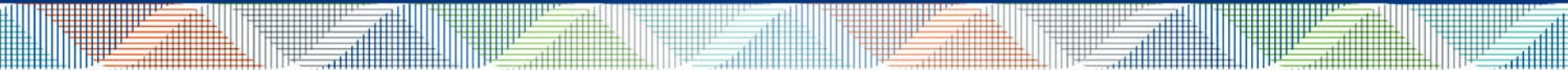


American Express Campus Analyze This 2017

Analyze This



Team Details

Team Name : No one

Name	Campus	Roll No.	Email Id
Prajjwal Jaiswal	IIT Kharagpur	14MT3EP01	prajjwaljaiswal.iitkgp@gmail.com
Manisha Meena	IIT Kharagpur	14CE31006	manu96meena@gmail.com

Imputation of missing value for “mvar12”

Feature Engineering: inclusion of new variable like ratio of total spend among four quarters

Outcome variable: 0 for none , 1 for Supp, 2 for Elite and 3 for Credit

UNDERSAMPLING to balance the dataset

Model using Extreme Gradient Boosting(xgb)

Parameter Tuning of the model

Strategy to decide final list

Created dataframe with probability of Supp, Elite, Credit and None for each Test id.

Calculated Maximum among Supp,Elite, Credit and None for each row

Excluded rows for which 'None' was maximum

Sorted in increasing order of Variable 'None'

Extracted top 1000 rows

Sorted in decreasing order of Maximum value among Supp, Elite and Credit

Details of each Variable used in the logic/mode/strategy

New variables were included to train the model like ratio of spendings in compared to different quarters.

Features with extremely low calculated feature importance were removed

Probablilty was taken in to account the final logic

Reasons for Technique(s) Used

Ratio of people who accept offer to those who don't accept is high (imbalanced dataset)

Normally classifiers don't perform well on these datasets(weak classifier)

Xgb uses a number of such weak classifiers to make a strong one. It learns from previously misclassified instances and imposes penalty for them in next classifies.

This becomes very fruitful in cases where model tends to predict single class (in imbalanced data)