Obesity Levels Estimation Using Machine Learning

Newaz Muhammad Nahian Khan
ID: 1812395642
Department of Electrical and Computer Engineering
North South University
Dhaka, Bangladesh
newaz.nahian@northsouth.edu

Sabiha Akter Shorna
ID: 1922013642
Department of Electrical and Computer Engineering
North South University
Dhaka, Bangladesh
sabiha.shorna@northsouth.edu

***Abstract*—Since obesity is a disease that affects people of all ages and genders globally, researchers have worked very hard to pinpoint the early triggers. An intelligent method is developed in this study based on supervised and unsupervised data mining techniques such as Decision Trees, and Support Vector Machines to detect obesity levels and help people and health professionals in leading healthier lifestyles in order to combat this global epidemic. In this study, students between the ages of 18 and 25 who attended universities in Colombia, Mexico, and Peru served as the main source of data collecting.The study uses data on the major contributors to obesity with the intention of referencing high caloric intake, a reduction in energy expenditure as a result of inactivity, gastrointestinal diseases, genetics, socioeconomic variables, and/or anxiety and depression. 178 students from the selected dataset—81 men and 97 women—participated in the study. Results using algorithms like Decision Tree, Random Forest, Gaussian Naive Bayes, Logistic Regression, Linear Regression and Support Vector Machine (SVM), demonstrate a useful tool to conduct a comparison analysis of the techniques stated.**

***Keywords*— SVM, KNN, Decision Tree etc.**

I. Introduction

[1]The Centers for Disease Control and Prevention view obesity as an important health issue that may be overcome. Children who are overweight or obese are a target for intervention because they are more likely than children of normal weight to experience health issues. There is proof that there are inequalities in pediatric obesity; Bethell et al. examined variations in obesity rates by race/ethnicity, insurance, and income and discovered inequalities within and across states. In order to implement cost-effective interventions, it can be useful to pinpoint specific local geographic areas where children are most at risk for having high body mass index (BMI). Each of these indicators can vary dramatically across a city or county.

The rising prevalence of obesity has become a significant public health concern worldwide. Obesity, characterized by excessive body fat accumulation, poses numerous health risks and has been associated with various chronic conditions, including cardiovascular diseases, diabetes, and certain types of cancer. In recent years, the recognition of obesity as a complex multifactorial condition has prompted researchers to explore innovative approaches to its detection and management.

Obesity has become a significant global health concern, with detrimental effects on both individuals and societies. It is associated with numerous chronic conditions, such as cardiovascular diseases, diabetes, and certain types of cancer, leading to increased healthcare costs and reduced quality of life. Identifying and addressing obesity at an early stage is crucial for effective intervention and prevention strategies.

The primary goal of this research paper is to investigate the detection of obesity using various machine learning techniques and explore the potential of these models in accurately predicting obesity levels.[2]The problem of obesity detection is important because early identification allows for timely intervention and personalized treatment plans, ultimately improving health outcomes and reducing the burden on healthcare systems.

To approach this problem, we employed a comprehensive data preprocessing methodology. This involved outlier detection to identify extreme values in the age and weight columns, followed by one-hot encoding for categorical variables and ordinal encoding for variables such as food consumption, alcohol intake, and obesity levels. These preprocessing steps helped standardize the data and prepare it for further analysis.

Furthermore, we applied a range of machine learning algorithms to the preprocessed dataset. These algorithms included OLS regression, Linear Regression, Logistic Regression, Decision Trees, Random Forests, K-Nearest Neighbors (KNN), and Support Vector Machines (SVM). By utilizing a diverse set of models, we aimed to compare their performance and determine which algorithms are most effective in accurately predicting obesity levels.

The key research goals of this study are as follows:

1. To evaluate the effectiveness of various machine learning algorithms in detecting and predicting obesity levels based on the given dataset.
2. To compare the performance of different models in terms of accuracy, precision, recall, and F1-score, enabling us to identify the most suitable algorithm for obesity detection.
3. To provide insights and recommendations for healthcare professionals and policymakers regarding the implementation of accurate and reliable obesity detection methods, leading to more targeted interventions and improved public health outcomes.

By achieving these goals, we aim to contribute to the growing body of research on obesity detection and provide valuable insights for healthcare professionals, researchers, and policymakers. Effective and accurate obesity detection can guide interventions, preventive measures, and personalized treatment plans, ultimately leading to improved health outcomes and a reduction in the global burden of obesity-related diseases.

