International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10
DOI: https://doi.org/10.17762/ijritcc.v11i10.8510
***Article Received:** 30 July 2023 **Revised:** 22 September 2023 **Accepted:** 09 October 2023*

Comparison of Different Machine Learning and Self-Learning Methods for Predicting Obesity on Generalized and Gender-Segregated Data

[1]Khushi Joshi, [2]Neha Aher, [3]Soha Arora, [4]Varnika Mulay, [5]Prof. Renuka Suryawanshi, [6]Dr. Nitin Pise, [7]Dr. Vitthal Gutte
[1,2,3,4,5,6,7]Department of Computer Science Engineering, MIT WPU, Pune, India
[1]gunnujoshi98@gmail.com, [2]nehaaher16@gmail.com, [3]sohaarora12@gmail.com, [4]varni.mulay@gmail.com,
[5]renuka.suryawanshi@mitwpu.edu.in, [6]nitin.pise@mitwpu.edu.in, [7]vitthalgutte2014@gmail.com

Abstract— Obesity is a global health concern with long-term implications. Our research applies numerous Machine Learning models consisting of Random Forest model, XGBT(Extreme Gradient Boosting) model, Decision Tree model, k-Nearest Neighbors technique, Support Vector Machine model, Linear Regression model, Naïve Bayes classifier and a neural network named Multilayer Perceptron on an obesity dataset so that we can predict obesity and reduce it. The models are evaluated on recall, accuracy, F1-score, and precision. The findings reveal the performance of the algorithms on generalised and gender-segregated data providing insights concerning feature selection and early obesity identification. This research aims to demonstrate the comparative study of obesity prediction for gender-neutral and gender-specific datasets.

Keywords- Obesity, Decision Trees, Machine Learning, Support Vector Machine, Naïve Baye's classifier, Random Forest.

Abbreviations
RFT : Random Forest
XGBT : XGBoost
DT : Decision Tree
KNN : k-Nearest Neighbors
SVM : Support Vector Machine
LR : Logistic Regression
NB : Naïve Bayes
MLP: Multilayer Perceptron

I. INTRODUCTION

Unhealthy eating habits accompanied by a lethargic lifestyle lead to an unhealthful weight gain. Obesity is a wellness concern that causes immoderate weight gain in a human being. It is an intricate condition driven by various factors such as physical fitness, daily food and water intake, exercise regimen, age, height, etc.

Being obese can be immensely harmful to any individual's health as it could result in a reduced life span and various life-threatening diseases like cardiovascular diseases, diabetes, respiratory issues, kidney disease, skin infections, etc. Whether a person is obese or not is decided by a numerical value i.e the person's BMI- Body Mass Index. BMI formula [5, 6]: BMI = weight (kg) / height (m2) [5, 6].

The World Health Organization (WHO) has set some standards for categorizing obese people. Anybody having a BMI of 30 or more is considered obese. The WHO has also specified a few levels of obesity shown in Table 1 [5, 6].

Table -1: Health standards specified by WHO.

Sr. No.	Category	BMI
1.	Person is underweight	< 18.5
2.	Person is healthy	18.5 – 24.9
3.	Person is overweight	25.0 – 29.9
4.	Person has obesity level 1	30.0 – 34.9
5.	Person has obesity level 2	35.0 – 39.9
6.	Person has obesity level 3	> 40

Our collection of data focuses on the data from sovereign nations of North and South America namely Colombia, Mexico and Peru. Hence, considering today's situation, we discovered that according to the Global Obesity Observatory (GOO) [1-3]: Peru carries a probability of 7.5/10, denoting

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10
DOI: https://doi.org/10.17762/ijritcc.v11i10.8510
***Article Received:** 30 July 2023 **Revised:** 22 September 2023 **Accepted:** 09 October 2023*

that every 7th person out of 10 is obese [1]. Mexico carries a probability of 8/10, denoting that every 8th person out of 10 is obese [2]. Colombia carries a probability of 7.5/10, denoting that every 7th person out of 10 is obese [3]. Looking at these figures, we realized the importance of designing a model that can predict obesity to reduce these figures and improve the overall fitness of the people. According to our research, the obesity rates by country in 2023 [4]: Colombia = 23.9% Peru = 21.1% Mexico = 30.6%. This study also focuses on presenting the comparative study of obesity prediction based on the individual's gender. It shows how the features affecting the person's obesity change as per gender.

