|  |  |  |  |
| --- | --- | --- | --- |
|  | Graduação em Banco de Dados  Case 1 – Cobrança e Python | | |
|  | | | |
| ***Nome dos Alunos:*** | | | |
| *Turma:* **A** | |  | *Data:* |
| *Professor Responsável:* **Roberto Ângelo 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**:
   1. Tire todas as dúvidas sobre o case com o professor;
   2. Faça um resumo para colocar na apresentação;
2. **Entendimento dos Dados**:
   1. Descrição rápida dos dados para colocar na apresentação;
   2. Utilize a base disponibilizada pelo professor e o dicionário para ilustrar essa fase;
   3. Coloque os itens anteriores na apresentação;
3. **Pré-processamento**:
   1. Tenha ao menos 10 variáveis pré-processadas para aplicar as técnicas;
   2. Para as novas colunas pré-processadas criadas, façam uma análise bivariada e univarida em uma tabela de análise simples;
   3. Justifique a razão das transformações sofridas por estas variáveis;
   4. Coloque as análises (Excel, Python...), justificativas e o código de transformação (Python) na apresentação;
4. **Modelagem**:
   1. Separe a base em 75% para treinamento do modelo e 25% para teste (Python);
   2. Rode diversas técnicas de aprendizado (Python);
   3. Coloque todas as evidências (códigos) desta etapa na apresentação;
5. **Avaliação**:
   1. Compare os erros de classificação e MSE (Python ou Excel);
   2. Selecione a técnica que possui o menor erro MSE;
   3. 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);
   4. Coloque todas as evidências destas avaliações na apresentação;
6. **Entrega**:
   1. Descreva como seria a entrega e o cálculo deste escore em produção;
   2. Coloque o código principal de entrega do modelo na apresentação ou fórmula se possível.