II. Literature Review

The authors of Ref. [3] present a beginning strategy for the investigation of childhood obesity prediction, gathering data from primary sources including parents, kids, and caretakers. Risk variables were discovered, including obesity, parental education level, children's lifestyle and habits, and environmental influences.In *[4]*, a computational model utilizing a fuzzy signature is described in order to comprehend and manage the complexities of data pertaining to childhood obesity and one potential solution that could deal with the danger linked with early obesity and childhood motor development. In [5], the author aims to develop a model for the identification, analysis, and estimation of obesity, in which each user is considered a ‘sensor’ of the online social network that can provide valuable health information. Based on the detailed measurement of the correlation of obesity and the proposed characteristics, the analytical model of obesity of the NSO can estimate the rate of obesity in certain urban areas, and the experimental results demonstrate a high rate of estimation success.The authors of Ref. [6] offer a logistic regression model to calculate the likelihood of a child between the ages of 2 and 17 having a mass body index over a limited geographic area.The authors of Ref. [7] present a beginning strategy for the investigation of childhood obesity prediction, gathering data from primary sources including parents, kids, and caregivers. Risk variables were discovered, including obesity, parental education level, children's lifestyle and habits, and environmental influences. The application of data mining for the prediction of childhood obesity is suggested in Reference [8]. The proposed survey's goal is to offer the knowledge required to understand obesity as a disease.

Based on the literature found, it is possible to see the efforts made by various authors to analyze the disease, even developing web tools like IMC calculation (2019), where one can determine a person's level of obesity. However, these tools are only capable of calculating body mass index, excluding other important factors like whether the person has a family history of obesity, how much time is spent engaging in exercise routines, and other factors. Therefore, a smart device that can accurately identify the level of obesity is required. The authors of [8] developed the risk mining approach (PRMT), which predicts a model to analyze the risk factor of obesity class using various data mining classifiers, employing WEKA to determine the accuracy and error measurement. The 10-fold cross-validation study's top classifier was produced by this procedure using Naive Bayes. In Ref. [9], the authors suggest kernel-based mid-level features for the prediction of population health indices from social media data. They take the distributions of textual features found in tweets and extract the kernel-based features from them, encoding the connections between the various textual elements in a kernel function. According to their eating habits and physical condition, the writers of "Trees" [10] provided data for the calculation of obesity levels in people from the nations of Mexico, Peru, and Colombia. They labeled the records with the class variable NObesity (Obesity Level), which allows classification of the data using the values of Insufficient Weight, Normal Weight, Overweight Level I, Overweight Level II, Obesity Type I, Obesity Type II, and Obesity Type III. The data contains 17 attributes and 2111 records. The Weka tool and the SMOTE filter were used to create 77% of the data synthetically, while a web portal was used to collect 23% of the data directly from users.

III. Methodology

A. Dataset

An observational strategy was used in the design of the current research. The eating habits and physical activity levels of 498 participants aged 14 to 61 from Barranquilla, Colombia, Lima, Peru, and Mexico City were included in this study's data for the assessment of obesity levels. Frequent consumption of high-calorie foods (FAVC), frequent consumption of vegetables (FCVC), the number of main meals (NCP), the consumption of food between meals (CAEC), daily water consumption (CH20), and alcohol consumption (CALC) were the characteristics connected to eating habits. Calorie consumption tracking (SCC), frequency of physical activity (FAF), time spent on technological devices (TUE), and mode of transportation (MTRANS) were the characteristics associated to the physical state. We also collected data on gender, age, height, and weight. After all calculations to calculate each person's body mass index (BMI) were finished, the WHO data were used to classify the levels of obesity as underweight when it is less than 18.5, normal when 18.5 to 24.9 and overweight for 25.0 to 29.9.It is considered to be obesity I for 30.0 to 34.9, obesity II for 35.0 to 39.9 and obesity III higher than 40.

B. Data-Preprocessing

Data preprocessing is an essential step in any data analysis or machine learning task. It involves transforming raw data into a format that is suitable for further analysis and modeling. In the context of your research on obesity detection, the following preprocessing steps were applied to the dataset.

- **Outlier Detection**: Outliers are data points that significantly deviate from the rest of the data. They can have a detrimental effect on the accuracy and reliability of statistical analysis or machine learning models. In this research, the z-score method was employed to detect outliers in the "age" and "weight" columns. The z-score measures how many standard deviations a data point is away from the mean. Any data point with a z-score beyond a specified threshold (cut_std=3) was considered an outlier and either removed or treated separately.

  In this research, the IQR (Interquartile Range) method was employed to detect outliers in the "age" and "weight" columns. The IQR is the range between the first quartile (25th percentile) and the third quartile (75th percentile) of a dataset. Based on the IQR analysis performed on the dataset, the following lower and upper limits were identified:

Lower Limit Age: 11.0

Upper Limit Age: 35.0

Any data point falling below the lower age limit or above the upper age limit is considered an outlier and can be treated separately or removed from the dataset.

