Application of Multiple Machine Learning Techniques in Classifying Obesity Level using Multivariate Dataset

(SAMPLE TITLE ABOVE)

\*\*Insert your name here

***Abstract*—** \*\*insert your statements here

1. Introduction

\*\*insert your statements here

1. Materials and Methodology

In this section, the research methodology and techniques are presented. The machine learning algorithms also discussed.

\*\*Insert your diagram here and explain it briefly.

*A. The Dataset*

\*\*Explain your datasets

*B. Data Visualization*

To fully understand the obesity dataset, for visualizing the data the study used the Scatter Plot for some of the attributes of the dataset.

\*\*SAMPLE

*Scatter Plot*

For the obesity dataset, the scatter plot visualization tool creates a two-dimensional plot. The researchers used an intelligent data visualization technique to generate useful projections, which resulted in the top-scoring pair of variables, Vocabulary richness W/C and Access, in terms of average classification accuracy.



Fig. 2 Scatter Plot for Age, Gender and Obesity Level Attributes

Figure 2 depicts the data in relation to determining which age and gender have a high degree of obesity. According to the graph, the female has a high risk of obesity, which is caused by a variety of factors. This means that at a specific age, a male is less likely than a female to gain weight.

*C. Machine Learning Classifiers*

\*\*Insert your machine learnings: SAMPLE BELOW

*AdaBoost*

Adaptive Boosting is a meta-algorithm for combining weak learners that adjusts to the 'hardness' of each training sample. It is applicable to both regression and classification problems. AdaBoost removes all occurrences with uncertain target values, continues categorical variables, removes empty columns, and imputes missing values with mean values as part of its preprocessing.

table ii. parameters used for the adaboost classifier

|  |  |
| --- | --- |
| **Parameters** | **Value** |
| Number of Estimators | 50 |
| Learning Rate | 1.00000 |
| Boosting Method | SAMME.R |
| Regression Loss Function | Linear |

There are four hyper-parameters in AdaBoost that should be optimized because they can affect model performance include the base estimator, max depth, number of estimators, and learning rate. For the AdaBoost model, decision trees were used as base estimators, and each decision tree-based model was built with a maximum leaf depth of 4 to avoid overfitting; the number of n estimator trees was set to 50. As a result, decision tree-based weak classifiers helped to generalize (increase) the performance of the strong classifier. To adjust the contribution of each model for a strong classifier, the learning rate was set to one [22]. This implies that through adjusting the value of the estimators, the better performance on the model.

1. Results and Discussion

\*\*Discussed the result or accuracy of each machine learning: The results of the four proposed machine learning algorithms, AdaBoost, Gradient Boosting, KNN, and Random Forest, were presented in this section. It was used in attributes obtained from a dataset of 2111 rows and 17 columns from the UCI ML repository, which included obesity causes.

1. Conclusion

\*\*insert your statements here

1. References
2. Konstantakopoulos, F., Georga, E. I., & Fotiadis, D. I. (2021). 3D Reconstruction and Volume Estimation of Food using Stereo Vision Techniques. *In 2021 IEEE 21st International Conference on Bioinformatics and Bioengineering (BIBE) (pp. 1-4). IEEE.*

\*\*SAMPLE ABOVE