Obeisty Level Prediction

Amin Majdi (amin_majdi@student.uml.edu)
Mingyu Wan (mingyu_wan@student.uml.edu)
Eric Zabele (eric_zabele@student.uml.edu)

I. Objective

Obesity is thought to be one of the most concerning health issues in the modern world. Thus, a good indicator that helps people to know about their obesity status is crucial. The objective of this project is to perform machine learning techniques to precisely categorize the obesity status of people into seven different levels (Insufficient Weight, Normal Weight, Overweight Level I, Overweight Level II, Obesity Type I, Obesity Type II and Obesity Type III) based on 16 different factors (gender, age, height, weight, family records, frequency of eating high caloric food, frequency of eating vegetables in meals, number of meals in a day, eating between meals, smoking habits, the daily amount of drinking water, monitoring the calories received from the daily meals, having physical activity, technological device usage in a day, alcohol drinking habits and the type of the transportation that the person uses frequently) [1].

II. Current State of Art

Many different machine learning algorithms and datasets have been used to analyze the factors that influence obesity.

Adnan et al. [2] investigated the childhood obesity problem that has a low ratio of negative samples compared to the positive samples by presenting a framework with a hybrid approaches based on Naïve Bayes for prediction and Genetic Algorithm for parameter optimization. Their approaches reached a 75% improvement in prevision.

Dugan et al. [3] predicted childhood obesity using RandomTree, RandomForest, ID3, J48, Naïve Baye and Bayes. Their results show that ID3 model has the best performance of precision of 85% and sensitivity of 89%.

Eduardo et al. [1] generated a dataset for obesity level estimation containing 17 attributes and 2111 instances and then trained their model using Decision Trees, Bayesian Network and Logistic Regression. Their results show that decision tree algorithm has the best precision (97.4%) to predict obesity level.

Cervantes and Palacio [4] selected 178 instances from the dataset generated in [1], 81 men and 97 women with ages between 18 and 25 years, to detect the obesity level using Simple K-Means, Decision Tree and Support Vector Machines algorithms. They compared these techniques by evaluation metrics of precision, recall, true positive rate, false-postive rate and ROC area and reached a higher precision (98.5%) than [1-3].

III. Approach

With our goal to classify obesity in mind, we will be using Python and supporting libraries such as Sklearn and Pandas to complete this project. Before feeding our dataset into an algorithm we will first perform some exploratory data analysis to get a better understanding of our data. Once we have a clear understanding, we will preprocess our data by looking for any perfect correlations for potential dropout, perform any necessary feature scaling and then split our data into training and testing data and then move on to training our models. This being a supervised learning problem, a few models we have in mind to potentially use are Naïve Bayes, ID3, SVM's, Random Forest, and possibly an ANN. Once trained, we can calculate our accuracy and compare our results against others who have used this dataset and subsequently try to improve our performance and analyze the contributing factors to obesity.

IV. Dataset

The dataset involved in this proposal is from UCI Machine Learning Repository.

V. Timeline

Specific Objectives	Three months		
	1	2	3
Do literature review to see how previous studies built, trained and analyzed the model regarding obesity level			
Use the aforementioned algorithms in the approach section to train the dataset and compare against other papers success.			
Refine our models to improve accuracy and then analyze the main factors contributing to obesity			

VI. Roless and Task

Amin Majdi and Mingyu Wan conceived this idea and wrote the manuscript. Eric Zabele carried out various algorithms to train the model.

References

[1] E. De-La-Hoz-Correa, F. E. Mendoza-Palechor, A. De-La-Hoz-Manotas, R. C. Morales-Ortega, and S. H. Beatriz Adriana, "Obesity Level Estimation Software based on Decision Trees," *Journal of Computer Science*, vol. 15, no. 1, pp. 67-77, 2019.

[2] M. H. B. M. Adnan, W. Husain, and N. A. A. Rashid, "A Hybrid Approach Using Naïve Bayes and Genetic Algorithm for Childhood Obesity Prediction," *2012 International Conference on Computer & Information Science (ICCIS)*, vol. 1, pp. 281-285, 2012.

[3] T. M. Dugan, S. Mukhopadhyay, A. Carroll, and S. Downs, "Machine Learning Techniques for Prediction of Early Childhood Obesity," *Appl Clin Inform*, vol. 6, no. 3, pp. 506-20, 2015.

[4] R. C. Cervantes and U. M. Palacio, "Estimation of obesity levels based on computational intelligence," *Informatics in Medicine Unlocked*, vol. 21, 2020.