Body performance

Sumaiah ali alghamdi

Abstract:

This project aimed to understand body performance according to bodies different variables and data that's related to how this body could look like, perform and some medical measurements values taken from them, then assign them to appropriate class. Assigning body performance to class are done by using machine learning models that will described in detailed later in the document. The data set that are trained and tested in this project provided by Kaggle and labeled using classification models, with sklearn Stacking Classifier and get approximately 75% accuracy.

Design:

This project is one of the T5 Data Science BootCamp requirements. Data provided by Kaggle has been used in this project and present 4 distinct classes which is: A, B, C and D, A is the best for classifying bodies performance. Classifying done using machine learning models will enable who ever interested to extract will performed plan to each class according to their data aiming to maintain and enhance the body performance.

Data:

The data set provided from Kaggle in .csv format, containing 13393 records, every record has 12 features each one of them containing important measurements that could helping classify person to appropriate class, this features cleaned to easy use and implement using classifying modeling in ML.

Algorithm:

Features engineering:

- Checking for missing values and fill it if necessary “there is no missing values”.
- Rename features to easier and valid names to call and use.
- Identifying label features unique values.
- Cleaning label feature and gender from spaces, lower cases if there is.
- Replace string values into numerical then change feature type to int.
- Selecting unique variable in label feature to be categorical features.

Models:

k-nearest-neighbor, logistic regression, random forest classifier, decision tree classifier were used before stacking model classifier that get the higher accuracy.

Model evaluation and selection:

The entire dataset containing 13393 records, every record has 12 features split into 80/20 train vs test, and all scores reported below were calculated with classification report on both splits.

Final results of using stacking modeling classifier:

Accuracy: 75%

Precision: 75%

Recall: 76%

F1: 75%

Tools:

- Pandas for data manipulation
- Scikit-learn for modeling
- Matplotlib and Seaborn for plotting

Communication:

The slides are provided besides details are provided at this document of the project. Feel free to any pull requests.

Conclusion:

Assigning every person to appropriate class according to their data and measurements helps who ever concerned to improve bodies performance by giving them studied plans that stands for every one problem solving such as : loosing/gaining weight, loosing/gaining fat, increase body fitness , power and strength …etc. or plan for maintaining the perfect body performance for whom assigned to the best class.