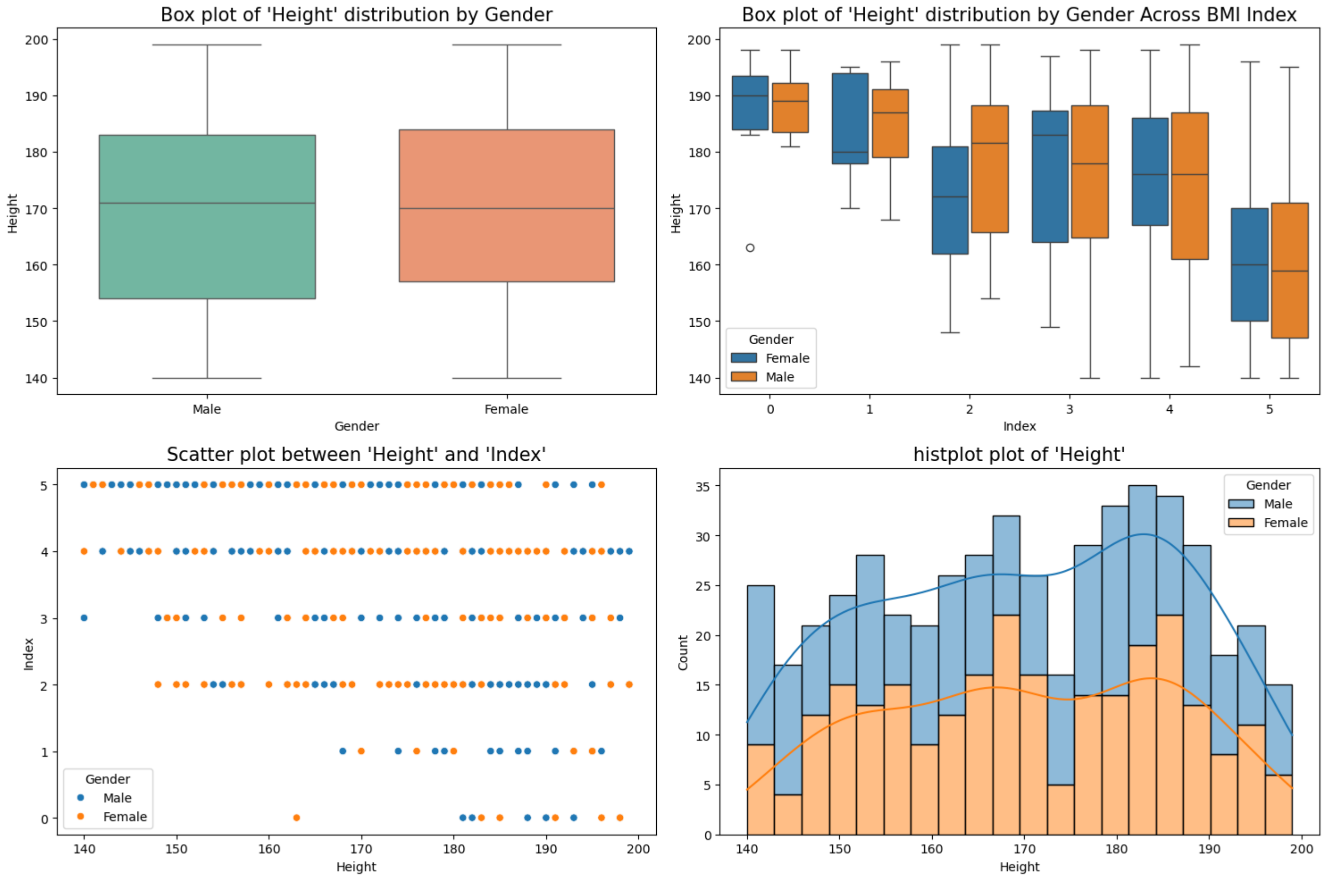

Figure 3: Visualization for the Height feature based on gender

4.2 Analysis of Weight feature

Analysis of "Weight" is performed. The distribution of weight based on gender seems identical as we see in the first box plot. They have similar lower bound, upper bound, Q1, Q2, and Q3. In the second figure, What we can extract is that people with low BMI index have low median weight for both males and females.

