

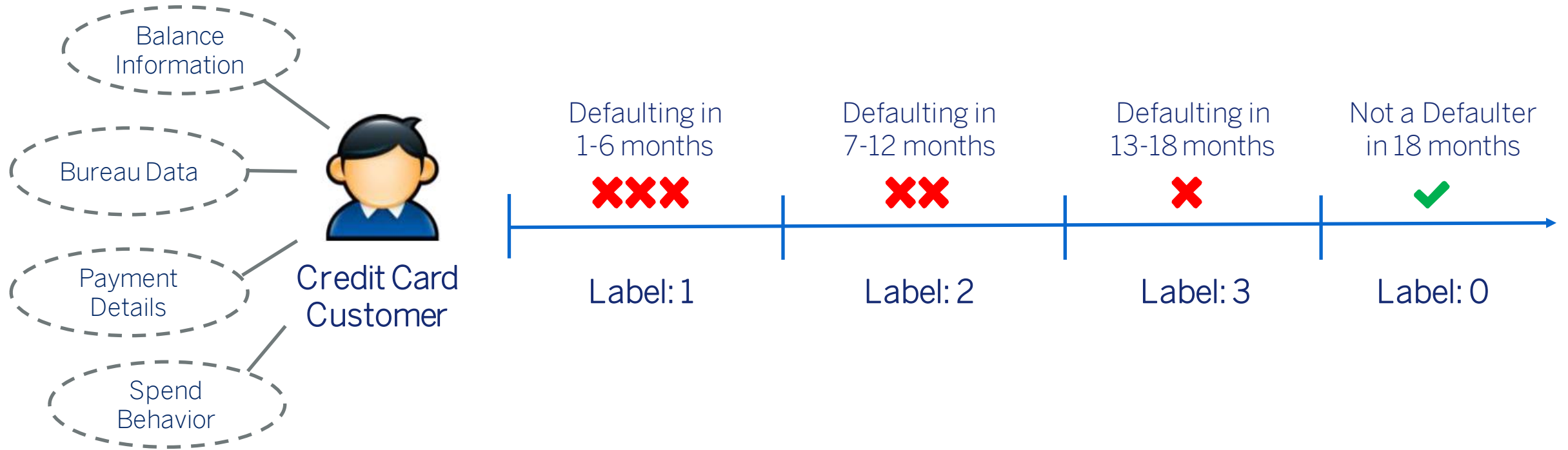
The American Express Campus Super Bowl Data Science Challenge

Overview: Credit Risk - Can you better the best?

- Over the years, American Express has been known for its impeccable record of maintaining industry best (lowest) credit and fraud loss numbers.
- In this year's challenge, we are looking for those who can better the results we already produce.
- Once a credit card is given to an individual, American Express keeps a tab on individual's ability to pay back the expenditure (spend) on the card.
- American Express manages credit risk through a model that predicts when a customer will default over the next few months. Predicting this '**Time to Default**' for a customer helps American Express in taking timely and appropriate actions to help the customer manage his/her debt levels.

Problem Statement

Given several dimensions of information about a Credit Card Customer, can you predict time to default for that customer?



Default = Inability to pay due amount on the credit card

Data

- For this challenge, we will be using information of randomly chosen customers in 3 different time periods as 'Training', 'Validation' and 'Test' datasets.
- Each customer will have a unique ID and several numeric as well as character features related to balance information, bureau data, payment details, spend behavior, delinquency attributes etc. The details of the variables can be found in the dictionary
- Default Flag in Training data will take values 0/1/2/3
 - Label 0: Not a defaulter in next 18 months
 - Label 1: Defaulting in 1-6 months
 - Label 2: Defaulting in 7-12 months
 - Label 3: Defaulting in 13-18 months

Training Data

- 458,913 rows
- 190 columns
 - A unique ID column
 - Default_Flag with values 0/1/2/3
 - 188 independent variables

Validation Data

- 467,966 rows
- 189 columns
 - A unique ID column
 - 188 independent variables

Test Data

- 456,655 rows
- 189 columns
 - A unique ID column
 - 188 independent variables

Stages of Competition

This competition will run in 3 rounds :



Round 1: Leaderboard Submissions

- Participants will train their models using labelled Training dataset
- Score the Validation dataset using their chosen model
- Upload IDs and predicted labels of 0/1/2/3 on Validation dataset in the form of a csv without a header
 - Submitted csv should have only 467,966 rows and 2 columns of ID and predicted label (no header)
 - There should not be any duplicates in ID column and should match with IDs shared in validation dataset
 - Predicted label should take values of 0/1/2/3 only
- **Evaluation Metric:** “Accuracy” of the prediction is calculated by comparing actual labels and predicted labels
- Teams will see their accuracy post each submission.
- Top 20 teams with highest Accuracy will be displayed on a Leaderboard

		Predicted Labels			
		0	1	2	3
Actual Labels	0	N00	N01	N02	N03
	1	N10	N11	N12	N13
	2	N20	N21	N22	N23
	3	N30	N31	N32	N33

$$\text{Accuracy} = \frac{N00 + N11 + N22 + N33}{\text{Total Number of Records}}$$

Round 2: Test Data Scoring

- Top 10 teams from Round-1 based on Leaderboard accuracy will move to Stage-2
- In Round-2, using their chosen model, participants will score Test dataset
- Upload IDs and predicted labels of 0/1/2/3 on Test dataset in the form of a csv without a header
 - Submitted csv should have only 456,655 rows and 2 columns of ID and predicted label (no header)
 - There should not be any duplicates in ID column and should match with IDs shared in Test dataset
 - Predicted label should take values of 0/1/2/3 only

Round 3: Virtual Interaction



- Top teams from Round-2 will be asked to share details of the codes used to arrive at the model
- Teams will also create a presentation detailing their approach and insights derived. They will be asked to present the same to a panel.
- Presentations will be evaluated based on novelty in approach, ease of execution of the solution and clarity in articulation.



Top 3 teams will be selected as winners based on Test Data scoring accuracy as well as scores from the presentation

The background of the image is a repeating pattern of the American Express World Service logo. The logo consists of a circular emblem with a globe in the center, surrounded by the words "AMERICAN EXPRESS" in a banner above and "WORLD SERVICE" in a banner below. The pattern is rendered in a light blue color on a darker blue background.

THANK YOU