HW2_Readme

November 18, 2021

1 HW #2: Modeling Patient Engagement

Team 1 :

- Shobhit Aggarwal (shobhita)
- Ho Kit Fu (hoikitf)
- Shiyuan Li (shiyuan2)

1.1 Hypothesis Testing

Testing hypothesis from Group 2

Jupyter Notebooks Submitted :

- **HW2_Part1_Hypothesis.ipynb**

1.2 Clustering

Cluster patients by similar sequences of engagement using k-medoids.

Use k = 3, time period as month, and the engagement level as highest level of engagement in the month (this was the method implemented by Deloitte in 2018)

Jupyter Notebooks Submitted :

- **HW2_Part2_EngMatrix.ipynb** - Preparing data matrix for clustering
- **HW2_Part2_Clustering.ipynb** - Use K-Medoids to cluster in three groups. Contains Silhoutte score and cluster visualisation

1.3 Predictive Modelling

Develop two models that use predictor variables for a patient in one time period to predict the engagement level the patient will fall into the next time period.

Group1: Based on the predictive variables in month i, predict the highest level of engagement in the time period in month i+1

Jupyter Notebooks Submitted :

- **HW2_Part3_Predicting.ipynb** - Contains random forest model and logistic regression model for predicting the values

Note : We see that for the first a few months, the prediction is pretty accurate. This might due to the reason that we don't have the data for every month of a patient. And the date of the data is almost the early months.