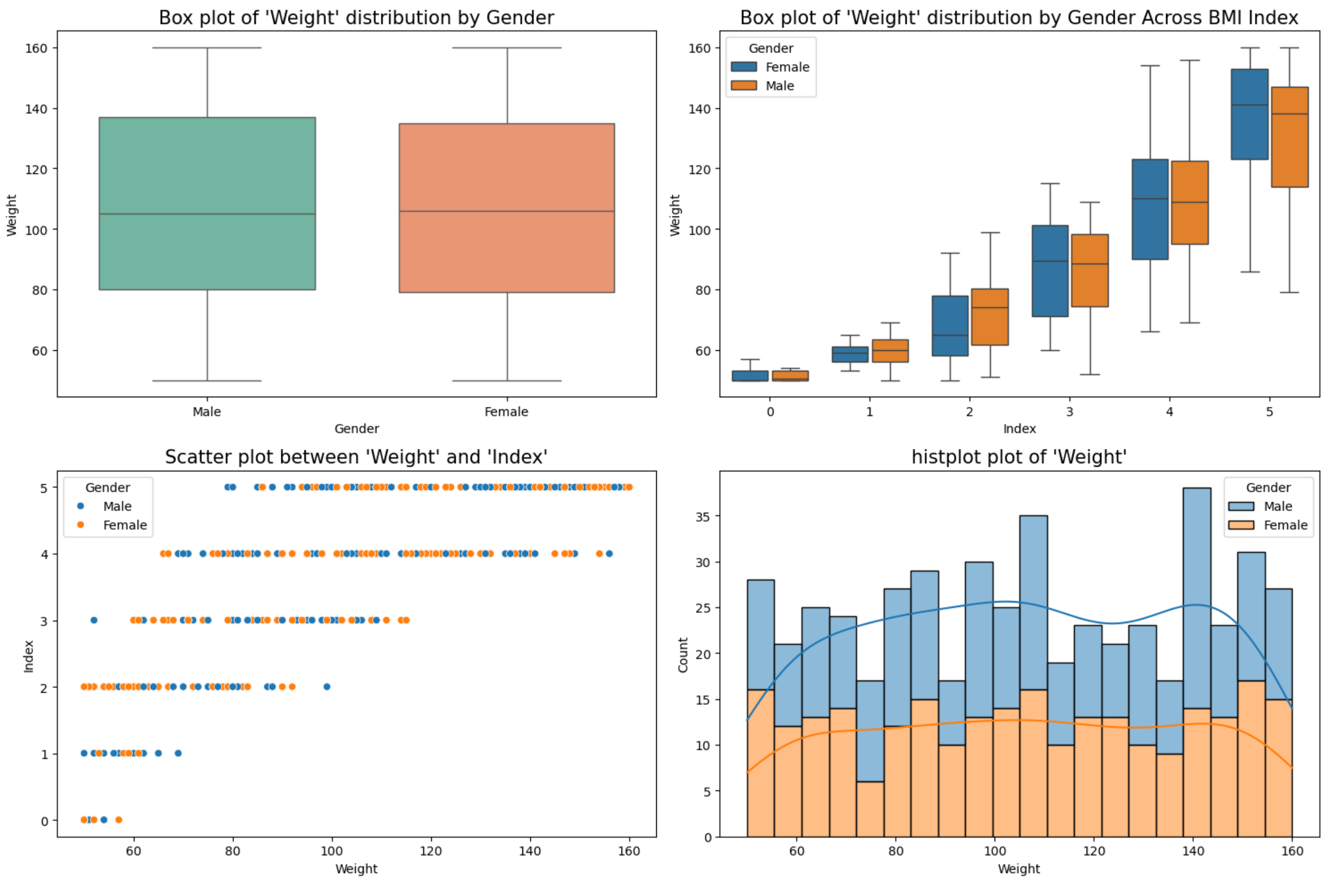

Figure 4: Visualization for the weight feature based on gender

4.3 Correlation Analysis

We calculated the correlation between "Height", "Weight", and "Index" features, We found that there is a positive but too weak association between height and weight features. This is good for the machine learning model as well because we will not face the problem of multi-collinearity. If we look at the correlation between height feature and target variable index, it has a value of -0.422 which means there is a moderate negative correlation between them. Upon looking correlation coefficient between weight and BMI Index we can see that the value is 0.804 which means there is a strong positive correlation between "Weight" and "BMI Index".

	Height	Weight	Index
Height	1.000000	0.000446	-0.422223
Weight	0.000446	1.000000	0.804569
Index	-0.422223	0.804569	1.000000

Table 3: Correlation table between features

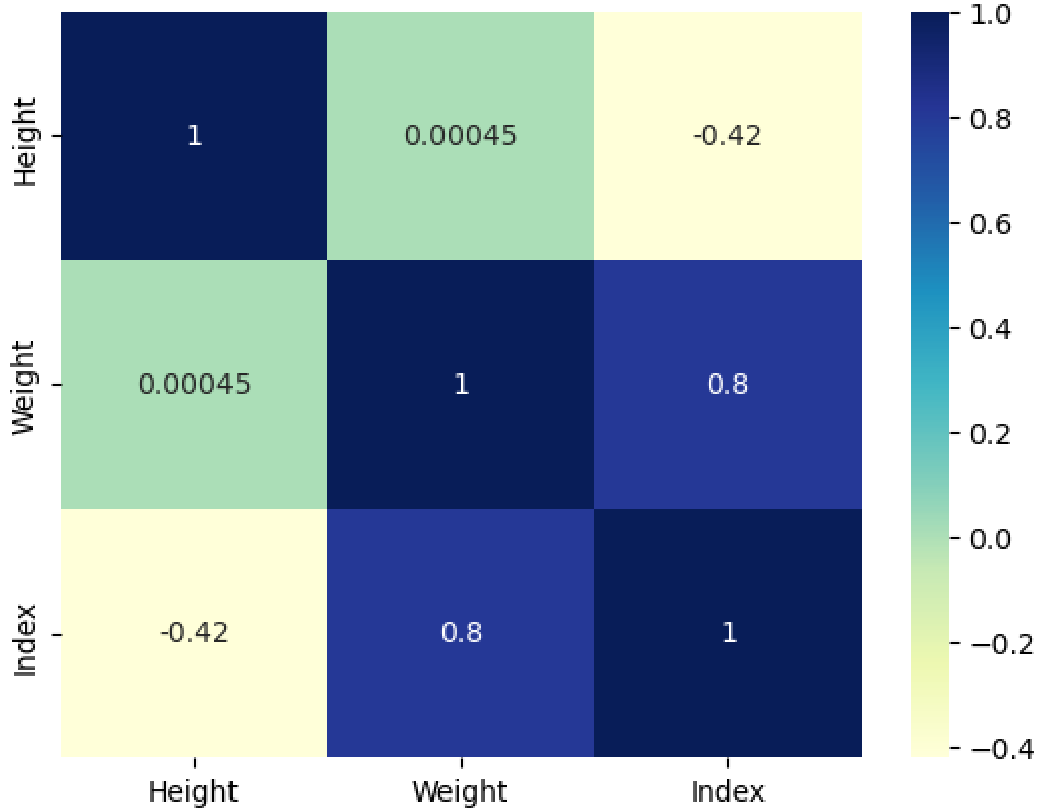

Figure 5: Correlation table between features

4.4 BMI Index Analysis

While analyzing the BMI Index target variables, we found that there is an imbalance in data between the categories.

