PROJECT SYNOPSIS

Calories Burn Prediction Analysis Using XGBoost Regressor and Linear Regression Algorithms

(Tentative project Name)

Submitted towards the partial fulfillment of the criteria for award of Post Graduate In Data Analytics by Imarticus

Submitted By:

HARIHARASUDHAN P (IL031248)

Course and Batch: Post Graduate in Data Analytics & Feb-2022(PGA-22)

Scope & Objective:

The overarching idea of this research project is to make a comparative study of machine learning algorithms to predict the calories burn during the workout. In this paper we first build a machine learning systems that can predict the amount of calories burnt during exercise. In today's world many people are inquisitive about the workout that they do and the weight loss plan that they take and how much calorie do they burn once they workout.

Business Problem Statement:

To predict the calories burn during the workout with Machine Learning, Here I will be using the XGBoost regressor and Linear Regression. As it is one of the best approaches for regression analysis.

Data Sources:

1. The repository that we used for dataset is Kaggle. There are two csv files which contains 15000 instances and 7 attributes.
2. The data set from Kaggle repository contains attributes of each person's details including their gender, age, workout duration, heart rate, body temperature, height and weight. This dataset is taken as the training data. And the second calories dataset contains target class which have the calories burned by corresponding person.

Analytics Tools:

- **Jupyter Notebook:** collaborative work capabilities.
- **Pandas:** A python data analysis library enhancing analytics and modelling.
- **Matplotlib, Seaborn, plotly:** A python machine learning library for quality visualizations.
- **Statsmodels:** A python that enables us to estimate and analyze various statistical Models.

Analytics Approach:

- To create a algorithms to comparative study among the proposed technique and check the performance of the model with statistical

KPIs,Timelines,Milestones(proposed):

KPIs:

- R-Square
- Adjusted R-Square
- Mean Square Error(MSE)
- Root Mean Square Error(RMSE)
- Mean Absolute Error(MAE)
- Score

Timelines:

Expected to complete the project by 15-08-2022

Milestone:

Expected processing and making prediction based on the dataset achieve a score of 95%.

File Format:

The datasets are provided in CSV format, with the following features:

Retrieved from kaggle:

1) User_Id
2) Gender
3) Age
4) Height
5) Weight
6) Duration
7) Heart_Rate
8) Body_Temp
9) Calories

Output:

- Best fit of calories burn during the exercise.