**Relatório final**

**Case – Dados**

***Autor: Gabriel Aurichio Nogueira***

Ao avaliar as métricas dos algoritmos de **Regressão Linear** e **K-Nearest Neighbors (KNN)**, podemos observar as limitações e o desempenho de ambos os modelos, com base nos resultados obtidos.

**Resultados da Regressão Linear**

O modelo de **Regressão Linear** apresentou um **Coeficiente de Determinação (R²)** de **0.01**, o que indica que ele explica apenas **1%** da variação dos ferimentos nos acidentes de carro. Esse valor baixo sugere que o modelo não consegue capturar bem as relações entre as variáveis independentes e a variável dependente (**Injury**). O **Erro Quadrático Médio (MSE)** foi de **0.25**, o que pode parecer baixo à primeira vista, mas, ao ser combinado com o R² baixo, demonstra que as previsões feitas pelo modelo não são precisas.

Além disso, os coeficientes baixos indicam que, segundo o modelo, as variáveis não possuem grande influência na previsão da variável dependente. Por exemplo, a variável **Alcohol Involved**, que tem impacto comprovado na capacidade cognitiva do condutor, não mostra aumento significativo no número de acidentes quando analisada de forma linear, o que pode ser observado nos gráficos.

Em termos de **melhoria**, a Regressão Linear não é o algoritmo ideal para esse tipo de cenário, especialmente quando as relações entre as variáveis não são lineares. A sugestão seria a utilização de **outros algoritmos de Machine Learning**, como **K-Nearest Neighbors (KNN)**, **Random Forest Classifier** ou **Logistic Regression**, que lidam melhor com dados não lineares.

**Resultados do K-Nearest Neighbors (KNN)**

Ao passar para o algoritmo de **K-Nearest Neighbors (KNN)**, a **acurácia obtida foi de 56.09%**, o que significa que o modelo acertou pouco mais da metade das previsões. Embora isso represente um desempenho razoável, ainda há **erros significativos**, especialmente quando comparamos a complexidade dos dados. No entanto, o modelo apresentou uma performance balanceada nas métricas gerais, mostrando um desempenho aceitável para as duas classes e deixando espaço para aprimoramentos, particularmente na previsão de ferimentos.

Um dos pontos principais do KNN é o parâmetro **n\_neighbors**, que influencia diretamente o desempenho do modelo. Ao aumentarmos o valor de **n\_neighbors**, observamos que a **acurácia aumentou**, mas houve um impacto negativo na **performance computacional**, já que o processamento dos dados se torna mais intensivo.

Uma melhoria significativa seria a aplicação do **KNN Ponderado**, que considera a **distância** dos vizinhos mais próximos ao fazer a predição final, atribuindo maior peso aos vizinhos mais próximos e menor peso aos vizinhos mais distantes. Isso pode ajudar a melhorar a acurácia do modelo, especialmente quando os dados apresentam padrões complexos.

**Conclusão Geral**

Conclui-se que, comparado com a Regressão Linear, o **K-Nearest Neighbors** apresentou uma **melhora significativa**, embora ainda haja margem para melhorias. O modelo KNN mostrou-se mais adequado para lidar com dados mais complexos e não lineares, mas sua performance pode ser aprimorada por meio de ajustes no valor de **n\_neighbors** e pela implementação do **KNN Ponderado**.

Em resumo, enquanto a **Regressão Linear** não conseguiu capturar a complexidade do problema, o **KNN** apresentou uma **performance razoável**, mas ainda precisa de ajustes para alcançar um desempenho ideal. A escolha do modelo adequado depende da natureza dos dados e da complexidade das relações entre as variáveis, sendo o **KNN** uma alternativa promissora em relação à Regressão Linear para este tipo de tarefa.