

Projeto 3 - D&A - AWS - Ciência de Dados Edição Fast Learning

Risco de inadimplência e crédito em empréstimos.

Muitas pessoas têm dificuldade em obter empréstimos devido a históricos de crédito insuficientes ou inexistentes, e frequentemente são exploradas por credores não confiáveis.

A Home Credit busca ampliar a inclusão financeira para essas pessoas, oferecendo uma experiência de empréstimo positiva e segura. Utilizando uma variedade de dados alternativos, como informações de telecomunicações e transacionais, a empresa prevê a capacidade de pagamento dos clientes.

Atualmente a Home Credit emprega métodos estatísticos e de aprendizado de máquina para essas previsões, mas desafia os cientistas de dados a explorar todo o potencial desses dados, garantindo que clientes com capacidade de pagamento não sejam rejeitados e que os empréstimos oferecidos sejam adequados para capacitar seus clientes a terem sucesso.

Os alunos devem considerar diferentes abordagens, como:

- **Regressões Logísticas:** Avaliar a performance de uma Regressão Logística, assim como a interpretação dela.
- **Modelos Baseados em Árvores (Escolha do Aluno):** Avaliar a performance de um algoritmo baseado em Árvore, assim como a interpretação dela.
- **SVM:** Avaliar a performance de um algoritmo de Máquina de Vetores de Suporte, assim como sua interpretação

Os alunos devem realizar as seguintes tarefas para concluir o projeto:

1. **Exploração de Dados:** Analisar e visualizar os dados disponíveis para entender melhor as características da variável "Target" e como ela se relaciona com as demais informações.
2. **Pré-processamento de Dados:** Realizar limpeza, transformação e seleção de características dos dados.
3. **Desenvolvimento do Modelo:** Implementar ao menos dois algoritmos, dos recomendados, utilizando técnicas de machine learning. Escolher entre um deles e os motivos da escolha, utilizando argumentos baseados em métricas.
4. **Avaliação do Modelo:** Avaliar o desempenho do modelo utilizando métricas apropriadas, especialmente Curva Roc e Matriz de Confusão, levantando os potenciais impactos que podem ser observados baseado nessas métricas.
5. **Apresentação dos Resultados:** Preparar um notebook apresentando os resultados obtidos, incluindo uma explicação sobre o funcionamento do modelo escolhido, as principais descobertas durante a análise exploratória de dados e a avaliação do desempenho do modelo.
6. **Vídeo da Apresentação:** Grave uma apresentação de até 20 minutos sobre o seu projeto, abordando todos os pontos que julgue relevante sobre o desenvolvimento, incluindo o funcionamento, descobertas, dificuldades, avaliação e exemplo de funcionamento.

Os alunos são encorajados a serem criativos e explorar diferentes técnicas e abordagens para a construção dos modelos de ML. Os projetos serão avaliados tanto pela precisão técnica quanto pela clareza na comunicação dos resultados e dos passos desenvolvidos.

Boa sorte!

Projeto 3 - D&A - AWS - Ciência de Dados Edição Fast Learning

Informações auxiliares sobre os datasets

application_{train|test}.csv:

- Este é o arquivo principal, dividido em dois arquivos para Treinamento (com TARGET) e Teste (sem TARGET).
- Dados estáticos para todas as aplicações. Uma linha representa um empréstimo em nossa amostra de dados.

bureau.csv:

- Todos os créditos anteriores do cliente fornecidos por outras instituições financeiras que foram relatados ao Birô de Crédito (para clientes que têm um empréstimo em nossa amostra).
- Para cada empréstimo em nossa amostra, há mesma quantidade de linhas em relação a quantos créditos o cliente teve no Birô de Crédito antes da data da aplicação.

bureau_balance.csv:

- Saldo mensal de créditos anteriores no Birô de Crédito.
- Esta tabela tem uma linha para cada mês de histórico de cada crédito anterior relatado ao Birô de Crédito - ou seja, a tabela tem (# empréstimos na amostra * # de créditos anteriores relativos * # de meses em que temos algum histórico observável para os créditos anteriores) linhas.

POS_CASH_balance.csv:

- Fotos de saldo mensal de empréstimos POS (ponto de vendas) anteriores e empréstimos em dinheiro que o requerente teve com a Home Credit.
- Esta tabela tem uma linha para cada mês de histórico de cada crédito anterior na Home Credit (crédito ao consumidor e empréstimos em dinheiro) relacionado aos empréstimos em nossa amostra - ou seja, a tabela tem (# empréstimos na amostra * # de créditos anteriores relativos * # de meses em que temos algum histórico observável para os créditos anteriores) linhas.

credit_card_balance.csv:

- Fotos de saldo mensal de cartões de crédito anteriores que o requerente tem com a Home Credit.
- Esta tabela tem uma linha para cada mês de histórico de cada crédito anterior na Home Credit (crédito ao consumidor e empréstimos em dinheiro) relacionado aos empréstimos em nossa amostra - ou seja, a tabela tem (# empréstimos na amostra * # de cartões de crédito anteriores relativos * # de meses em que temos algum histórico observável para os cartões de crédito anteriores) linhas.

previous_application.csv:

- Todas as aplicações anteriores para empréstimos Home Credit de clientes que têm empréstimos em nossa amostra de dados.
- Há uma linha para cada aplicação anterior relacionada aos empréstimos em nossa amostra de dados.

installments_payments.csv:

- Histórico de pagamentos para os créditos previamente liberados na Home Credit relacionados aos empréstimos em nossa amostra de dados.
- Há uma linha para cada pagamento realizado mais uma linha para cada pagamento perdido. Uma linha é equivalente a um pagamento de uma prestação OU uma prestação correspondente a um pagamento de um crédito anterior da Home Credit relacionado aos empréstimos em nossa amostra de dados.

HomeCredit_columns_description.csv:

- Este arquivo contém descrições para as colunas nos vários arquivos de dados.

