Comparative Analysis of Machine Learning Algorithms for Predicting the Classification of Quetelet Index

Stella Ogunleye and Olawale Awe
Department of Mathematical Sciences, Anchor University Lagos, Nigeria

- The Quetelet Index, generally known as the BMI (Body Mass Index), is a basic weight-for-height index that is commonly used to identify humans as underweight, normal-weight, overweight or obese.
- The purpose of this study is to compare the performance of four machine learning algorithms for predicting the classification of quetelet index among adolescents and young adults via data obtained from a survey administered to 150 adolescents across several departments in a university community in Africa.

Related publications:

❖ G, Iossa S (2015) Fat quality influences the obesogenic effect of high fat diets.
❖ Chiong, R., Fan, Z., Hu, Z., & Chiong, F. (2021). Using an improved relative error support vector machine for body fat prediction.
❖ DeGregory,Thomas, D. M. (2018). A review of machine learning in obesity.

stogunleye@gmail.com

3rd Conference on Statistics and Data Science
Salvador, Brazil (online)
October 28-30, 2021

Problem, Data, Objectives

- We are surrounded by huge amounts of large-scale high dimensional

Data (Li and Liu et al., 2016). It is therefore desirable to reduce the dimensionality of data for many learning tasks due to the curse of dimensionality. Feature selection has shown its effectiveness in many applications by building simpler and more comprehensive model, improving learning performance, and preparing clean, understandable data. This is the problem this project is set to solve, particularly with respect to feature selection involving attributes associated with the classification of BMI using the Boruta algorithm before determining the best machine learning algorithm suitable for the classification of BMI.

- The aim of this project is to comparatively explore the performance of four (4) machine learning algorithms for the classification of quetelet index.

Specific objectives of this study are to:

(i) Determine the most important variables for predicting/classifying BMI using the Boruta algorithm.

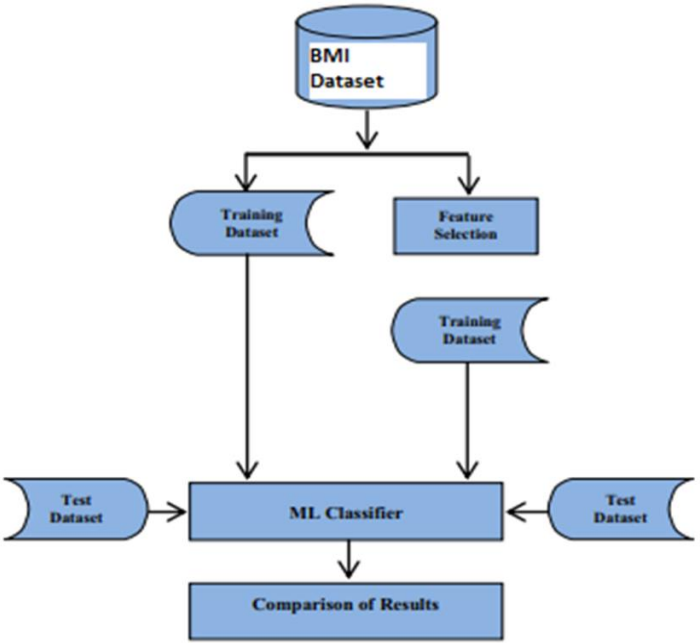
(iii)Compare the performance of four (4) machine learning algorithms on BMI classification.

stogunleye@gmail.com

3rd Conference on Statistics and Data Science
Salvador, Brazil (online)
October 28-30, 2021

Methods

- The implementation of the ML algorithms and feature selection methods were carried out using the R software.
- The practical was carried out using feature selection preceding the applications of 4 ML classifiers.), the BMI dataset was subjected to feature selection in order to remove the redundant and non-relevant features.
- The resulting dataset was then divided into training/test subsets in the usual ratio of 70/30 percent. Several classification experiments were conducted and the results were compared for the scenario.



stogunleye@gmail.com

3rd Conference on Statistics and Data Science
Salvador, Brazil (online)
October 28-30, 2021

Results and Conclusions

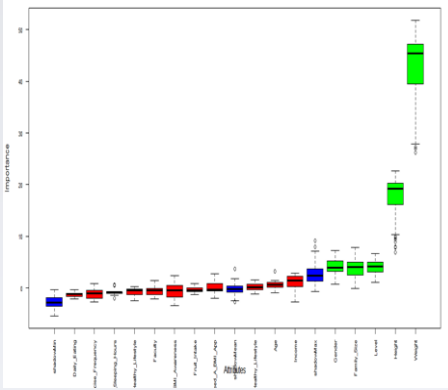
Thus, from the results obtained, we concluded that:

- The Linear Discriminant Analysis appeared to be the best ML algorithm.
- Females are more obese and overweight than Males.
- Feature selection is able to improve BMI predictions when the right features are selected e.g. weight, height, family size, gender and level.
- Majority of the respondents are slightly aware of what BMI is all about.

A crosstabulation of Gender and BMI categories

			BMI_Cat			
			Normal Weight	Overweight	Obese	Total
Gender	Male	Count	46	22	6	74
		% Of Total	30.7%	14.7%	4.0%	49.3%
	Female	Count	41	17	18	76
		% Of Total	27.3%	11.3%	12.0%	50.7%
Total		Count	87	39	24	150
		% Of Total	58.0%	26.0%	16.0%	100.0 %

Awareness		Frequency	Percent
	Not Aware	66	44.0
	Slightly Aware	43	28.7
	Aware	21	14.0
	Highly Aware	20	13.3
	Total	150	100.0



Performance Metrics	LDA	NAÏVE BAYES	SVM	RF
Accuracy	94	84	75	82
Error	6	26	25	18

3rd Conference on Statistics and Data Science
Salvador, Brazil (online)

stogunleye@gmail.com