BMI Index	count
0	13
1	22
2	69
3	68
4	130
5	198

Table 4: BMI Index Frequency count

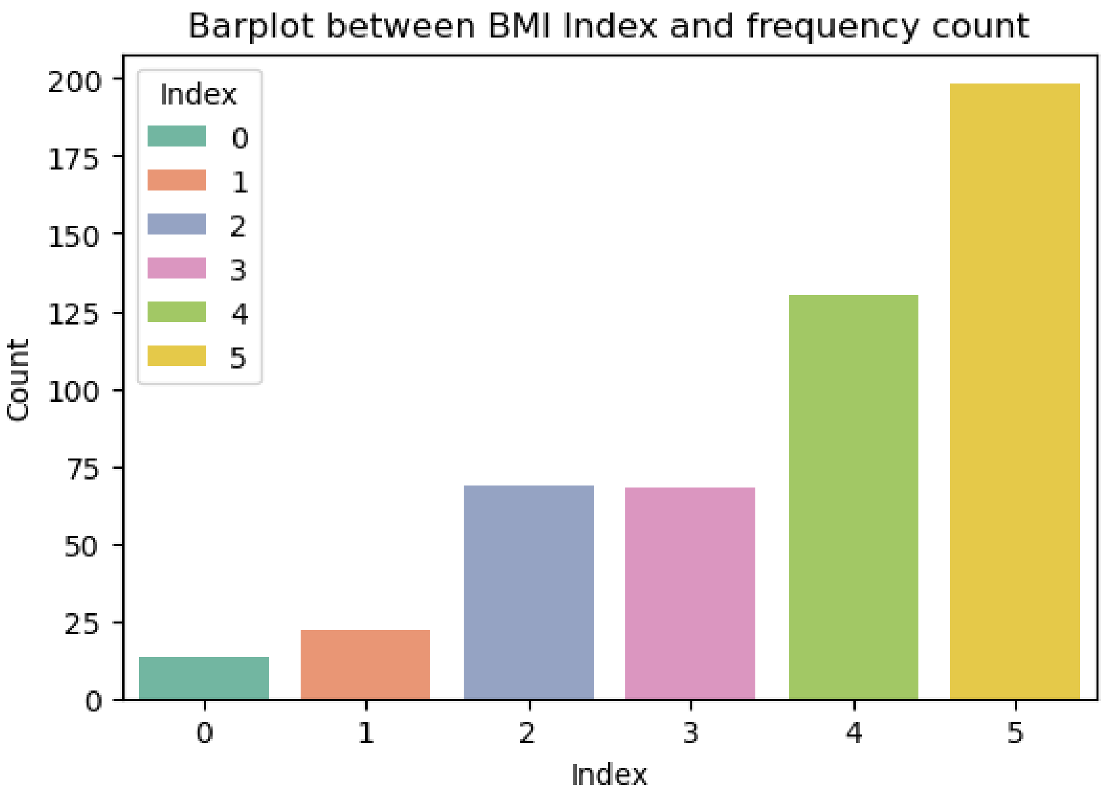

Figure 6: Bar plot for BMI index frequencies

5 Data Preprocessing

5.1 Resampling Techniques

While looking at the target variable records, I found that there is an imbalance in the records. There were few records only for some categorical classes in comparison to others. There are multiple strategies to address this issue and the strategy is oversampling and undersampling.

- Oversampling the Minority Class
- Undersampling the Majority Class

Before doing any resampling, we have the "Gender" categorical field as a feature. So, I proceeded with one hot encoding and dropped the Gender_Female field to avoid the dummy variable trap. I split the records into train test split to address the following issues:

- Preventing Data Leakage

- Maintaining the Integrity of the Test Set

5.1.1 Oversampling the Minority Class

Oversampling the minority class is a technique to address the issue of imbalance in the categorical class in datasets, particularly in classification tasks. When one class has fewer instances than another, then there can be the possibility of biased model performance. This is because the model may only focus on predicting the majority classes. Oversampling helps to balance the class distribution by increasing the number of instances in the minority class. Some oversampling techniques are explained below:

- **Random Oversampling**
  Random Oversampling is a technique to address an imbalance in categorical classes in datasets by increasing the number of instances in the minority class. Initially, It determines the minority category/class (the one with fewer instances) and majority category. It then randomly selects instances from the minority class and then duplicates them and finally balances the data based on the categories.

Gender_Male	count
0	488
1	478

Table 5: Gender and Frequency after random sampling

Index	count
0	161
1	161
2	161
3	161
4	161
5	161

Table 6: BMI Index and Frequency after random sampling

- **SMOTE (Synthetic Minority Over-sampling Technique)**
  SMOTE is a more advanced oversampling technique that addresses the imbalance in categorical classes by generating synthetic data points for the minority classes rather than duplicating existing ones. It generates more diverse samples for the minority classes which eventually helps to improve the performance of the model. It also reduces the risk of overfitting as it does not duplicate the existing records but rather generates new synthetic data points using the K-Nearest Neighbor (KNN) algorithm and by interpolating between the selected minority instance and a randomly chosen neighbor.

Gender_Male	count
0	566
1	400

Table 7: Gender and Frequency after SMOTE sampling

Index	count
0	161
1	161
2	161
3	161
4	161
5	161

Table 8: BMI Index and Frequency after SMOTE sampling

5.1.2 Undersampling the Majority Class

Undersampling is a technique used to address class imbalance by reducing the number of instances in the majority class to match the minority class, creating a more balanced dataset.

- Random Undersampling
  Random Undersampling is a technique to address class imbalance by randomly removing instances from the majority class until the dataset becomes balanced with respect to the minority class.

Gender_Male	count
1	36
0	24

Table 9: Gender and Frequency after Random undersampling

Index	count
0	10
1	10
2	10
3	10
4	10
5	10

Table 10: BMI Index and Frequency after Random undersampling

Note: For our classification models, we choose SMOTE oversampling.

5.2 Standardization and Normalization

Feature scaling is the data pre-processing technique to normalize or standardize the data of independent variables/features. This technique ensures that all data from independent variables are on a similar scale. There are two main techniques for feature scaling. They are as follows:

- **Mean Max Normalization**
  Mean-max Normalization is a scaling technique. It brings all the features from the dataset within a similar range. It involves calculating the mean and subtracting the mean from every feature. This feature is important for the machine learning algorithms which are sensitive to the feature scales, e.g.: gradient descent. Normalization ensures that all features contribute equally to the machine learning model's learning process and eventually lead to faster convergence and improve the model's performance. Following is the formula for mean-max normalization.

$$x' = \frac{x - \min(X)}{\max(X) - \min(X)} \tag{2}$$

- **Z-score standardization**
  Z-score standardization is a feature scaling technique used to transform data into a standard normal distribution with a mean of 0 and a standard deviation of 1. This will calculate the z-score using the formula mentioned below for each data point.

