

## Relatório com síntese de acurácia obtida para cada modelo

### 1. Preparação dos Dados:

- **Leitura dos Dados:** Carregamos dois conjuntos de dados: as notas da pesquisa e as transcrições dos áudios de atendimento.
- **Tratamento de Dados Faltantes:** Removemos os valores nulos na coluna "Entrega Serviço".
- **Criação de Variáveis Dummy:** Transformamos a coluna Entrega Serviço em variáveis dummy para facilitar a análise.
- **Extração de Informações:** Processamos as transcrições para extrair informações importantes.
- **Combinação de Datasets:** Combinamos os datasets para criar um dataset completo.

### 2. Preenchimento de Dados Faltantes:

- **Regressão Linear e Polinomial:** Usamos regressão linear e polinomial para prever os valores faltantes na variável Entrega de Serviço.
- **Regressão Polinomial:** A regressão polinomial funcionou melhor (R-quadrado de 0.94).
- **Preenchimento dos Valores Faltantes:** Preenchemos os valores faltantes com a regressão polinomial.

### 3. Cálculo do Net Promoter Score:

- **Cálculo da Soma das Notas:** Calculamos a soma das notas para cada registro.
- **Cálculo do Percentual Médio:** Calculamos o percentual médio das notas para cada registro.
- **Categorização do NPS:** Categorizamos os clientes como Promotores, Detratores ou Neutros.

### 4. Visualização dos Resultados:

- **Gráfico de Barras:** Criamos um gráfico de barras para mostrar a distribuição das categorias NPS.
- **Conclusões:** O gráfico mostrou que o número de clientes Neutros e Promotores é semelhante, com apenas alguns Detratores.

### 5. Modelagem Preditiva:

- **Árvore de Decisão e Random Forest:** Usamos árvore de decisão e Random Forest para prever a categoria do NPS.
- **Avaliação dos Modelos:** Avaliamos a acurácia dos modelos com métrica de acurácia e matriz de confusão.

### Conclusões:

A acurácia da árvore de decisão foi 0.6 com critério Gini e para árvore de decisão com o critério sendo Log loss foi 0.7. Para Random forest o cenário também não mudou, para melhorar a performance de ambos os modelos para melhor avaliação do modelo poderíamos aumentar a quantidade e melhoria dos dados. É válido ressaltar, que é provável que obtemos esse resultado devido à baixa quantidade de features e max\_depth, tendo em vista um dataset pequeno para treinar os modelos.