II. LITERATURE SURVEY

Uses Deep learning framework to predict future obesity trends for ages between 3–20 years from the medical history of children (Electronic Health Records) using LSTM network architecture. Its capability is limited to forecast predictions for a maximum of 3 years ahead, contains only non-grouped features and the number of features is not reduced or standardized [8].

Application of ML techniques for a prediction model for the detection of individuals who are obese or may be overweight. Attributes include physical condition and eating habits. ML classification algorithms tested: DT, SVM, k-NN, NB, XGBT model, RFT ensemble technique, as well as extreme gradient boosting model. RFT outperformed other algorithms achieving the best result with 77.69% accuracy [9].

This study proposed to predict obesity using different ML models utilizing LR, SVM, RFT ensemble technique, (ANN)-artificial neural network, KNN and k-means clustering to categorize the individuals into BMI groups during various stages of pediatric care including regular checkups, multiple early and random checkups with an accuracy of 89%, 77%, and 89% [10].

The objective is to evaluate the effectiveness of ML methods, specifically Classification and Regression Trees, LR, and NB in addressing disproportionate data by employing the SMOTE (Synthetic Minority Oversampling Technique) model. The goal is to predict obesity status using risk factors from the dataset. Dataset extracted from the RISKESDAS survey; LR achieves the highest accuracy metric at 72%. The CART approach attained the most sensitivity at 82% and the maximum F1-score stood at 72% [11].

Estimated levels of obesity in people belonging to Colombia, Peru and Mexico. Consisting of 17 distinct features and 2111 separate records the dataset is moderately diverse, 77% of which were synthetically generated using the SMOTE filter and Weka. The algorithm used is DT [12].

The research performed the mutual information classification technique, χ-square method (χ= Chi), F-Classify algorithm, and for obesity estimation with an average accuracy of 93.06%, 89.04%, 90.32%, and 86.52% for chosen features. Bayesian optimization techniques are implemented and comparison is done using a trained neural network to identify risk factors [13].

Applied nine ML algorithms including RFT algorithm, KNN, MLP neural network, SVM, DT classifier, Adaptive Boosting, LR and NB and Gradient Boosting Classifier for prediction of obesity. The highest accuracy of 97.09% was achieved for the LR model whereas lowest accuracy was recorded for "Gradient Boosting" Algorithm at 64.08% [14].

Used five machine learning algorithms consisting of the RFTs, XGBT Classifier, DT model, KNN- k-nearest Neighbor, and SVM on Publicly available datasets. XGBT achieved the highest accuracy of 99.05% while "KNN" stood at a promising low of 95.74% [15].

III. METHODOLOGY

A. Dataset Description

The dataset required for this study was acquired from the UCI Machine Learning Repository and includes data from sovereign nations of North and South America namely - Colombia, Mexico and Peru. It comprises 17 different features alongside 2111 distinct instances [7].

The features included in the dataset are mentioned in **Table -2** [13]:

Table -2: Dataset features description

Sr. No.	Features	Meaning
1.	FCVC	Tracking vegetable intake
2.	CAEC	Tracking of food consumption between meals
3.	NCP	Number of main meals
4.	FAVC	Tracking high-caloric food intake
5.	SCC	Calorie consumption tracking
6.	FAF	Tracking physical movements
7.	TU E	Time consumption of electronic devices
8.	MTRANS	Mode of Transport
9.	CH20	Daily water intake
10.	CALC	Alcohol consumption

A. Method

In this study, the first step was to apply basic preprocessing algorithms to check for clean data. After confirming that the data was clean, the next step was to calculate the BMI of every instance and add it as a new feature in the dataset. As Machine Learning (ML) algorithms consider only numeric data for

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10
DOI: https://doi.org/10.17762/ijritcc.v11i10.8510
***Article Received:** 30 July 2023 **Revised:** 22 September 2023 **Accepted:** 09 October 2023*

analysis, a label encoder from the sklearn preprocessing library is applied to convert all categorical data to numeric data.

For the purpose of identifying the correlation present in the data, Pearson's correlation technique is implemented. The correlation obtained above is plotted as a heatmap using the Seaborn and Matplotlib library provided by Python.

After recognizing the highly correlated attributes, the dataset is divided into separate dataframes, X and y, where X: consists of the feature data and y: depicts the target attribute. The dataset is then splitted into training and testing chunks such that 80% of the data is selected for the training phase and 20% is reserved for the testing stage; the same is implemented by applying the train_test_split function present in Python sklearn library [23].