$$z = \frac{x - \mu}{\sigma} \tag{3}$$

6 Result and Observation

6.1 classification models on BMI Categories

6.1.1 Logistic Regression

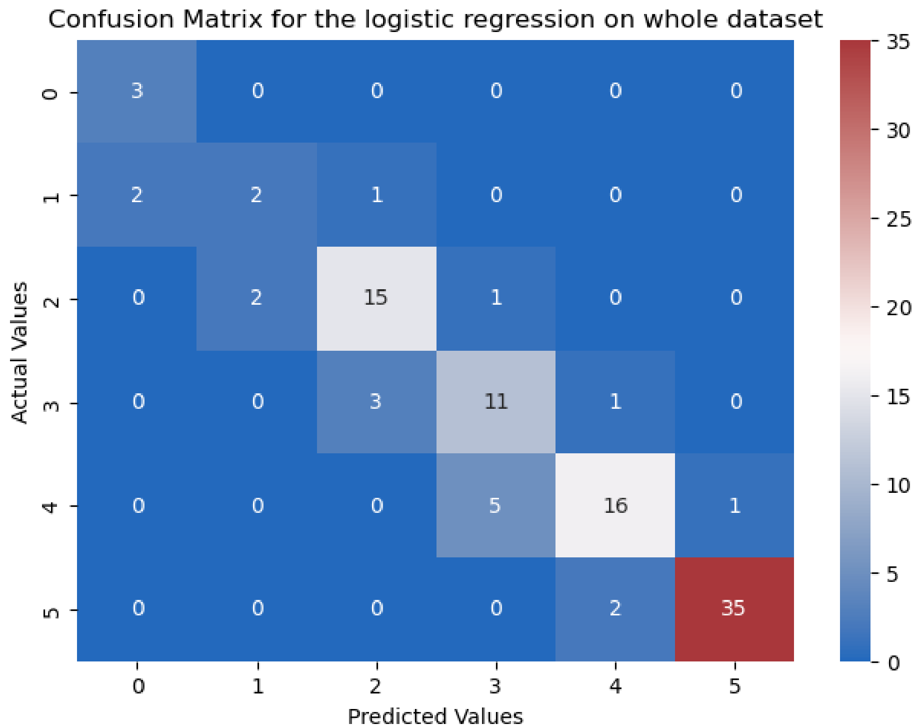

Figure 7: Confusion Matrix for logistic regression model on the whole dataset

Class	Precision	Recall	F1-Score	Support
0	0.60	1.00	0.75	3
1	0.50	0.40	0.44	5
2	0.79	0.83	0.81	18
3	0.65	0.73	0.69	15
4	0.84	0.73	0.78	22
5	0.97	0.95	0.96	37
Accuracy		0.82 (100)		
Macro Avg	0.73	0.77	0.74	100
Weighted Avg	0.83	0.82	0.82	100

Table 11: Classification Report for logistic regression model

The above confusion matrix and classification report evaluate the performance of logistic regression model on whole dataset. The classification report indicates that model has as overall accuracy of 82%. Model has macro average scores for the model are (Precision: 0.73, Recall: 0.77, F1-Score: 0.74) treat all classes equally, while the weighted averages (Precision: 0.83, Recall: 0.82, F1-Score: 0.82) reflect the class distribution, showing better alignment with the dataset's imbalance.

6.1.2 K-Nearest Neighbor

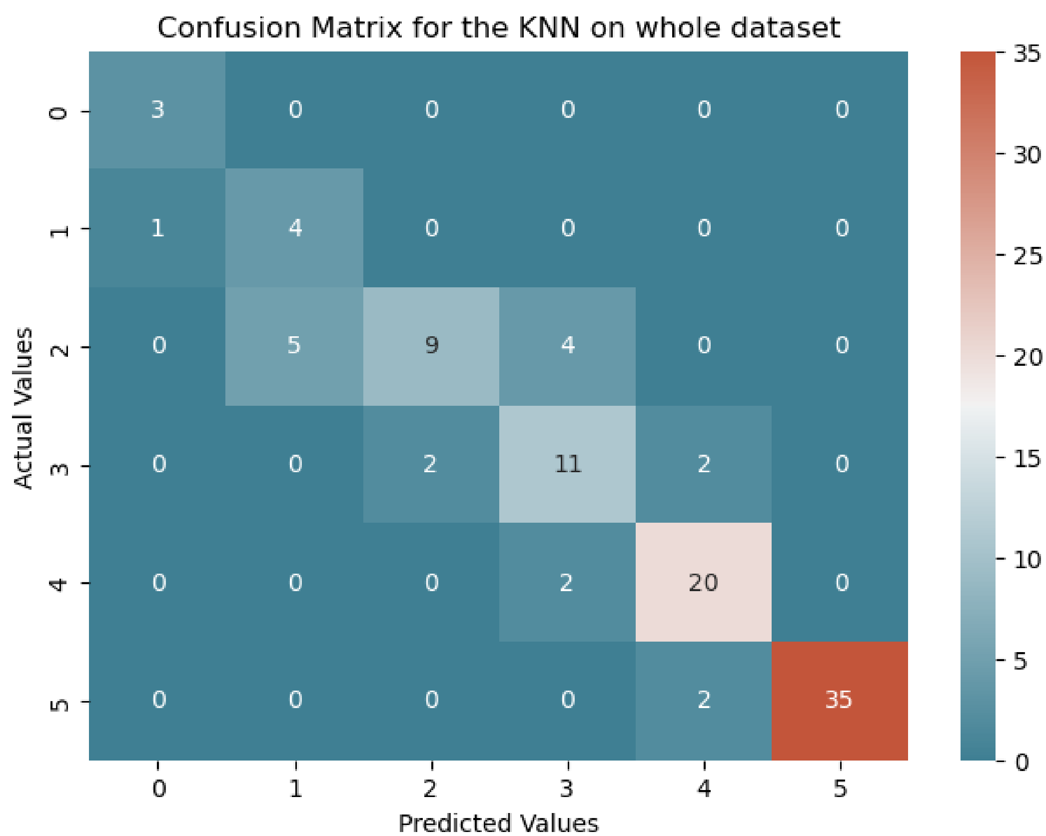

Figure 8: Confusion Matrix for K-Nearest Neighbor model on the whole dataset

Class	Precision	Recall	F1-Score	Support
0	0.75	1.00	0.86	3
1	0.44	0.80	0.57	5
2	0.82	0.50	0.62	18
3	0.65	0.73	0.69	15
4	0.83	0.91	0.87	22
5	1.00	0.95	0.97	37
Accuracy		0.82 (100)		
Macro Avg	0.75	0.81	0.76	100
Weighted Avg	0.84	0.82	0.82	100

Table 12: Classification Report for k-nearest neighbor model

The above confusion matrix and classification report evaluate the performance of K-nearest neighbor model on whole dataset. The classification report indicates that model has as overall accuracy of 82%. Model has following macro average scores (Precision: 0.75, Recall: 0.81, F1-Score: 0.76) and following the weighted averages scores (Precision: 0.84, Recall: 0.82, F1-Score: 0.82).

