**Relatório de Análise de Resultados: Previsão de Custos Médicos**

**1. Introdução**

* Objetivo do Relatório: Este relatório tem como objetivo analisar os resultados de um modelo de previsão de custos médicos, destacando os insights obtidos e validando estatisticamente o modelo.
* Descrição do Modelo: O modelo utilizado foi o XGBRegressor, que é um método de aprendizado de máquina baseado em árvores de decisão. Para preparar os dados, foi utilizado o LabelEncoder para codificar as variáveis categóricas e o StandardScaler para padronizar as variáveis numéricas.

**2. Análise dos Resultados**

* Desempenho do Modelo: 87,71% de R², $ 2365,86 de erro médio absoluto e 19079500,87 de erro médio Quadrático.

**Métricas de Avaliação:**

* R²: Indica a proporção da variabilidade dos custos médicos explicada pelo modelo.
* MAE (Mean Absolute Error): Média dos erros absolutos entre as previsões e os valores reais.
* RMSE (Root Mean Squared Error): Raiz quadrada da média dos erros quadráticos, fornecendo uma medida da precisão do modelo.

**Validação Cruzada:**XGBRegressor mostrou melhor desempenho em termos de erro médio e poder preditivo, segue então o desempenho dos outros modelos testados.

* XGBRegressor(OneHotEncoder): 87.61% $ 2459,16 e 19233794.22
* RNA -> MLPRegressor(LabelEncoder): 87,60%, $ 2490,83 e 19242616,03.
* Random Forest(LabelEncoder): 87,24%, $ 2556,20 e 19805433,17 .
* Regressão Linear(OneHotEncoder): 78,359% , $ 4181,19 e 33596915.85.
* Regressão Linear Lasso(OneHotEncoder) 78,358%, $ 4181,29 e 33597789,30.
* Regressão Linear Ridge(OneHotEncoder) 78,32%, $ 4193.19 e 33645393,49 .
* Regressão Linear(LabelEncoder): 78,33%, $ 4186,50 e 33635210,43.

**3. Insights Obtidos**

* **Relação Fumante-Custos:** Foi identificado que há uma relação de 80% entre ser fumante e os custos médicos, indicando que indivíduos fumantes tendem a ter custos significativamente mais altos.
* **Impacto das outras caracteristicas :** A idade e a IMC também mostraram-se fatores importantes, mas menos impactantes em comparação ao fator "fumante" sendo respectivamentes 30% e 20%,enquanto as outras caracteristicas como sexo, quantidade de filhos e região não chegaram a 10 %.

**4. Validação Estatística**

* Com base no p-valor de 60,78%, concluímos que o modelo não é estatisticamente significativo.
* Considerando o intervalo de confiança entre $11.673,65 e $14.357,52, e a média dos custos médicos de $13.270,42, podemos afirmar que as previsões estão bem alinhadas com os dados observados.

**5. Conclusão**

* Resumo dos Resultados: Podemos Perceber então que há uma relação entre aumentos dos custos médicos com idade, Indice de Massa Corporal e se há pessoa é fumante.