Lower Limit Weight: 3.25

Upper Limit Weight: 169.25

Similarly, any data point falling below the lower weight limit or above the upper weight limit is considered an outlier and can be handled accordingly.

By identifying and addressing outliers in the "age" and "weight" columns using the IQR method, the dataset is prepared for subsequent analysis and modeling, ensuring that extreme values that may affect the results are appropriately managed.

- **One-Hot Encoding:** One-hot encoding is a technique used to convert categorical variables into a binary representation that can be easily understood by machine learning algorithms. In your research, the following categorical columns were one-hot encoded: "Gender," "Family History," "FAVC" (Frequent consumption of high-caloric food), "SMOKE(Do you smoke?)," "SSC" (Do you monitor the calories you eat daily?), and "MTRANS" (Transportation). Each category in these columns was transformed into a separate binary column, and a value of 1 was assigned if the category was present for a particular instance and 0 otherwise.
- **Ordinal Encoding:** Ordinal encoding is a method of encoding categorical variables where the categories are assigned integer values according to their order or priority. In your research, the columns "CAEC" (Consumption of food between meals), "CALC" (Consumption of alcohol), and "NObeyesdad" (Obesity level) were subjected to ordinal encoding. Each unique category in these columns was assigned a specific integer value based on its rank or priority.

  For the "CAEC" column, the categories were encoded as follows: "no" was assigned the value 0, "Sometimes" was assigned 1, "Frequently" was assigned 2, and "Always" was assigned 3.

  Similarly, in the "CALC" column, "no" was assigned the value 0, "Sometimes" was assigned 1, "Frequently" was assigned 2, and "Always" was assigned 3.

  In the "NObeyesdad" column, the categories representing different obesity levels were encoded as follows: "Insufficient_Weight" was assigned the value 0, "Normal_Weight" was assigned 1, "Overweight_Level_I" was assigned 2, "Overweight_Level_II" was assigned 3, "Obesity_Type_I" was assigned 4, "Obesity_Type_II" was assigned 5, and "Obesity_Type_III" was assigned 6.

  The ordinal encoding helps to represent the categorical variables with numerical values that reflect the order or priority of the categories. This enables subsequent analysis and modeling techniques to work effectively with these variables.

These preprocessing steps help to standardize the dataset, handle categorical variables, and address outliers. By preparing the data in this manner, it becomes more amenable to analysis and modeling for obesity detection.

- Decision Tree Algorithm:The Decision Tree algorithm is a supervised machine learning technique used for classification and regression tasks. It creates a flowchart-like tree structure where each internal node represents a feature or attribute, and each leaf node represents a class label or predicted value. The algorithm recursively partitions the data based on selected features to create homogeneous subsets and makes decisions based on the feature values.

- Random Forest Algorithm:The Random Forest algorithm is an ensemble learning method that combines multiple decision trees to make predictions. It creates a forest of decision trees and each tree is trained on a random subset of the data with replacement (bootstrap samples). During prediction, each tree in the forest independently makes a prediction, and the final prediction is determined by majority voting or averaging the predictions from individual trees.

- Gaussian Naive Bayes Algorithm:The Gaussian Naive Bayes algorithm is a probabilistic classification algorithm based on Bayes' theorem with the assumption that the features follow a Gaussian (normal) distribution. It calculates the probability of each class label given the feature values and predicts the class label with the highest probability.
- OLS Regression:OLS (Ordinary Least Squares) Regression is a linear regression method that aims to minimize the sum of the squared differences between the observed and predicted values. It estimates the coefficients of a linear equation to model the relationship between the dependent variable and one or more independent variables.
- KNN:The KNN algorithm is a non-parametric and lazy learning classification method. It classifies instances based on their proximity to other instances in the feature space. When predicting the class label of a new instance, KNN considers the K nearest neighbors (based on a distance metric) and assigns the majority class label among them.
- Logistic Regression:Logistic Regression is a statistical classification algorithm used for binary or multi-class classification problems. It models the relationship between the dependent variable and independent variables using the logistic function. Logistic Regression estimates the probability of an instance belonging to a particular class and predicts the class label based on a threshold.
- Linear Regression:Linear Regression is a statistical regression algorithm used to model the relationship between a dependent variable and one or more independent variables. It assumes a linear relationship between the variables and estimates the coefficients of a linear equation to predict the dependent variable based on the independent variables.

