

## **Relatório de Aprendizado e Conclusões do Projeto de Predição dos Sobreviventes do Titanic**

### **Etapa 1: Importação dos Dados**

*Aprendizado:*

- Apreendi a carregar um dataset diretamente da biblioteca Seaborn e a usar métodos como head(), info() e describe() para ter uma visão geral da estrutura dos dados.

*Conclusão:*

- Essa etapa foi fundamental para entender o que cada coluna representa e identificar que alguns campos apresentavam valores ausentes, o que indicava a necessidade de uma limpeza posterior.
- 

### **Etapa 2: Exploração e Limpeza dos Dados**

*Aprendizado:*

- Descobri a importância de tratar dados faltantes e como usar métodos como fillna() para substituir valores ausentes com a mediana ou a moda, conforme a coluna.
- Compreendi que nem todas as colunas são úteis para a predição e que é preciso remover aquelas que podem atrapalhar o modelo (como colunas redundantes ou irrelevantes).
- Apreendi a converter dados categóricos em numéricos utilizando o LabelEncoder, preparando os dados para serem processados pelos modelos de Machine Learning.

*Conclusão:*

- A limpeza dos dados é uma etapa crítica, pois dados inconsistentes ou irrelevantes podem comprometer o desempenho dos modelos. Essa etapa garante que a base de dados esteja “limpa” e adequada para o treinamento.
- 

### **Etapa 3: Seleção de Features e Divisão dos Dados**

*Aprendizado:*

- Entendi como selecionar as variáveis que realmente influenciam na predição (features) e definir a variável alvo (target).
- Apreendi a dividir os dados em conjuntos de treino e teste utilizando o train\_test\_split(), o que é essencial para avaliar a performance do modelo em dados não vistos.

*Conclusão:*

- A escolha correta das features e a divisão adequada dos dados evitam problemas como o overfitting, garantindo que o modelo aprenda de forma generalizada e performe bem em novos dados.
- 

### **Etapa 4: Treinamento dos Modelos**

*Aprendizado:*

- Experimentei dois modelos de Machine Learning: a Regressão Logística, que é mais simples e estatística, e o Random Forest, que é um conjunto de árvores de decisão.

- Compreendi que cada modelo possui suas particularidades e pode reagir de maneira diferente aos dados.

*Conclusão:*

- Treinar modelos variados permite comparar suas performances e escolher o que melhor se adapta ao problema. Essa diversidade de abordagens enriquece a análise e a confiabilidade das previsões.
- 

## **Etapas 5: Avaliação dos Modelos**

*Aprendizado:*

- Aprendi a usar métricas como acurácia, matriz de confusão e relatório de classificação para medir o desempenho dos modelos.
- Entendi como cada métrica pode indicar pontos fortes e áreas de melhoria do modelo.

*Conclusão:*

- A avaliação detalhada dos modelos é essencial para identificar possíveis ajustes e aprimoramentos, garantindo que o modelo seja confiável e eficaz na previsão.
- 

## **Etapas 6: Visualização dos Resultados**

*Aprendizado:*

- Descobri como gráficos podem auxiliar na compreensão dos dados, demonstrando, por exemplo, quais features são mais importantes e como a sobrevivência se distribui entre diferentes grupos (sexo e classe).

*Conclusão:*

- As visualizações não só ajudam a validar as escolhas das features, mas também oferecem insights valiosos que podem guiar futuras melhorias no modelo e na análise dos dados.
- 

## **Etapas 7: Predição para Novos Passageiros**

*Aprendizado:*

- Aprendi a implementar uma função que utiliza um modelo treinado para prever a sobrevivência de novos passageiros, aplicando na prática o conhecimento adquirido.

*Conclusão:*

- Essa etapa demonstra a aplicabilidade do modelo em cenários reais, reforçando a importância de construir pipelines que possam ser usados em situações do mundo real.
-

**Conclusão Geral:**

Ao longo deste projeto, foi possível compreender e vivenciar cada fase do pipeline de Machine Learning. Desde a importação e limpeza dos dados até a avaliação dos modelos e a visualização dos resultados, cada etapa contribuiu para desenvolver um modelo robusto e aplicável a um problema real.

O que mais me marcou foi perceber como a qualidade dos dados e a escolha das features influenciam diretamente o desempenho dos modelos. Essa experiência reforçou a importância de um bom pré-processamento e da análise cuidadosa dos resultados, que juntos formam a base de uma solução de Machine Learning bem-sucedida.