6.1.3 Support Vector Machine

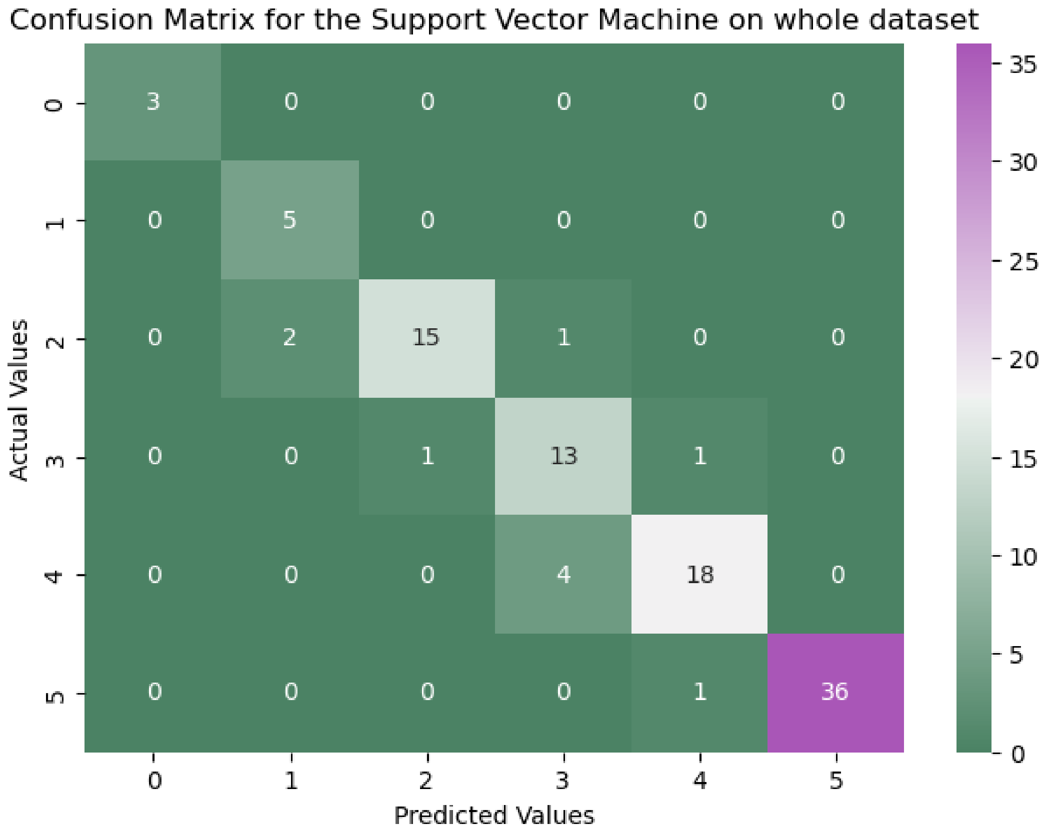

Figure 9: Confusion Matrix for Support Vector Machine model on the whole dataset

Class	Precision	Recall	F1-Score	Support
0	1.00	1.00	1.00	3
1	0.71	1.00	0.83	5
2	0.94	0.83	0.88	18
3	0.72	0.87	0.79	15
4	0.90	0.82	0.86	22
5	1.00	0.97	0.99	37
Accuracy		0.90 (100)		
Macro Avg	0.88	0.92	0.89	100
Weighted Avg	0.91	0.90	0.90	100

Table 13: Classification Report for support vector machine model

The above confusion matrix and classification report evaluate the performance of support vector machine model on whole dataset. The classification report indicates that model has as overall accuracy of 90%. Model has following macro average scores (Precision: 0.88, Recall: 0.92, F1-Score: 0.89) and following the weighted averages scores (Precision: 0.91, Recall: 0.90, F1-Score: 0.90).

6.1.4 Model Comparision on entire dataset

We tried to compare the performance between logistic regression, K-Nearest neighbor, and Support Vector Machine while using whole dataset. We used accuracy weighted Average precision, Recall and F1-Score for comparison.

Model	Accuracy	Precision	Recall	F1-Score
logistic regression	0.82	0.83	0.82	0.82
K-Nearest Neighbor	0.82	0.84	0.82	0.82
Support Vector machine	0.90	0.91	0.90	0.90

Table 14: Classification Models Comparision using weighted averages based on whole dataset

While comparing 3 models, If we look at all four performance matrices accuracy, precisions, recall and F1-Score of the models, Support vector machine has highest scores in all parameters. Support vector machine is performing best among 3 models on whole dataset.

6.2 Classification Models on BMI Categories specific to Male

BMI data are filtered for the males only and we got around 254 records. Male data was further split into train and test records. Upon looking into frequencies, the following findings were made.

Index	count
0	6
1	15
2	28
3	32
4	59
5	105

Table 15: BMI Indexes and frequency for male records

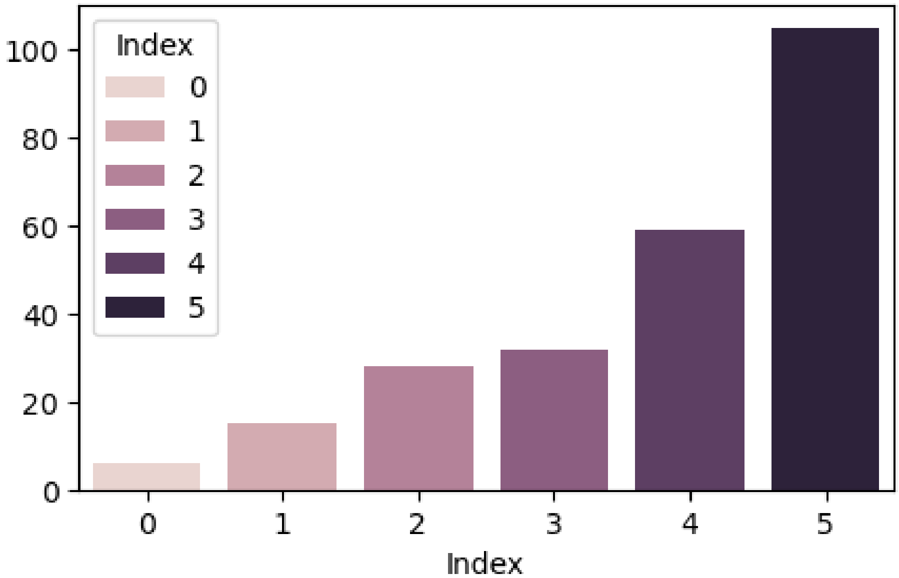

Figure 10: Barplot between BMI Index and frequency count for male records only

SMOTE sampling is performed on the dataset to address the imbalance in the dataset. After performing SMOTE sampling, data distribution becomes equal among each BMI categorical class.