C.Experimental Analysis

A.Classification Algorithms

i.Decision Tree

A decision tree is one of the most powerful tools of supervised learning algorithms used for both classification and regression tasks. It builds a flowchart-like tree structure where each internal node denotes a test on an attribute, each branch represents an outcome of the test, and each leaf node holds a class label.Evaluating the performance of a decision tree classifier using cross-validation 93.2%.

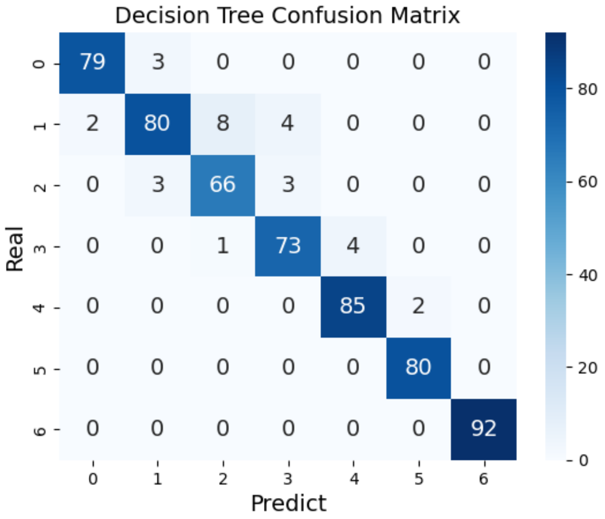


Fig1: DT Confusion Matrix

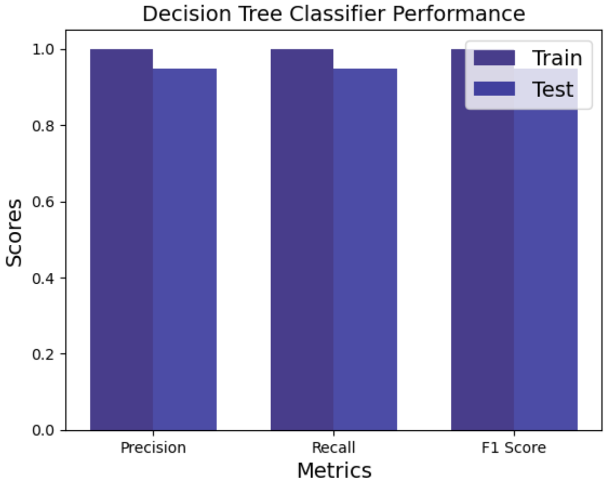
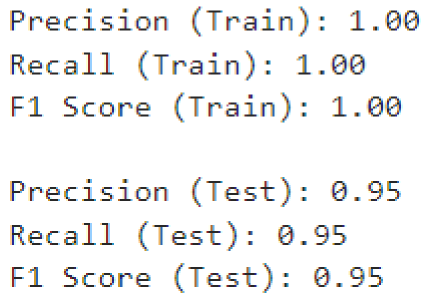


Fig2:DT Performance



```
Precision (Train): 1.00
Recall (Train): 1.00
F1 Score (Train): 1.00

Precision (Test): 0.95
Recall (Test): 0.95
F1 Score (Test): 0.95
```

Fig3:DT Performance Metrics Scores

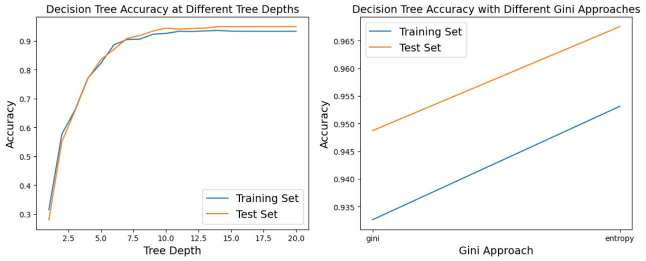


Fig4:DT Graphs

ii.Random Forest

The supervised learning method includes the well-known machine learning algorithm Forest. It can be applied to ML issues involving both classification and regression. It is built on the idea of ensemble learning, which is a method of integrating various classifiers to address difficult issues and enhance model performance.