ML provides a method called the SMOTE- 'Synthetic Minority Oversampling Technique' to reduce the problem of overfitting caused by random oversampling. Python provides an imblearn library that includes the SMOTE function in the over_sampling method for achieving the above purpose. Subsequently, the RFT Classifier is used to showcase the attribute importance score of all features of the dataset. The model is then fitted on the train data frame and the feature_importances_ function is applied to get the required output. Feature importance is a method in ML that displays the score of every feature that determines its importance in the dataset. There exist multiple ML techniques and models that are used to calculate these scores. The output is then plotted as a bar plot using the data visualization libraries provided by Python for better understanding. The features are displayed on the Y-axis, and the importance is displayed on the X-axis. The following step is to implement the ML algorithms on the separated training phase data followed by testing the built models on our reserved test data.

1. *LR Model:* Described as a statistical model implemented for prediction analytics and classification. It works on independent variables in a dataset. The output is probability-based, and the dependent variable is between 0 and 1.

   Logit(π) = 1/ (1+ exp(-π))
   *ln (pi/(1-π)) = Beta0 + Beta1*X_1 + … + Bk*Kk*

2. *DT:* It is a non-parametric, supervised ML algorithm applied for solving regression problems as well as classification problems. It has a tree representation containing root nodes, branches, leaf nodes, and internal nodes. It uses a divide-and-conquer strategy and conducts greedy searches within a tree. Information gain and Entropy criteria is calculated for every node in order to select the best attributes.

   Information Gain(S,a)= Entropy(S)-Σ(|Sy |/|S|)Entropy(Sy)

3. *RFT*: RFT is a supervised learning technique. It can be defined as a method implementing Ensemble learning. It enhances predictive accuracy by combining multiple classifiers to address problems effectively.

4. *SVM*: It is a Supervised ML algorithm which performs exceptionally to resolve classifying problems. It is used for linear or non-linear classification as well as outlier detection. SVM finds an optimal hyperplane that separates different classes in target features.

   {For linear SVM classifier, we use:
   $y = \{1: wTx + b \geq 0$
   $y = \{0: wTx + b < 0$

5. *k-NN*: KNN has a non-parametric approach. It is a supervised machine learning classifier. It relies on proximity-based analysis for grouping individual data points and hence, performs classification. For distance calculation, it uses Euclidean, Manhattan, Minkowski, and Hamming distance formulas.

   Euclidean Distance = $d(x,y)=\sqrt{i=1n(y_i - x_i)^2}$

   Manhattan Distance = $d(x,y) = (i=1m|x_i-y_i|)$

   Minkowski Distance = $(i=1n|\, x_i - y_i|)^{1/p}$

6. *NB:* It serves as a classifier which works on probability grounded in the principles of Baye's Theorem concerning Bayesian statistics. It calculates the conditional probability for classification. It does not learn the feature importance of classifying different classes.

   Posterior probability =[(Conditional probability)(Prior Probability)]/evidence

7. *XGBT*: XGBT is an application of Gradient Boost-DT. For the model, trees are generated in a chronolological way. Weights are assigned to all independent variables which serve as input for the tree that predicts the result.

   $y_i = \sum_{k=1}^{K} f_k(x_i)$, $f_k \in F$

8. *Neural Network:* We have executed the MLP Neural Network characterized by a feedforward architecture with multiple layers. All the layers relate to each other and each layer communicates with adjacent

layers. Perceptron to Neural networks is the Neuron counterpart to a human brain. MLP consists of node values, activation functions, and node weights to compute network results

To determine which model predicts the accurate result, ML provides a few measures:

1. *Accuracy*

It represents the proportion of accurate predictions out of the total number of predictions generated.

$$\text{Accuracy} = \frac{\text{Number of Correct predictions}}{\text{Total no. of . prediction}}$$

2. *Precision*

It represents the proportion of correct Positives outputs to the total Positives that the model generated.

$$\text{Precision} = \frac{\text{True positive}}{\text{True positive + false positive}}$$

3. *Recall*

It signifies the ratio of positive outcomes to the total number of pertinent samples that were expected to be positive.

$$\text{Recall} = \frac{\text{True positive}}{\text{True positive + false negative}}$$

4. *F1 score*

It is the Harmonic mean of Precision and Recall, providing insights into the model's precision and resilience.

