

Desafio Machine Learning

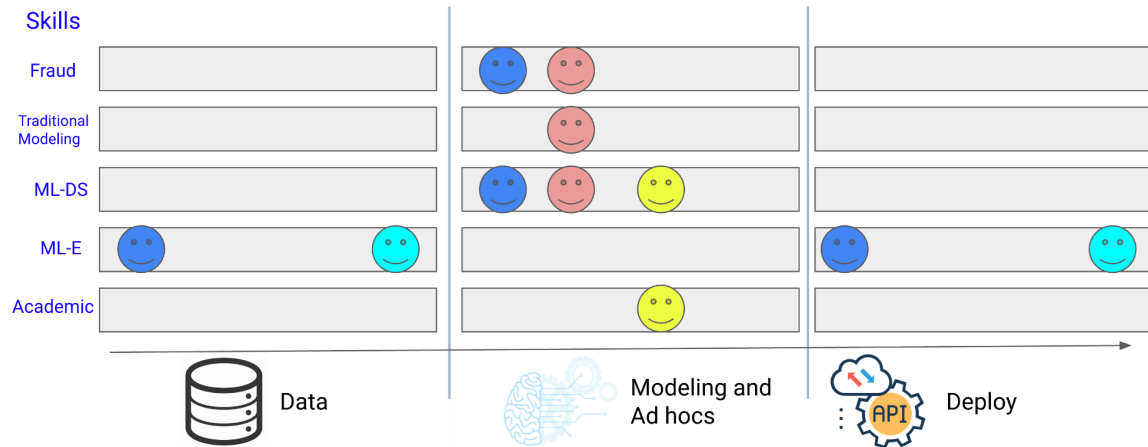
O arquivo dados.csv possui os dados para este projeto. Sobre ele pedimos que:

- 1) Desenvolva um modelo para prever a variável Fraude (1-Sim, 0-Não). Quais técnicas testou? Qual selecionou como melhor abordagem? Explique!
- 2) Compare o desempenho do seu algoritmo com o desempenho de um modelo anterior, cujo resultado é encontrado na coluna Score do dataset. Quais métricas usou e qual sua conclusão?
- 3) Como você pode garantir que o desempenho do modelo no laboratório vai ser um proxy para o desempenho do modelo em produção?
- 4) Supondo que o desempenho produtivo seja muito diferente do esperado, quais você acha que são as causas mais prováveis?
- 5) Com a informação adicional de que uma fraude custa 10 vezes mais para a empresa do que o bloqueio incorreto de um pagamento, o que você faria de diferente?
- 6) Queremos encontrar o ponto de corte ideal para o modelo já treinado apresentado na parte (Score, onde 0 (zero) é o risco mais baixo e 100 (cem) o mais alto). Sabemos que MELI ganha 10% do valor de um pagamento aprovado corretamente e a cada fraude aprovada perdemos 100% do valor do pagamento. Defina o ponto de corte que maximize os lucros do MELI, de forma que todos os pagamentos com menor Score sejam aprovados e os maiores sejam rejeitados.
- 7) Se o modelo precisar responder online, no menor tempo possível, o que isso mudaria suas decisões no item 1?
- 8) Como colocar esse modelo em produção?

Observações:

- A coluna "Fraude" é o flag para aprender.
- A coluna "Score" corresponde ao resultado do algoritmo a ser substituído.
- O formato dos números atende ao delimitador "." (ponto) para casas decimais e "," (vírgula) para milhares.

Considere que a cor define um perfil específico e os skills/atividades esperadas deste profissional.



Skills:

Fraud: Já possui conhecimento prévio de prevenção a fraude

Traditional Modeling: Tem experiência em desenvolver modelos para fraude/credito/crm utilizando metodologias tradicionais do mercado financeiro (regressão logística).

ML-DS: Machine Learning Data Scientist: Possui conhecimento das técnicas novas de modelagem como XGBoosting e outras e domina linguagem python

ML-E: Machine Learning Engineering: Possui experiência em manusear dados para construção de features stories e já participou de implantações de modelos de machine learning.

Academic: Possui pós-graduação em nível mestrado/doutorado voltado para áreas relacionadas com data science.

Atividades:

Data: Extrair e manipular dados e construção de feature stories

Modeling e Ad-hocs: Construção de modelos e estudos voltados para mitigar fraude

Deploy: Colocar modelos em produção, visando melhores práticas

Qual perfil melhor te define? Qual você gostaria de atuar no MELI? Comente!