The Hong Kong Polytechnic University

Department of Electrical and Electronics Engineering

EIE4430 Honours Project

2024-2025 Semester 1

Student Name: Chan Hou Ting Constant (21034774d)

Project Title: **Machine learning model to predict the risk of diabetes**

Progress Report (1/11/2024)

**Works did in past month**

I added AdaBoost and Decision Tree to the codebook, also applied SHAP value (summary plot) and ROC to each model to see feature importance and performance of the models. There were some troubles when I applied SHAP value to AdaBoost as the shap package did not support the AdaBoost. Then, I sorted the codebook to made it more understandable and applied the model to the dataset for the prediction. I found the accuracy of the prediction on the dataset and test set were close to each other. The following tasks were to solve the problem of applying SHAP value to AdaBoost and found the suitable hypermeters to get the best results.

**SHAP value**

一張含有 文字, 圖表, 螢幕擷取畫面, 字型 的圖片

自動產生的描述

**ROC**

一張含有 文字, 圖表, 行, 螢幕擷取畫面 的圖片

自動產生的描述

**Accuracy (Prediction vs test set)**

**XGBoost**

一張含有 文字, 螢幕擷取畫面, 字型, 數字 的圖片

自動產生的描述

一張含有 文字, 螢幕擷取畫面, 數字, 字型 的圖片

自動產生的描述

**~=** 80.36% Accuracy [100-(151(predict error)/769(total count))\*100]

**Random Forest**

一張含有 文字, 螢幕擷取畫面, 字型, 數字 的圖片

自動產生的描述

一張含有 文字, 螢幕擷取畫面, 數字, 字型 的圖片

自動產生的描述

~= 79.45% Accuracy [100-(158/769)\*100]

**Logistic Regression**

一張含有 文字, 螢幕擷取畫面, 字型, 數字 的圖片

自動產生的描述

一張含有 文字, 螢幕擷取畫面, 數字, 字型 的圖片

自動產生的描述

~= 76.72% Accuracy [100-(179/769)\*100]

**AdaBoost**

一張含有 文字, 螢幕擷取畫面, 字型, 數字 的圖片

自動產生的描述

一張含有 文字, 螢幕擷取畫面, 數字, 字型 的圖片

自動產生的描述

~= 79.84% Accuracy [100-(155/769)\*100]

**Decision Tree**

一張含有 文字, 螢幕擷取畫面, 字型, 數字 的圖片

自動產生的描述

一張含有 文字, 螢幕擷取畫面, 數字, 字型 的圖片

自動產生的描述

~= 77.37% Accuracy [100-(174/769)\*100]