F1 = 2 *(1/(1/precision)+(1/recall))

5. *Confusion matrix*

Generates one matrix output that gives a comprehensive evaluation of the model based on the following results:

5.1 True Positive (TP): Forecasted “YES” while actual output was also “YES”

5.2 False Positive (FP): Forecasted “YES” but the actual output was “NO”

5.3 False Negative (FN): Forecasted “NO” but actual output was “YES”

5.4 True Negative (TN): Forecasted “NO” while actual output was also “NO”

The model which gives the most accurate score is identified based on the above parameters. Considering this study, one of the main objectives is to identify the hidden patterns in the data to develop an accurate model for predictive analysis. ML provides a technique called clustering that assists in this process. Clustering is an ML method that falls under unsupervised ML. It is a way in which clusters of unlabelled, similar data are formed to display and recognize the patterns. Some of the clustering techniques used in this study are:

1. *K-Means clustering*

K-means is a method that aims at forming clusters based on the similarity between the data. It makes sure that the clusters are distinct and that the data points belonging to a cluster are like each other and different from the data points belonging to other clusters.

2. *Elbow Method*

The elbow method uses the Within Cluster Sum of Squares method for finding out the ideal count of clusters.

$$\text{WCSS} = \sum_{P_i \text{ in Cluster1}} \text{distance}(P_i\, C_1)^2 + \sum_{P_i \text{ in Cluster2}} \text{distance}(P_i\, C_2)^2 + \sum_{P_i \text{ in CLuster3}} \text{distance}(P_i\, C_3)^2$$

Following the clustering method, we have also calculated the silhouette score of the model. A silhouette score shows how good the clustering technique we have used. The score value ranges from [-1, 1]. The values nearer to 1 are the best whereas the score nearer to -1 are the worst. Scores nearer to 0 indicate that the clusters will be overlapping.

Self-organizing maps, formerly referred to as Kohonen maps, are an instance of artificial neural network (ANN) that serves as the basis for algorithms that use unsupervised learning that minimize dimensions, acquire insights from data that is highly dimensional, pinpoint patterns granting a starting place for additional decision making, perceive a cluster's location which is basically the positioning of neurons, the edge clusters that demonstrate outliers, and cluster separation which indicates how the data points of some spectra are separated. We have the statistics for each of the neurons that comprise the SOM in our interpretation, providing the index of its location, mean, standard deviation, and the number of data points.

Dimensionality reduction is also accomplished using the Principal Component Analysis (PCA) method. This study applies PCA to the data, reducing it to two dimensions and then visualizing the result.

This entire procedure is then applied to the male and female datasets separately to attain a comparative analysis of the

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10
DOI: https://doi.org/10.17762/ijritcc.v11i10.8510
***Article Received:** 30 July 2023 **Revised:** 22 September 2023 **Accepted:** 09 October 2023*

same. The model is validated by providing some unseen data as input and getting the predicted result as output.

IV. SYSTEM ARCHITECTURE

B. Former System Design

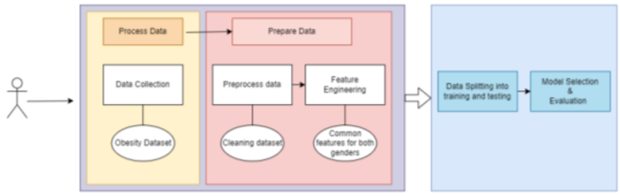

Fig -1: Former system design.

A. Suggested System Design

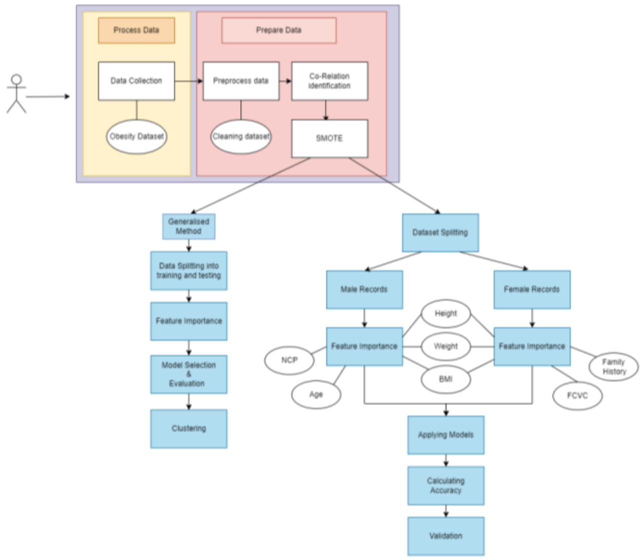

Fig -2: Suggested system design

Figure 2 represents the working of our model in a diagrammatic format.

