Programming assignment #5

Course: CHE1148H - Data Process Analytics

Coding options

Python in .ipynb notebooks is recommended for this assignment.

1 Partial dependency plots

Generate the partial dependency plots for the top 5 features in your best Random Forest model from the previous assignment. Use the population distribution (%) as the primary y-axis and the response rate (%) as the secondary y-axis. To identify the best bin edges you can use a single tree model with the feature of interest **or** you can pick values that you feel are relevant. Make sure that appropriate scaling of the axis is used for the best visualization and interpretation. **+3 Bonus points** if you identify and plot any interesting 2D plot (i.e. two variables at a time).

Deliverables:

1 notebook: 1 graph for each feature and Markdown language comments.

2 LIME and SHAP interpretations

In Feb-2014, clients CS1350 and CS1200 emailed your customer service department complaining about the company's decision to market to them (or the lack of it). The customer service department has asked for your help to figure out why the clients received and did not receive a marketing offer, respectively. You will implement the LIME algorithm to calculate what particular features contributed to our decision for these two clients. **+3 Bonus points** if you manage to calculate and plot the SHAP values for these two clients.

Deliverables:

1 notebook: one dashboard per client showing the factors that contributed to our decision; comments and interpretation of the results.