Index	count
0	84
1	84
2	84
3	84
4	84
5	84

Table 16: BMI Indexes and frequency for male records after SMOTE sampling

6.2.1 Logistic Regression

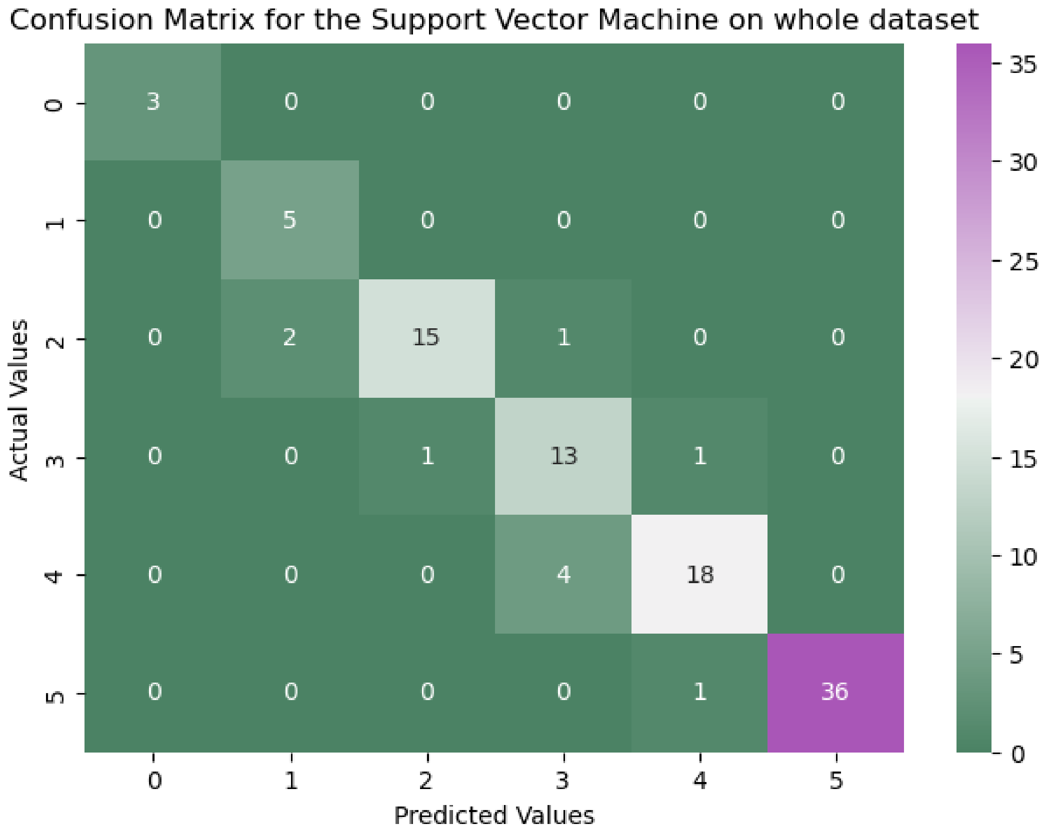

Figure 11: Confusion Matrix for male-specific logistic regression model

Class	Precision	Recall	F1-Score	Support
0	0.00	0.00	0.00	1
1	0.00	0.00	0.00	3
2	0.86	1.00	0.92	6
3	1.00	0.83	0.91	6
4	0.91	0.83	0.87	12
5	0.91	0.95	0.93	21
Accuracy		0.84 (49)		
Macro Avg	0.61	0.60	0.61	49
Weighted Avg	0.84	0.84	0.84	49

Table 17: Classification Report for male-specific logistic regression model

The above confusion matrix and classification report evaluate the performance of logistic regression model on male specific dataset. The classification report indicates that model has as overall accuracy of 84%. Model has following macro average scores (Precision: 0.61, Recall: 0.60, F1-Score: 0.61) and following the weighted averages scores (Precision: 0.84, Recall: 0.84, F1-Score: 0.84).

6.2.2 K-Nearest Neighbor

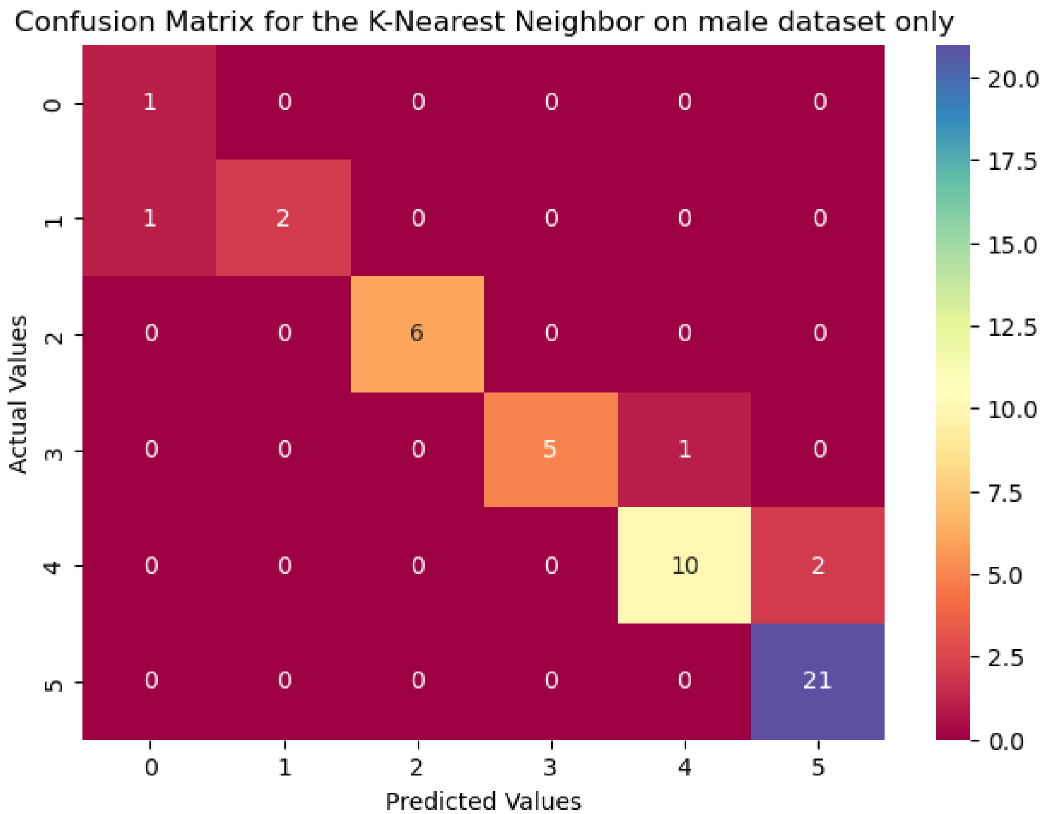

Figure 12: Confusion Matrix for male-specific k-nearest neighbor model

Class	Precision	Recall	F1-Score	Support
0	1.00	1.00	1.00	1
1	1.00	1.00	1.00	3
2	1.00	0.83	0.91	6
3	0.75	1.00	0.86	6
4	0.82	0.75	0.78	12
5	0.90	0.90	0.90	21
Accuracy		0.88 (49)		
Macro Avg	0.91	0.91	0.91	49
Weighted Avg	0.88	0.88	0.88	49

Table 18: Classification Report for male-specific k-nearest neighbor model

The above confusion matrix and classification report evaluate the performance of KNN model on male specific dataset. The classification report indicates that model has as overall accuracy of 88%. Model has following macro average scores (Precision: 0.91, Recall: 0.91, F1-Score: 0.91) and following the weighted averages scores (Precision: 0.88, Recall: 0.88, F1-Score: 0.88).

