



Graduação em Banco de Dados e Sistemas de Informação
Case 1 – Cobrança e Python

Nome dos Alunos:

Turma: A

Data:

Professor Responsável: Roberto Angelo Fernandes Santos

Observações:

- Este exercício pode ser realizado individualmente ou em dupla.



CASE CONSULTORIA DE PRIORIZAÇÃO EM COBRANÇA - DESCRIÇÃO

Empresa Contratante: CobraNada Recuperação de Ativos Ltda.

Origem da Carteira: Títulos Vencidos de um Banco.

Forma de Remuneração: Para cada título recuperado o banco paga R\$ 50,00.

Descritivo do Problema: CobraNada atualmente não utiliza nenhum tipo de ordenação baseada em Data Mining para orientar a priorização das chamadas de seus atendentes de Call Center.

Premissa: Utilizar ao menos duas técnicas para gerar o score, comparar e utilizar a melhor delas.

Sabe-se que a capacidade mensal de chamadas só permite o acionamento de 70% das dívidas da carteira disponível, ou seja, 30% dos devedores (títulos em atraso) deixam de ser acionados todo mês.

A priorização (ordenação) atual para acionamento de cobrança é baseada na ordem do arquivo de dívidas, desta forma, cobra-se os primeiros 70% de títulos do arquivo. O desempenho da CobraNada é medido pelo número total de contratos que são pagos por seus clientes.

Solução Desejada: A CobraNada gostaria de um modelo de Data Mining baseada em um **escore** de 0 a 100, que ordenasse todas as dívidas de acordo com a propensão de pagamento. Ou seja, quanto maior este escore, maior deverá ser a chance de pagamento da dívida. Assim, os 70% títulos com maior propensão de serem pagos serão acionados prioritariamente. **Os títulos para serem considerados bons devem ter sido pagos em até 120 dias de atraso.**

ETAPAS – CRISP-DM SIMPLIFICADO

Gerem uma apresentação seguindo os passos abaixo:

1. Entendimento do Negócio:

- a. Tire todas as dúvidas sobre o case com o professor;
- b. Faça um resumo para colocar na apresentação;

2. Entendimento dos Dados:

- a. Descrição rápida dos dados para colocar na apresentação;
- b. Utilize a base disponibilizada pelo professor e o dicionário para ilustrar essa fase;
- c. Coloque os itens anteriores na apresentação;

3. Pré-processamento:

- a. Tenha ao menos 10 variáveis pré-processadas para aplicar as técnicas;
- b. Para as novas colunas pré-processadas criadas, façam uma análise bivariada e univariada em uma tabela de análise simples;
- c. Justifique a razão das transformações sofridas por estas variáveis;
- d. Coloque as análises (Excel, Python...), justificativas e o código de transformação (Python) na apresentação;

4. Modelagem:

- a. Separe a base em 75% para treinamento do modelo e 25% para teste (Python);
- b. Rode diversas técnicas de aprendizado (Python);
- c. Coloque todas as evidências (códigos) desta etapa na apresentação;

5. Avaliação:

- a. Compare os erros de classificação e MSE (Python ou Excel);
- b. Selecione a técnica que possui o menor erro MSE;
- c. Estime o retorno total da recuperação calculando o retorno no conjunto de teste e extrapolando para a base total com a técnica com menor MSE, considerando que serão cobrados apenas os 70% melhores scores (Livre);
- d. Coloque todas as evidências destas avaliações na apresentação;

6. Implementação:

- a. Descreva como seria a entrega e o cálculo deste score em produção;
- b. (Opcional) Coloque o código principal de entrega do modelo na apresentação ou fórmula se possível.