Campaign Response Model

Homework 08

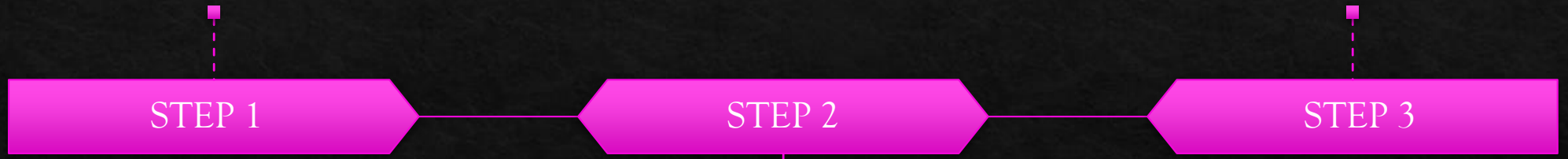
Step for Model XGBoost Tuning

Campaign response analysis should be observes and analyzes recent behaviors, or some behaviors, it will be more accurate.

K-fold Cross Validation(CV) is to divide the data into k equal parts to create and test the model (train + validate), calculate the mean accuracy or error (i.e. model performance) before the model can be used to predict the test set data.

K-fold Cross Validation(CV) provides a solution to this problem by dividing the data into folds and ensuring that each fold is used as a testing set at some point.

By dividing the data into k- folds =5 , 7 and 10.



Clean data : Original Data is dated from 11-May-2011 to 16-Mar-2015, selected data only from 1-Jan-2013 to16-Mar-2015.

Noted: The larger the data, the larger the k-fold value, the more accurate results, but it takes much longer times to train the data.

Conclusion for Model XGBoost Tuning

XGBoost	Original Data (11-May-2011 to 16-Mar-2015)	K-fold = 5	Clean Data (1-Jan-2013 to 16-Mar-2015)	K-fold = 5	Clean Data (1-Jan-2013 to 16-Mar-2015)	K-fold = 7	Clean Data (1-Jan-2013 to 16-Mar-2015)	K-fold = 10
XGBoost model – SMOTE RFM test set f-1 score	0.67		0.70		0.72		0.72	
XGBoost model – SMOTE CLV test set f-1 score	0.66		0.72		0.70		0.70	
Best AUC Score	0.7089		0.7481		0.7511		0.7520	