6.2.3 Support Vector Machine

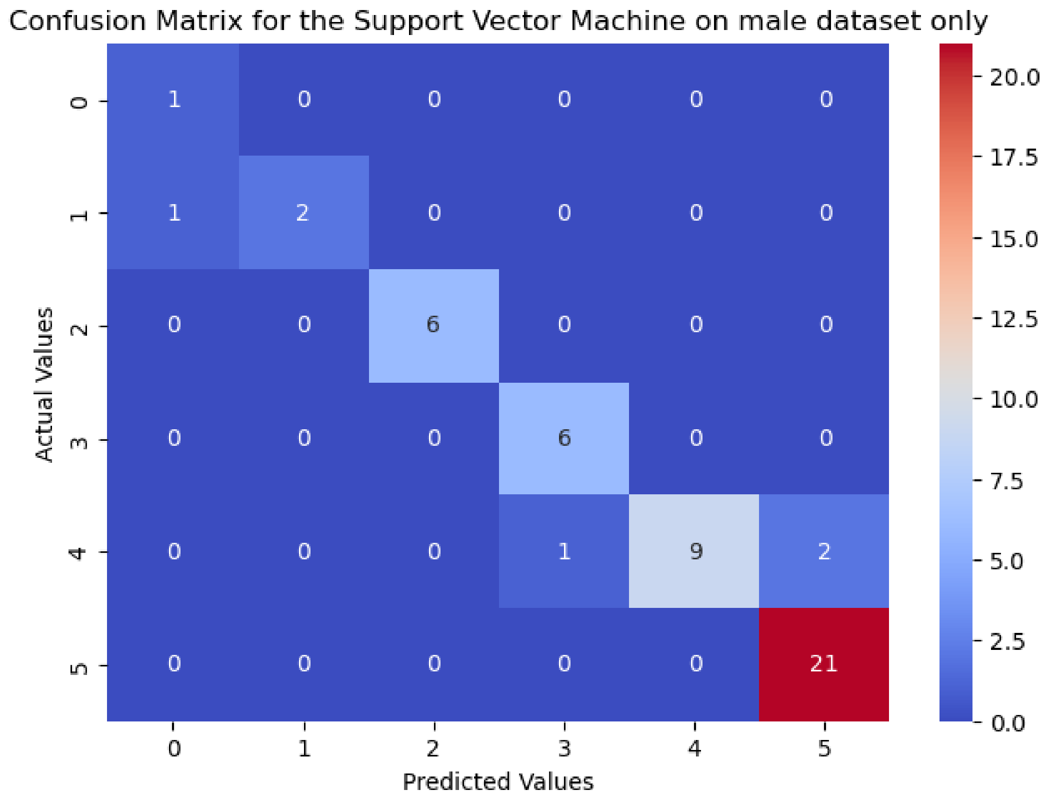

Figure 13: Confusion Matrix for male-specific support vector machine model

Class	Precision	Recall	F1-Score	Support
0	0.00	0.00	0.00	1
1	0.75	1.00	0.86	3
2	1.00	1.00	1.00	6
3	1.00	1.00	1.00	6
4	0.91	0.83	0.87	12
5	0.91	0.95	0.93	21
Accuracy		0.92 (49)		
Macro Avg	0.76	0.80	0.78	49
Weighted Avg	0.90	0.92	0.91	49

Table 19: Classification Report for male-specific support vector machine model

The above confusion matrix and classification report evaluate the performance of support vector machine model on male specific dataset. The classification report indicates that model has as overall accuracy of 92%. Model has following macro average scores (Precision: 0.76, Recall: 0.90, F1-Score: 0.78) and following the weighted averages scores (Precision: 0.90, Recall: 0.92, F1-Score: 0.91).

6.2.4 Model Comparision on male-speific dataset

Model	Accuracy	Precision	Recall	F1-Score
logistic regression	0.84	0.84	0.84	0.84
K-Nearest Neighbor	0.88	0.88	0.88	0.88
Support Vector machine	0.92	0.90	0.92	0.91

Table 20: Classification Models Comparision using weighted averages based on male-specific dataset

While comparing 3 models, If we look at all four performance matrices accuracy, precisions, recall and F1-Score of the models, Support vector machine has highest scores in all parameters. Support vector machine is performing best among 3 models on male specific dataset.

6.3 Classification Models on BMI Categories specific to Female

BMI data are filtered for the females only and we got around 255 records. Female data was further split into train and test records. Upon looking into frequencies, the following findings were made.

Index	Count
0	7
1	7
2	41
3	36
4	71
5	93

Table 21: BMI Indexes and frequency for female records

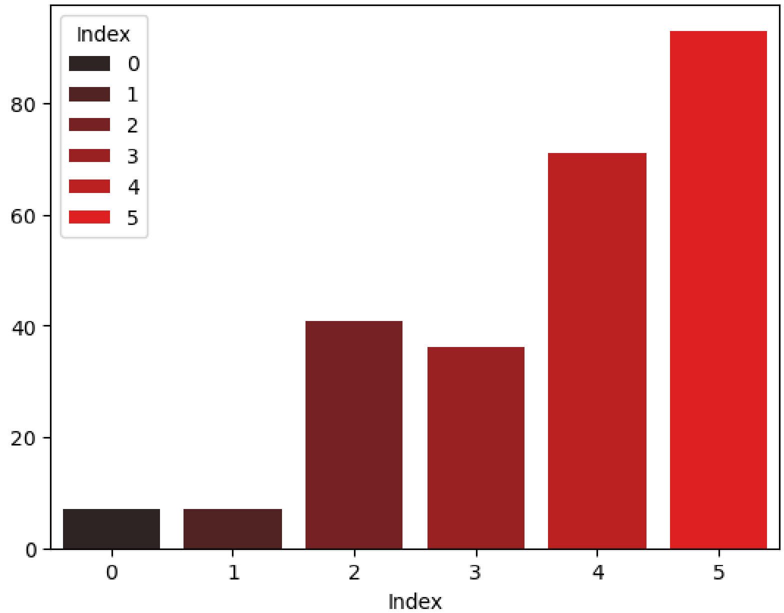

Figure 14: Barplot between BMI Index and frequency count for female records only

SMOTE sampling is performed on the dataset to address the imbalance in the dataset. After performing SMOTE sampling, data distribution becomes equal among each BMI categorical class.