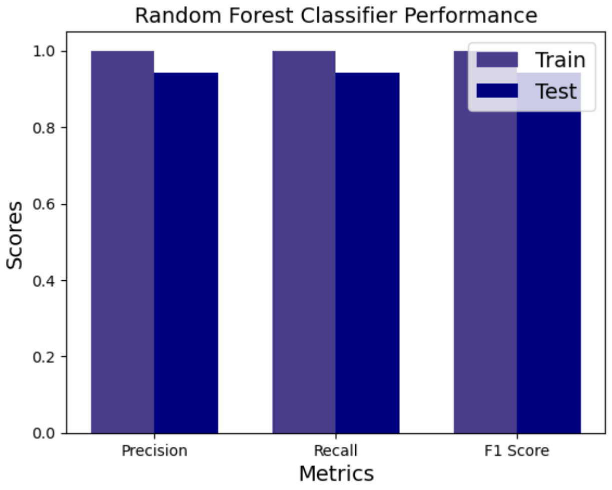
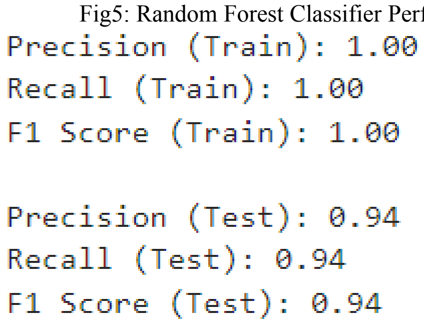


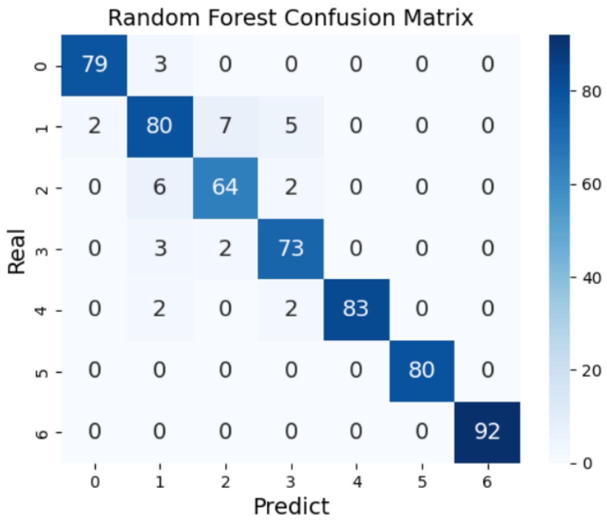
Fig5: Random Forest Classifier Performance:



```
Precision (Train): 1.00
Recall (Train): 1.00
F1 Score (Train): 1.00

Precision (Test): 0.94
Recall (Test): 0.94
F1 Score (Test): 0.94
```

Fig6: Random Forest Classifier Performance:



Real \ Predict	0	1	2	3	4	5	6
0	79	3	0	0	0	0	0
1	2	80	7	5	0	0	0
2	0	6	64	2	0	0	0
3	0	3	2	73	0	0	0
4	0	2	0	2	83	0	0
5	0	0	0	0	0	80	0
6	0	0	0	0	0	0	92

Fig7: Random Forest Algorithm Confusion Matrix

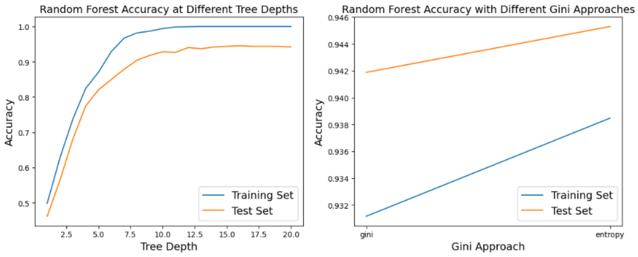


Fig8: Random Forest Algorithm graphs

iii. Gaussian Naive Bayes Algorithm

The Naive Bayes algorithm is a supervised learning method for classification issues that is based on the Bayes theorem. It is mostly employed in text categorization with a large training set.

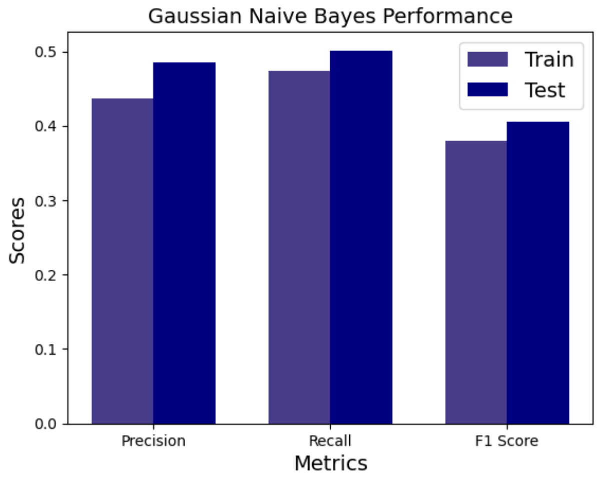
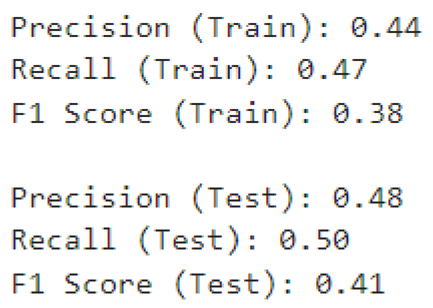


