**背景**

交易欺诈作为信用卡行业面临的主要贷后风险业务问题，每年都使信用卡行业遭受巨额损失。基于大数据机器学习开发出高效的交易欺诈识别模型一直是金融行业的主要挑战之一。本次大赛以此作为主题，具体的数据和要求如下：

**内容**

本次建模的目标是识别交易是否欺诈，提供的数据集为某行某月一部分信用卡客户在两天内发生的交易，训练集包含256327条交易记录，有443条为欺诈交易，测试集28480条，需要参赛者提交预测结果。数据集高度不平衡，正类（欺诈）占所有交易的0.172%。

数据集中只包含数值型变量，由于数据保密性问题，我们不能提供原始特征和数据上的更多背景信息，除‘Time’和‘Amount’外的其他变量都进行了PCA变换，其中特征V1，V2，…,V28是由PCA得到的主成分。特征‘Time’包含每个交易和第一天00：00：00之间经过的秒数，‘Amount’是该笔交易的金额。特征‘Class’是预测的目标变量，在欺诈的情况下取值为1，否则为0。

**提交方式**

每个参赛队伍需要提交测试集的预测结果的csv文件，文件名为”队伍名\_预测结果.csv”，内容格式如下:

Index（序号），Pred(预测概率)，Class(结果)

1，0.92, 1

2，0.13, 0

3，0.24, 0

4，0.55, 0

5，0.12, 1

…

对于此次的信用卡交易欺诈预测，我们采用Precision-Recall Curve (AUPRC)作为评价方式。

每只队伍都需要提交相应的比赛报告，你们可以使用你们喜欢的工具（Jupyter Notebook，Microsoft office等）来生成你们的报告，但你们的报告必须能采用HTML，PDF，PPT，WORD这些通用格式。你们的代码应采用一个单独的压缩文件包提交，且与你们的预测结果文件和报告分开。你们的代码将不会被评分，但你的结果必须包含你的代码，否则不会被打分。

我们强烈建议你们采用Jupyter Notebook来进行数据的处理、探索、建模。

请尽量采用python，R作为建模工具，其他语言在决赛环境中可能无法使用。如果有什么其他的问题，请联系群里的工作人员，我们会给你们提供一些建议。