Professional Certificate in Machine Learning (ML) and Artificial Intelligence (AI)

Capstone Project
Predicting Weight Loss

Presented by: Rudy Silva Mera




Overview

- Business Understanding
- CRISP-DM
- AI Application / Solutions
- Important Findings
- Recommendations

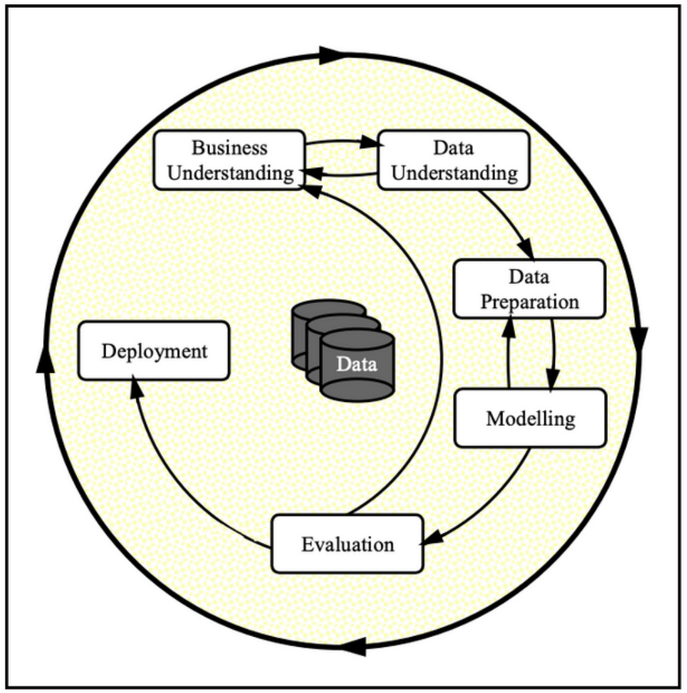
Business Understanding

- From a business perspective, I will be trying to identify the key drivers for this prediction question,

" Can I build ML models to predict the calories of an individual, whether a person would lose weight or not if we use features like exercise types, food intake, water consumption, supplement, dieting, etc? "

- For the built model or ensemble of models, can we use these models to build Artificial Intelligence (AI) applications that consumers can use to predict their weight loss/calorie loss over some time as opposed to the current method, where you do the following
  - Enter your details like weight, height, etc
  - Enter your target/goals with regards to weight loss
  - Track and log your exercise, food and water consumption
  - Application displays the weight loss and progress towards your target
- These AI applications can be incorporated into applications for websites, and physical devices that consumers can
- use to determine the factors that contribute to weight loss and help them on fitness programs.

CRISP-DM



CRISP-DM is an open standard process model for ML projects. This model is used during this project to build the ML models. The model consists of six phases:

- Business understanding: In this phase, the goal is to understand the business problem that the data mining project is trying to solve. includes understanding the business goals and the data that is available
- Data understanding: With the available data, we use several Exploratory Data techniques to understand the data that is available with regard to format, quality, and distribution of factors that would help with the predictions
- Data preparation: the data is prepared for the ML models. This includes cleaning the data, removing errors, and transforming the data into a format that can be used by the ML algorithms.
- Modeling: In this phase, we build several models. The model is used to predict the future values of the data or to classify the data into different categories.
- Evaluation: the model is evaluated to see how well it performs. This includes measuring the accuracy of the model and the performance of the model.
- Deployment: the model is deployed to production. This includes making the model available to the users and making sure that the model can handle the volume of data that it will be exposed to.

AI Applications

AI Application 1
Linear Regression (LR)

- LR is a way to estimate the relationship between two variables. This can make predictions about future values of the variables (i.e., Calories loss)
- The application uses variables like Total Distance and No of Steps taken
- Algorithm: GradientBoostingRegressor

Train Score: 85%
Test Score: 61%

AI Application 2
Decision Tree Classifier (DTC)

- DTC is a ML model that can be used to classify data. It does this by building tree of decisions based on the value of a variable (I.e., no of steps etc.).
- The tree is then used to classify new data like weight loss flag
- Algorithm: Decision Tree Classifier

Train Score: 62%
Test Score: 63%

AI Application 3
ML Classifications

- The goal of classification is to identify patterns (i.e., no of steps taken etc.) in the data and use those patterns to predict the class of new data points (i.e., weight loss)

- Algorithm: K-Nearest Neighbor (KNN)

Train Score: 75%
Test Score: 59%

Important Findings

- Three notebooks are included in this project which demonstrates that several Machine Learning (ML) models can be used to predict the number of calories loss and whether a person can lose weight.
- The models used data from 33 Fitbit users over a period of a month. The model prediction provided training accuracy of 65% to 100% depending on the ML algorithms. Predict weight loss and number of calories lost with an accuracy of 65% to 100%.
- Each notebook includes a Findings section for the models, but a summary is as follows:

1) For ML application 1, we tried several models with all the variables and a couple of strong inputs (i.e., Total Steps and Total Distance). The model provided an accuracy score of 62% to 100% for training and 37% to 67% for test. Important Variables were Total Steps and Total Distance

2) For ML application 2, the models produced an average accuracy score in the 60% range

3) For ML application 3, the K-Nearest Neighbor (KNN) had the best training score of 75% and a test score of 58%

Important Findings

- The recommendation would be to use an ML application with the GradientBoostingRegressor algorithm with a training score of 86% and a testing score of 61%. This provides room for training with more data especially data with users with steps over 30k steps and distances over 5 miles.
- For the Next steps based on the room to improve the models, I would recommend collecting more data where users record the following:

1) Daily water consumption
2) Calories intake (i.e., Protein, Carbs, etc.)
3) User demographics (i.e., Age, Gender, Weight, etc.)

- This would be used to train the model to provide better predictions for users based on some features listed above.
- This ML model would be deployed as an AI application and Weight loss applications can use the model for users to determine what they need to do to lose weight or estimate the amount of Calorie loss based on those features.

Thank you for your attention