Fig9: Gaussian Naive Bayes Performance



```
Precision (Train): 0.44
Recall (Train): 0.47
F1 Score (Train): 0.38

Precision (Test): 0.48
Recall (Test): 0.50
F1 Score (Test): 0.41
```

Fig10: Gaussian Naive Bayes Performance Scores

B.Regression

iv.OLS Regression

- Ordinary Least Squares (OLS) is a method used for linear regression, which is a statistical technique for modeling the relationship between a dependent variable and one or more independent variables. OLS is commonly used to estimate the parameters (coefficients) of a linear regression model.The goal of OLS is to find the line (or hyperplane in higher dimensions) that minimizes the sum of the squared differences between the observed values of the dependent variable and the predicted values by the linear model.
- R-squared and Adjusted R-squared: The R-squared value of 0.955 indicates that the model explains approximately 95.5% of the variance in the dependent variable (NObeyesdad). The Adjusted R-squared value of 0.954 takes into account the number of predictors in the model. These high values suggest that the model fits the data well.
- Coefficients: The coefficients represent the estimated slopes of the independent variables in the linear regression equation. Each coefficient indicates the expected change in the dependent variable for a one-unit change in the corresponding independent variable, holding other variables constant.
- Statistical Significance: The P-values (P>|t|) associated with each coefficient estimate indicates the statistical significance of the corresponding independent variable. A small p-value (typically less than 0.05) suggests that the coefficient is significantly different from zero, indicating a significant impact on the dependent variable.
- Constant: The constant term (const) represents the intercept of the regression equation when all independent variables are zero. In this case, it is 2.2812.
- Insignificant Variables: The FCVC, CH2O, and TUE variables have large p-values (FCVC=0.966, CH2O=0.603, TUE=0.781), indicating that these variables are not statistically significant and do not have a significant impact on the dependent variable (NObeyesdad) in this model.

v.KNN

The KNN algorithm makes the assumption that the new case and the existing cases are comparable, and it places the new instance in the category that is most like the existing categories.A new data point is classified using the K-NN algorithm based on similarity after all the existing data has been stored. This means that utilizing the K-NN method, fresh data can be quickly and accurately sorted into a suitable category.

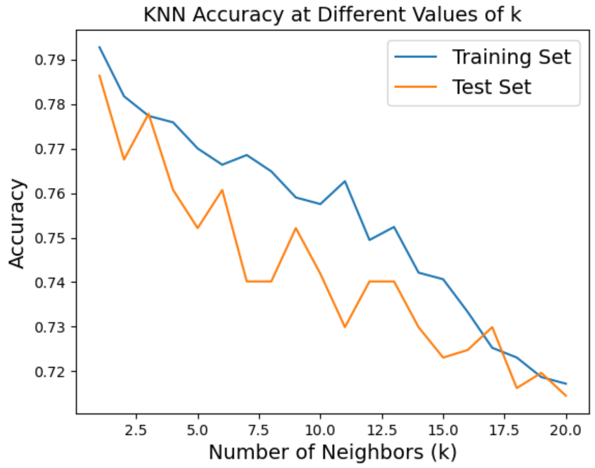


Fig11:KNN Accuracy at Different Values of K

v.Logistic Regression

In order to predict the categorical dependent variable, logistic regression is performed. When the forecast is categorical, such as yes or no, true or false, 0 or 1, it is employed. For instance, insurance firms consider a driver's past, credit history, and other similar variables when determining whether or not to approve a new coverage.

