Neustar Data Science Exercise

Context

This is a simulated dataset within the realm of credit card fraud decision. As a credit card issuer, you need to understand whom to give credit cards to. You have a dataset at your disposal to help you make those choices.

Questions

Perform an analysis of the following dataset; we want to understand:

- ➔ How much fraud is in this dataset?
- ➔ What are the other variables distributed?
- ➔ How well can we explain the present fraud using these variables?
- ➔ What kind of model would you use to predict incidences of this fraud phenomenon and why would you use that kind of model over all others?
- ➔ Implement the model you suggest above, what are its performance characteristics?

These questions are a framework to guide you in approaching this problem, but fundamentally we want you to showcase the extent of your abilities and knowledge. Please enclose any additional analysis you believe to be relevant and it will be considered.

Instructions

1. Please use Jupyter notebook/PyCharm or any IDE with Python or R to perform your analysis.Share across the code with comments to us (.py,.ipyb etc files)
2. For the questions asked on the exercise and any additional analysis which you want to list out please use a Powerpoint deck or a Word/PDF document to answer the same.This deck/document should be sent along with the code
3. Variables description for the dataset –card.txt , are provided in Table 1 below.

Good luck!

Table 1: Variable description

Column Name	Example value	Field explanation
id	0	Unique identifier for each row
dist_latest_transaction_address_km	2	Distance between place of application and the latest transaction with another credit card prior to the application date
email	djellybrande2@yolasite.com	email address used to apply
application_date	2019-03-17 03:57:12.247891	date of application
site_visits_A	0	site visits on web site A
site_visits_B	0	site visits on web site B
site_visits_C	0	site visits on web site C
credit_limit	$1,032.00	Credit limit in dollars
number_of_transactions	23	Integer number of transactions
is_fraud	0	Flag for fraud present