V. RESULTS AND DISCUSSIONS

A. Generelized data (Male+Female)

1.*Correlation Analysis*

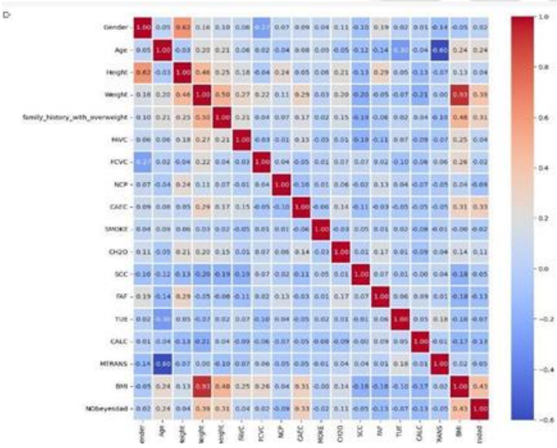

Fig -3: Heatmap showing correlation analysis of our dataset.

The above graph shown in Fig. 3 depicts the correlation between BMI and weight as the highest.

2. Feature Importance

Feature importance is a measure that organizes the features based on the impact they have in predicting the target value.

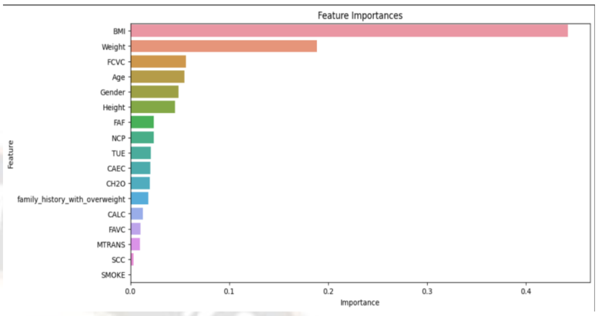

Fig -4: Feature importance using STD scaler.

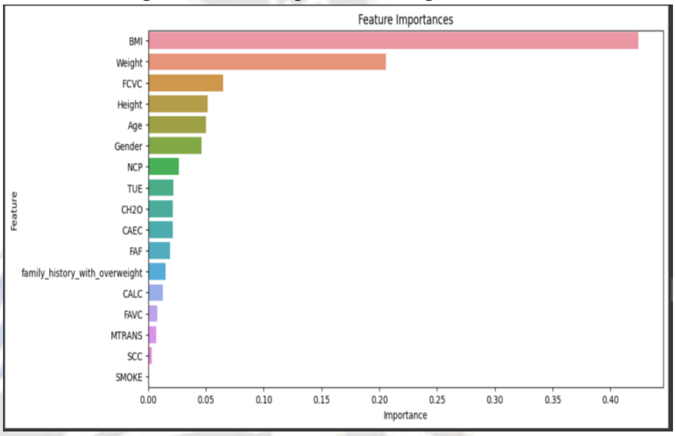

Fig -5: Feature importance using min max scaler

Feature importance shows that the importance of features like CALC, FAVC, MTRANS, SCC, and smoke are the least in both, the Minmax scaler and STD scaler, whereas features like BMI, weight, FCVC, height, age, and gender are the most important.

3. Models Accuracy

Comparison between model accuracy obtained after applying min max scaler and STD scaler.

Table -3: Model accuracies obtained after applying min max scaler.

Sr No.	Algorithms	Accuracy Score
1.	**DT**	**0.97**
2.	**LR**	**0.77**
3.	**SVM**	**0.87**
4.	**RFT**	**1.00**
5.	**NB**	**0.89**
6.	**k-NN**	**0.77**
7.	**XGBT**	**0.98**
8.	**Neural Network**	**0.96**

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10
DOI: https://doi.org/10.17762/ijritcc.v11i10.8510
***Article Received:** 30 July 2023 **Revised:** 22 September 2023 **Accepted:** 09 October 2023*

Table -4: Model accuracies obtained after applying standard scaler.