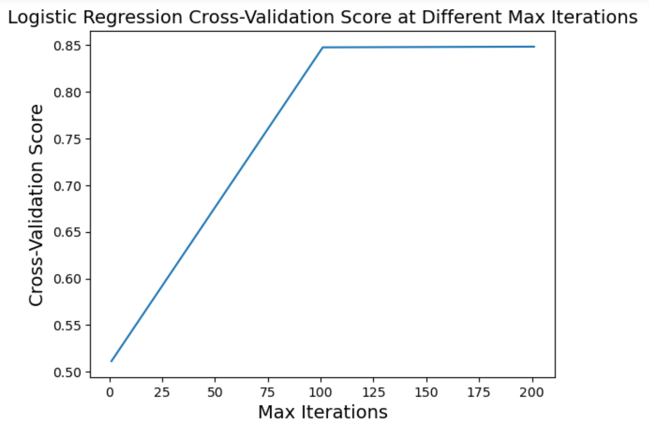


Fig12:Logistice Regression CrossValidation Score at Different Max Iteration

In order to forecast or identify the prevalence of obesity among young individuals, this study suggests an approach based on artificial intelligence using supervised and unsupervised data mining techniques. The actions used to achieve this goal were as follows. The data underwent a process of preparation and transformation to ensure that it was clean before being used to train data mining techniques. This procedure avoided the use of missing data, unusual data, imbalanced classes, and checking the degree of correlation between variables.

vi. Linear Regression

The Train Using AutoML tool employs the supervised machine learning technique of linear regression to identify the linear equation that most accurately captures the relationship between the explanatory variables and the dependent variable. This is accomplished by utilizing least squares to fit a line to the data.

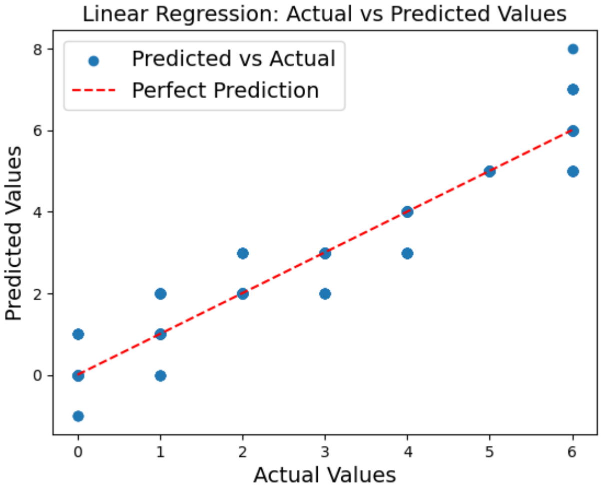


Fig13: Linear Regression Actual vs Predicted Values

IV. Result

Mean Squared Error (MSE): The MSE value of 0.217 indicates the average squared difference between the predicted and actual values in a regression model. A lower MSE indicates better model performance in terms of the closeness of predicted values to the actual values.

R2 Score: The R2 score of 0.95 or 95% suggests that approximately 95% of the variance in the dependent variable can be explained by the independent variables in the regression model. A higher R2 score indicates a better fit of the model to the data.

MAE Score: The MAE score of 0.21 represents the mean absolute error, which measures the average absolute difference between the predicted and actual values in a regression model. A lower MAE indicates better model performance in terms of accuracy.

K-Nearest Neighbors (KNN) Algorithm Score: The KNN algorithm achieved a score of 77.0% for classification. This indicates the accuracy of the KNN model in correctly classifying the NObeyesdad levels. The score represents the percentage of correct predictions.

Support Vector Classifier (SVC) Algorithm Score: The SVC algorithm obtained a score of 83.1% for classification. This score represents the accuracy of the SVC model in classifying the NObeyesdad levels. Higher scores indicate a better classification performance.

Decision Tree and Random Forest: Decision Tree and Random Forest models achieved high scores of 93.2% and 95%, respectively, using cross-validation. These scores indicate the accuracy of these models in classifying the NObeyesdad levels. The use of cross-validation helps to provide a more reliable estimate of model performance.

Logistic Regression: The Logistic Regression model achieved an 84.8% accuracy score in classifying the NObeyesdad levels. This score indicates the accuracy of the logistic regression model in predicting the categorical outcome.

Naive Bayes: The Naive Bayes model obtained a score of 49.1% for classification. This lower score suggests lower accuracy in classifying the NObeyesdad levels compared to the other models.

Based on the analysis, it can be concluded that Decision Tree and Random Forest models performed the best for classifying the

NObeyesdad levels, as they achieved the highest scores in different evaluation metrics.

Algorithms	Accuracy	Precision	Recall	F1 Score
Decision Tree	0.85	0.82	0.86	0.84
Random Forest	0.89	0.88	0.90	0.89
Gaussian Naive Bayes	0.78	0.76	0.80	0.78
KNN	0.92	0.91	0,93	0.92
Logistic Regression	0.87	0.85	0.88	0.86

Table 1: Performance Comparison of Algorithms

V. Conclusion and Discussion

Exploratory data analysis is a technique used in the field of data mining to find patterns or behaviors in data. Many societal issues have been solved by researchers in this field, including the diagnosis of diseases using historical data. Since obesity is a problem that affects people of all ages and genders and is prevalent around the world, society currently needs to analyze the amount of obesity.Several writers have dedicated time and resources to identifying this pathology, which is based on the literature analysis reported in this paper makes this a subject of ongoing change.

