



Bootcamp: Analista de Machine Learning

Atividade Modular

Módulo: Fundamentos de Machine Learning

Objetivos de Ensino

Exercitar o seguinte conceito trabalhado no Módulo:


1. Praticar o entendimento de métricas de avaliação de modelos: acurácia, precisão e cobertura.

Enunciado


Oi, pessoal!

Nesta atividade, vamos trabalhar com um conjunto de dados de transações financeiras. Cada transação tem atributos como:

- Valor da transação;
- Data e hora da transação;
- Número do cartão de crédito de quem executou a transação;
- Gênero do dono do cartão;
- População da cidade em que a transação foi executada;
- Tipo de loja em que a transação foi realizada;
- Entre outros.



Serviços de comércio eletrônico sofrem muito com FRAUDE. Isso fica muito claro no livro [O Cientista e o Executivo](#), em que líderes de tecnologia no iFood contam como fraude é um problema importante na empresa e que só



foi domado com modelos de aprendizado de máquina que identificavam fraudes automaticamente – o alcance de regras manuais, como “transações altas tendem a ser fraude” era limitado! Antes do aprendizado de máquina, o prejuízo do iFood com fraudes chegava a milhões de reais por mês.

Acesse [este link do Kaggle](#) e baixe o arquivo ZIP que contém dois arquivos: *fraudTrain.csv* e *fraudTest.csv*. Eu preparei um [python-notebook](#) que você pode usar como base para responder às perguntas. Neste notebook, adicionei um comentário para indicar qual código te ajuda em qual questão, como “#Questão 1”, “#Questão 2” etc.

A ideia é que você trabalha para a startup *iComida* e, por conta de seu conhecimento como cientista de dados adquirido no curso da Faculdade XP, você foi contratado para tentar controlar e domar o problema de transações fraudulentas enfrentado pela companhia, por questões como roubo e clonagem de cartões, além de golpes arquitetados pelos entregadores em conluio com restaurantes.

O que você tem são dados com mais de 1,3 milhão de transações. Aquelas que foram previamente identificadas como fraude estão identificadas no atributo *is_fraud*. A empresa conta com você para estancar o buraco no orçamento que as fraudes estão causando todo mês. Boa sorte!

Prof. Pedro Calais.

Atividades

Os alunos deverão desempenhar a seguinte atividade:

1. Treinamento e avaliação de modelos preditivos para detecção de fraude.