Sr No.	Algorithms	Accuracy Score
1.	**DT**	**0.97**
2.	**LR**	**0.90**
3.	**SVM**	**0.92**
4.	**RFT**	**0.99**
5.	**NB**	**0.89**
6.	**k-NN**	**0.83**
7.	**XGBT**	**0.98**
8.	**Neural Network**	**0.96**

The result of the above comparison shown in Table 3 and Table 4 is that the STD scaler gives more accurate results than MinMax scaler. This is because standardization is a good choice when the distribution of the data is approximately normal or when the algorithms being worked on assume normally distributed data, such as linear regression.

To enhance the accuracy scores of the proposed model, we removed the least important features and rescaled the data to get the following output (refer to Table 5):

Table -5: Model accuracy using an STD scaler after eliminating the least important features.

Sr No.	**Algorithms**	**Accuracy Score**
1.	**DT**	**0.96**
2.	**LR**	**0.89**
3.	**SVM**	**0.92**
4.	**RFT**	**0.99**
5.	**NB**	**0.92**
6.	**k-NN**	**0.87**
7.	**XGBT**	**0.98**
8.	**Neural Network**	**0.97**

4.Clustering - Elbow method

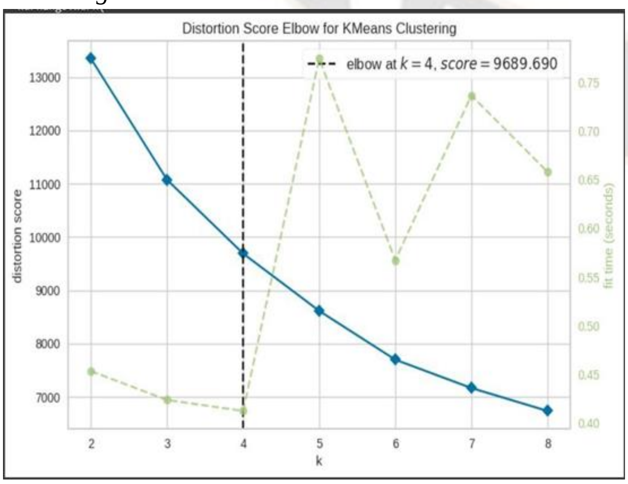

Fig -6: Elbow Method

Elbow method is a technique used in k-means clustering to identify the count of clusters i.e., the value of k. The point where the elbow bends is the k value. As shown in Fig. 6, the x-axis demonstrates the count of clusters (k value), and the y-axis demonstrates the Within Clusters Sum of Squares (WCSS) value for each value of k.

The output after implementing k-means clustering demonstrates that the count of clusters formed should be 4 (as k=4). The following clusters are formed after applying the k-means function on the dataset.

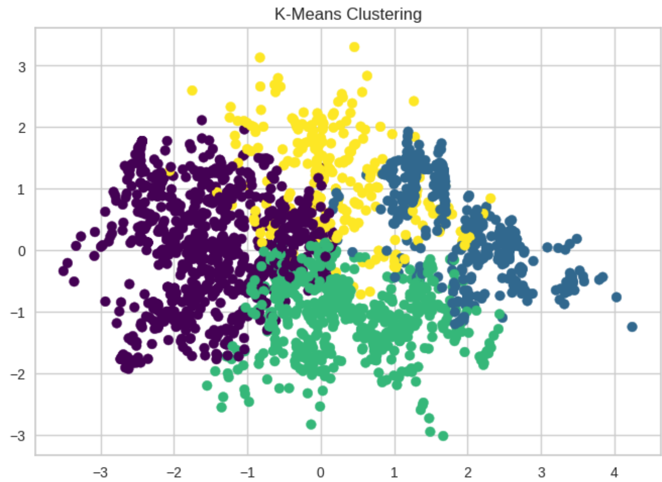

Fig -7: K-Means Clustering

Figure 7 clearly shows 4 different clusters. After further calculations and getting the cluster summary, the total number of records in each cluster came to be:
Cluster 0: 819 records
Cluster 1: 482 records
Cluster 2: 537 records
Cluster 3: 273 records
Subsequently, after mapping the records to the clusters we observed that cluster 0 was formed by combining records of labels 0 and 1 (and a few outliers). Cluster 1 was formed by combining records of labels 3 and 4. Cluster 2 was formed by combining records of labels 2, 3, and 6. Similarly, cluster 3 was formed by combining records of labels 2, and 6. This proves that labels 0 and 1, 3 and 4, 2 and 6 are like each other.
Labels Description is shown in Table 6 [7]:

Table -6: Labels Description

Sr No	**Category**	**]Label**
1.	Person is underweight	0
2.	Person is healthy	1
3.	A person has obesity type 1	2
4.	A person has obesity type 2	3
5.	A person has obesity type 3	4
6.	Person is overweight (level 1)	5
7.	Person is overweight (level 2)	6

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10
DOI: https://doi.org/10.17762/ijritcc.v11i10.8510
***Article Received:** 30 July 2023 **Revised:** 22 September 2023 **Accepted:** 09 October 2023*

The silhouette score for the clustering method is also calculated which comes to be around 0.30. Any score near 0 depicts overlapping clusters.

5.Self - Organizing Maps

Self-organizing maps, formerly referred to as Kohonen maps, are an instance of artificial neural network (ANN) that serves as the basis for algorithms that use unsupervised learning that minimise dimensions, acquire insights from the data that is highly dimensional, pinpoint patterns granting a starting place for additional decision making, perceive a cluster's location which is basically the positioning of neurons, the edge clusters that demonstrate outliers, and cluster separation which indicates how the data points of some spectra are separated. We have the statistics for each of the neurons that comprise the SOM in our interpretation, providing the index of its location, mean, standard deviation, and the number of data points.

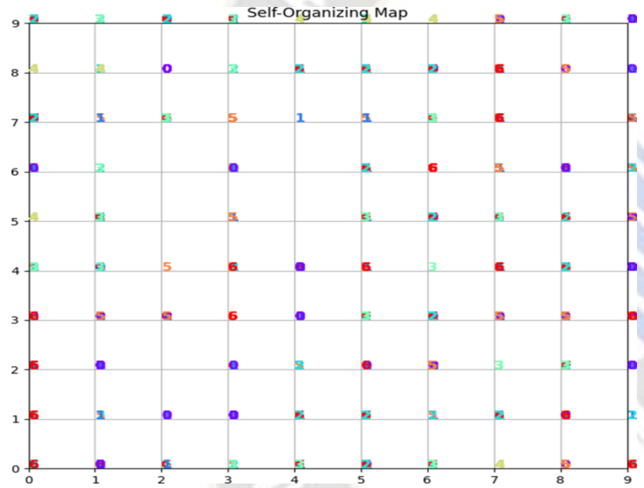

Fig -8: Self - Organizing Map

C. Female data v/s Male data comparative analysis

1.Feature importance (top 5 features)

Table -7: Feature importance applied to female data separately.

Sr No	**Feature**	**Importance Score**
1.	BMI	0.378
2.	Weight	0.277
3.	FCVC	0.049
4.	Height	0.046
5.	Family history	0.038

Table -8: Feature importance applied to male data separately.

Sr No	**Feature**	**Importance Score**
1.	BMI	0.419
2.	Weight	0.247
3.	Age	0.071
4.	Height	0.049
5.	NCP	0.036

The above comparison shown in Table 7 and Table 8 proves that the features affecting the prediction of the model changes according to the gender of the person. The 5 most important features of female data are BMI, weight, FCVC, height, and family_history_with_overweight. Whereas the 5 most important features of male data are BMI, weight, age, height, and NCP.

2.Model accuracy

Table -9: Comparison of model accuracy applied to female data separately.

Sr No	Algorithm	Accuracy Score
1.	DT	0.95
2.	LR	0.90
3.	SVM	0.91
4.	RFT	0.99
5.	NB	0.86
6.	k-NN	0.83
7.	XGBT	0.98
8.	Neural Network	0.95

Table -10: Comparison of model accuracy applied to male data separately.

Sr No	Algorithm	Accuracy Score
1.	DT	0.90
2.	LR	0.92
3.	SVM	0.93
4.	RFT	1.0
5.	NB	0.84
6.	k-NN	0.86
7.	XGBT	0.99
8.	Neural Network	0.95

Table 9 and Table 10 shows a comparative analysis of the accuracy scores obtained after applying the algorithms to the male and female data separately. The accuracy of the male data turned out to be slightly better than the female data as the number of instances of the male dataset is higher than that of the female.