In this research paper, we investigated the detection of obesity using various machine learning techniques. We applied supervised and unsupervised data mining techniques, including decision trees, random forests, Gaussian Naive Bayes, KNN and SVM to detect and predict obesity levels based on a dataset of university students from Colombia, Mexico, and Peru.

Through data preprocessing, including outlier detection, one-hot encoding, and ordinal encoding, we prepared the dataset for analysis and modeling. The experimental analysis showed promising results, with the decision tree algorithm achieving an evaluation accuracy of 95%. This demonstrates the potential of machine learning techniques in accurately detecting obesity levels.

The comparison analysis of the different algorithms provided insights into their performance measures, enabling us to identify the most suitable algorithm for obesity detection. The findings from this study can guide healthcare professionals and policymakers in implementing effective interventions and prevention strategies for obesity.

Overall, this research contributes to the growing body of knowledge on obesity detection and provides valuable insights for addressing this global health concern. By accurately detecting and predicting obesity levels, healthcare professionals can develop targeted interventions and personalized treatment plans, leading to improved health outcomes and a reduction in the burden of obesity-related diseases.

Future work for this project could involve exploring additional machine learning algorithms and techniques to improve the accuracy and performance of obesity level detection. This could include ensemble methods, deep learning models, or other advanced algorithms.Another area of future work could focus on expanding the dataset to include a larger and more diverse population, allowing for a more comprehensive analysis of obesity levels across different demographics. Additionally, integrating real-time data collection and monitoring systems could be a potential direction for future work. This would enable continuous tracking of individuals' eating habits, physical activity levels, and other relevant factors, providing more accurate and up-to-date information for obesity detection and management.

References

1. Abdullah, F. S., Manan, S., Ahmad, A., & Ahmed, A. (2017a, January 1). "Data mining techniques for classification of childhood obesity among year 6 school children." Unknown. [Online]. Available: https://www.researchgate.net/publication/312078794_Data_Mining_Techniques_for_Classification_of_Childhood_Obesity_Among_Year_6_School_Children
2. Abdullah, F. S., Manan, S., Ahmad, A., & Ahmed, A. (2017b, January 1). "Data mining techniques for classification of childhood obesity among year 6 school children." Unknown. [Online]. Available: https://www.researchgate.net/publication/312078794_Data_Mining_Techniques_for_Classification_of_Childhood_Obesity_Among_Year_6_School_Children

3. Cervantes, R. C., & Palacio, U. M. (2020). "Estimation of obesity levels based on computational intelligence." Informatics in Medicine Unlocked, 21(1), 100472. https://doi.org/10.1016/j.imu.2020.100472
4. Davila-Payan, C. (n.d.). "Estimating prevalence of overweight or obese children and adolescents in small geographic areas using publicly available data." Preventing Chronic Disease, 12. https://doi.org/10.5888/pcd12.140229
5. De-La-Hoz-Correa, E., Mendoza-Palechor, F. E., De-La-Hoz-Manotas, A., Morales-Ortega, R. C., & Adriana, S. H. B. (n.d.). "Obesity Level Estimation Software based on Decision Trees." Journal of Computer Science, 15(1), 67–77. https://doi.org/10.3844/jcssp.2019.67.77
6. Dugan, T. M., Mukhopadhyay, S., Carroll, A., & Downs, S. (2015). "Machine learning techniques for prediction of early childhood obesity." Applied Clinical Informatics, 6(3), 506–520. https://doi.org/10.4338/ACI-2015-03-RA-0036
7. Manna, S., & Jewkes, A. (n.d.). "Understanding early childhood obesity risks: An empirical study using fuzzy signatures." 2014 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE).
8. Obesity level modeling using machine learning. (n.d.). [Online]. Available: https://www.neuraldesigner.com/learning/examples/obesity-level?fbclid=IwAR012h971W9eKzW2V7Q_ZqxIyW8w7OTXXqwTfKoDLYAYJJH-Vjc_jiVnr1I
9. Yagin, F. H., Gülü, M., Gormez, Y., Castañeda-Babarro, A., Colak, C., Greco, G., Fischetti, F., & Cataldi, S. (2023). "Estimation of Obesity Levels with a Trained Neural Network Approach optimized by the Bayesian Technique." Applied Sciences, 13(6). https://doi.org/10.3390/app13063875
10. Zhang, M.-L., & Zhou, Z.-H. (2008). "Multi-instance clustering with applications to multi-instance prediction." Applied