Index	count
0	74
1	74
2	74
3	74
4	74
5	74

Table 22: BMI Indexes and frequency for female records after SMOTE sampling

6.3.1 Logistic Regression

	Predicted 0	Predicted 1	Predicted 2	Predicted 3	Predicted 4	Predicted 5
Actual 0	1	0	0	0	0	0
Actual 1	0	1	1	0	0	0
Actual 2	0	1	5	1	1	0
Actual 3	0	0	3	4	0	0
Actual 4	0	0	0	1	13	0
Actual 5	0	0	0	0	0	19

Table 23: Confusion Matrix for female-specific logistic regression model

Class	Precision	Recall	F1-Score	Support
0	1.00	1.00	1.00	1
1	0.50	0.50	0.50	2
2	0.56	0.62	0.59	8
3	0.67	0.57	0.62	7
4	0.93	0.93	0.93	14
5	1.00	1.00	1.00	19
Accuracy		0.84 (51)		
Macro Avg	0.78	0.77	0.77	51
Weighted Avg	0.85	0.84	0.84	51

Table 24: Classification Report for female-specific logistic regression model

The above confusion matrix and classification report evaluate the performance of logistic regression model on female specific dataset. The classification report indicates that model has as overall accuracy of 84%. Model has following macro average scores (Precision: 0.78, Recall: 0.77, F1-Score: 0.77) and following the weighted averages scores (Precision: 0.85, Recall: 0.84, F1-Score: 0.84).

6.3.2 K-Nearest Neighbor

	Predicted 0	Predicted 1	Predicted 2	Predicted 3	Predicted 4	Predicted 5
Actual 0	1	0	0	0	0	0
Actual 1	0	1	1	0	0	0
Actual 2	0	1	5	1	1	0
Actual 3	0	0	0	6	1	0
Actual 4	0	0	0	1	12	1
Actual 5	0	0	0	0	1	18

Table 25: Confusion Matrix for female-specific k-nearest neighbor model

Class	Precision	Recall	F1-Score	Support
0	1.00	1.00	1.00	1
1	0.50	0.50	0.50	2
2	0.83	0.62	0.71	8
3	0.75	0.86	0.80	7
4	0.80	0.86	0.83	14
5	0.95	0.95	0.95	19
Accuracy		0.84 (51)		
Macro Avg	0.81	0.80	0.80	51
Weighted Avg	0.85	0.84	0.84	51

Table 26: Classification Report for female-specific k-nearest neighbor model

The above confusion matrix and classification report evaluate the performance of KNN model on female specific dataset. The classification report indicates that model has as overall accuracy of 84%. Model has following macro average scores (Precision: 0.81, Recall: 0.80, F1-Score: 0.80) and following the weighted averages scores (Precision: 0.85, Recall: 0.84, F1-Score: 0.84).

6.3.3 Support Vector Machine

	Predicted 0	Predicted 1	Predicted 2	Predicted 3	Predicted 4	Predicted 5
Actual 0	1	0	0	0	0	0
Actual 1	0	1	1	0	0	0
Actual 2	0	2	4	1	1	0
Actual 3	0	0	0	7	0	0
Actual 4	0	0	0	0	14	0
Actual 5	0	0	0	0	1	18

Table 27: Confusion Matrix for female-specific support vector machine model

Class	Precision	Recall	F1-Score	Support
0	1.00	1.00	1.00	1
1	0.33	0.50	0.40	2
2	0.80	0.50	0.62	8
3	0.88	1.00	0.93	7
4	0.88	1.00	0.93	14
5	1.00	0.95	0.97	19
Accuracy		0.88 (51)		
Macro Avg	0.81	0.82	0.81	51
Weighted Avg	0.89	0.88	0.88	51

Table 28: Classification Report for female-specific support vector machine model

The above confusion matrix and classification report evaluate the performance of support vector machine model on female specific dataset. The classification report indicates that model has as overall accuracy of 88%. Model has following macro average scores (Precision: 0.81, Recall: 0.82, F1-Score: 0.81) and following the weighted averages scores (Precision: 0.89, Recall: 0.88, F1-Score: 0.88).

6.3.4 Model comparision on female-specific dataset

Model	Accuracy	Precision	Recall	F1-Score
logistic regression	0.84	0.85	0.84	0.84
K-Nearest Neighbor	0.84	0.85	0.84	0.84
Support Vector machine	0.88	0.89	0.88	0.88

Table 29: Classification Models Comparision using weighted averages based on female-specific dataset

While comparing 3 models, If we look at all four performance matrices accuracy, precisions, recall and F1-Score of the models, Support vector machine has highest scores in all parameters. Support vector machine is performing best among 3 models on female specific dataset.

6.4 Hyper-parameter Tuning's impact on Model Performances

Hyperparameters are the configuration variables that are set before training any machine learning models and Hyper parameter tuning is the process of identifying best values for the hyper-parameters such that model performance is improving. Hyper-parameter tuning plays a critical role in optimizing and improving model performances. It guides how the model is structured and trained. In case of logistic regression, regularization parameter and regularization type are the hyper-parameters. In case of KNN, number of neighbors is the hyper-parameter. In case of support vector machine, regularization parameter, kernel type, and gamma are the hyper-parameters. Changing the hyperparameter changes the model performances and hyperparameter tuning refers to optimizing the values of these parameters so that model performance improves.

6.5 Gender-specific Models and General Models Comparision

Model	Accuracy	Precision	Recall	F1-Score
logistic regression (General)	0.82	0.83	0.82	0.82
K-Nearest Neighbor (General)	0.82	0.84	0.82	0.82
Support Vector machine (General)	0.90	0.91	0.90	0.90
logistic regression (Male)	0.84	0.84	0.84	0.84
K-Nearest Neighbor (Male)	0.88	0.88	0.88	0.88
Support Vector machine (Male)	0.92	0.90	0.92	0.91
logistic regression (female)	0.84	0.85	0.84	0.84
K-Nearest Neighbor (female)	0.84	0.85	0.84	0.84
Support Vector machine (female)	0.88	0.89	0.88	0.88

Table 30: Gender-specific and General classification models Comparisions using weighted averages based on female-specific dataset

While comparing the logistic regression only performance across entire dataset, male-specific dataset and female-specific dataset, logistic regression (male) and logistic regression (female) have highest accuracy of 84%. While looking at the precision, recall and f1-score performance matrices, gender-specific logistic regressions are outperforming general logistic regression. While looking into the table 30, Gender-specific KNN models and support vector machine models are also outperforming the general classification model.

Gender-specific classification models are outperforming general classification models. This happen because of variation in data distribution between genders. There is chance that gender-specific models captures complex interaction between height and weights that are relevant to each gender. General classification models might introduce bias due to the averaging of effects across genders.

7 Conclusion

In conclusion, performances varies for each classification models. This is because each classification models are using different algorithms and cost functions. In general, Support vector machine is performing better than logistic regression and KNN models.