Finally, we validated the model by passing some unseen data as input. The unseen data input was in the form of an array [0, 30, 165, 60, 3.0, 20.0] which gave the following prediction displayed in Table 11 below.

Table -11: Model accuracies were obtained for validation of the model.

Sr No.	**Algorithms**	**Predicted label**
1.	LR	4
2.	DT	4
3.	RFT	3
4.	SVM	5
5.	k-NN	3

6.	NB	2
7.	XGBT	4
8.	Neural Network	4

Observing the predicted labels for the provided input we can see that most of the algorithms predict Obesity Type III (refer to Table 6).

VI. CONCLUSION

We were successfully able to predict the type of obesity of a person using certain ML algorithms. This study concludes that obesity is a disease that is influenced by various factors and the importance of these features varies based on the gender of the person. The main finding of this research was that the common features that determine the obesity levels of both males and females are weight and height and its comparative analysis .

Further expansion of this research could include improving the silhouette score for clustering and predicting behavioral change based on their diet and lifestyle. This will create better clusters and help in identifying the patterns in the dataset and the similarity between our data points.

REFERENCES

1. Peru, World Obesity. Available online: https://data.worldobesity.org/country/peru-171/
2. Mexico, World Obesity. Available online: https://data.worldobesity.org/country/mexico-139/
3. Colombia, World Obesity. Available online: https://data.worldobesity.org/country/colombia-43/
4. Obesity Rates by Country, wise voter. Available online: https://wisevoter.com/country-rankings/obesity-rates-by-country/
5. Obesity, World Health Organization. Available online: https://www.who.int/health-topics/obesity#tab=tab_1
6. A healthy lifestyle - WHO recommendations (6 May 2010), World Health Organization Available online:
7. https://www.who.int/europe/news-room/fact-sheets/item/a-healthy-lifestyle---who-recommendations
8. Estimation of obesity levels based on eating habits and physical condition . (2019). UCI Machine Learning Repository.
9. https://doi.org/10.24432/C5H31Z.
10. Gupta M, Phan TT, Bunnell HT, Beheshti R. Obesity Prediction with EHR Data: A deep learning approach with
11. interpretable elements. ACM Trans Comput Healthc. 2022 Jul;3(3):32. doi: 10.1145/3506719. Epub 2022 Apr 7. PMID:
12. 35756858; PMCID: PMC9221869.
13. [9] Elias Rodrígueza , Elen Rodrígueza , Luiz Nascimentoa,b , Aneirson da Silvaa and Fernando Marinsa. 2021.
14. Machine learning techniques to predict overweight or obesity. IDDM
15. Mondal, P.K.; Foysal, K.H.; Norman, B.A.; Gittner, L.S. Predicting Childhood Obesity Based on Single and Multiple
16. Well-Child Visit Data Using Machine Learning Classifiers. Sensors 2023, 23, 759.
17. Thamrin SA, Arsyad DS, Kuswanto H, Lawi A and Nasir S (2021) Predicting Obesity in Adults Using Machine
18. Learning Techniques: An Analysis of Indonesian Basic Health Research 2018. Front. Nutr. 8:669155. doi:
19. 10.3389/fnut.2021.669155
20. Fabio Mendoza Palechor, Alexis de la Hoz Manotas, Dataset for estimation of obesity levels based on eating habits and physical condition in individuals from Colombia, Peru and Mexico, Data in Brief, Volume 25, 2019, 104344, ISSN
21. 2352-3409.
22. Yagin, F.H.; Gülü, M.; Gormez, Y.; Castañeda-Babarro, A.; Colak, C.; Greco, G.; Fischetti, F.; Cataldi, S. Estimation of Obesity Levels with a Trained Neural Network Approach optimized by the Bayesian Technique. Appl. Sci. 2023, 13, 3875.
23. Faria Ferdowsy, Kazi Samsul Alam Rahi, Md. Ismail Jabiullah, Md. Tarek Habib, A machine learning approach for
24. obesity risk prediction, Current Research in Behavioral Sciences, Volume 2, 2021, 100053, ISSN 2666-5182.
25. Musa Fati Anisat, Federick Duniya Basaky, E.O. Osaghae, Obesity Prediction Model Using Machine Learning
26. Techniques, Journal of Applied Artificial Intelligence, 2022, Volume 3, Issue 1 : 24 – 3, ISSN: 2709-5908, DOI :10.48185/jaai.v3i1.47