**Etapa 3 – Teste lógico**

**1 - Você recebeu um conjunto de dados de clientes da nossa empresa, mas ele está repleto de valores ausentes e inconsistências. Como você abordaria esse problema? Descreva pelo menos três técnicas de limpeza e pré-processamento de dados que você poderia aplicar e explique quando cada uma delas seria apropriada.**

a) Atribuição de Dados Ausentes:

A atribuição pode ser usada quando, proporcionalmente ao volume total de registros, há poucos elementos faltantes e é possível estimar com base nos dados existentes.

Esta técnica é eficaz quando a técnica b afetaria negativamente o modelo com perda significativa de informações.

Exemplo: Preencher valores ausentes na idade dos clientes com a média ou mediana das idades existentes.

Este método é bastante aplicado no exercício de predição do Titanic.

b) Eliminação de Registros com Valores Ausentes:

Podemos usar a eliminação quando os valores ausentes representam uma porção significativa do conjunto de dados ou quando a presença desses valores não é crítica para a análise ou modelagem.

Exemplo: Se tivermos um campo de Endereço com muitos valores ausentes, se a análise não depender dos endereços poderemos eliminar da análise.

c) Detecção e Tratamento de Outliers:

Esta técnica é apropriada quando existem valores extremos nos dados que podem distorcer análises ou modelos.

Exemplo: Identificar e remover ou transformar outliers em dados de histórico de compras de clientes para evitar distorções nos resultados.

**2 - Nossa empresa deseja prever a taxa de churn (abandono de clientes) com base em dados históricos de uso do produto. Como você abordaria esse problema de modelagem preditiva? Descreva o processo que você seguiria, incluindo a seleção de recursos relevantes, a escolha de um algoritmo de aprendizado de máquina e a avaliação do desempenho do modelo.**

Prever a taxa de churn é crucial para empresas que desejam manter sua base de clientes. Abaixo listo um macroprocesso para identificação deste comportamento:

1. Coletar dados históricos de uso do produto, incluindo informações sobre clientes, interações, comportamentos e características.

Alguns recursos relevantes podem ser:

* Tempo de uso do produto.
* Frequência de uso.
* Histórico de compras.
* Comportamento de navegação.
* Feedback do cliente.
* Dados demográficos.

1. Pré-processamento dos Dados:

* Tratar valores ausentes, outliers e dados inconsistentes.
* Normalizar ou padronizar os dados numéricos.
* Separar os dados em conjuntos de treinamento e teste.

1. Escolha do Algoritmo para prever churn, dentre os mais comuns, listo:

* Regressão Logística
* Árvores de Decisão
* Random Forests

1. Treinamento do Modelo:

* Ajustar o modelo aos dados de treinamento.
* Otimizar hiper parâmetros (por exemplo, profundidade da árvore, taxa de aprendizado).

1. Avaliação do Desempenho do Modelo a partir de métricas como:

* Acurácia
* Precisão
* Recall
* F1-score
* Curva ROC

1. Validar o modelo com dados de teste.

*Obs.: Churn foi o tema do meu trabalho de conclusão da pós graduação em Ciência de Dados, porém o resultado foi que, com os dados disponibilizados pela empresa, não havia correlação possível entre o comportamento do cliente e churn.*

**3 - Após treinar um modelo de previsão de churn, como você avaliaria sua eficácia? Descreva pelo menos três métricas de avaliação de modelo que você consideraria relevantes para este caso específico e explique por que cada uma delas é importante.**

**Acurácia:**

É a proporção de verdadeiros positivos e verdadeiros negativos em relação ao total de previsões. É uma medida inicial útil para entender o desempenho geral do modelo.

**Curva ROC:**

Demonstra em gráfico a performance do modelo de classificação em todos os limiares de classificação. A área sob a curva ROC e mede a capacidade do modelo de distinguir entre as classes. Quanto maior a área abaixo da curva, melhor o modelo em prever corretamente.

**F1-Score:**

É a média harmônica entre precisão (proporção de verdadeiros positivos entre todas as previsões positivas) e recall (proporção de verdadeiros positivos entre todos os casos positivos reais).

**4 - Nossa empresa está considerando alterar o preço de sua assinatura mensal. Como você ajudaria a empresa a tomar essa decisão com base em dados? Quais são os fatores que você consideraria e como você os analisaria para determinar o impacto potencial dessa mudança nos negócios?**

Antes de realizar ajuste nos preços, é importante avaliar um ou mais itens da lista abaixo:

1. Analisar como mudanças anteriores de preços afetaram a aquisição e retenção de clientes.
2. Comparar os preços com os dos concorrentes e avaliar o posicionamento de mercado da empresa.
3. Analisar se o produto oferece valor suficiente para justificar um preço mais alto.
4. Segmentar clientes por comportamento de compra e sensibilidade a preços.

Pode ser interessante realizar um teste A/B e usar análise de sentimento para medir a reação dos clientes às mudanças de preço.