Relatório: Previsão de Chuvas na Austrália

# Relatório: Previsão de Chuvas na Austrália  
  
## Introdução  
O objetivo deste projeto é desenvolver um modelo preditivo para prever se choverá no dia seguinte em diferentes regiões da Austrália. O problema é formulado como uma tarefa de classificação binária, onde a variável-alvo é \*\*RainTomorrow\*\*.  
  
### Motivação  
- Identificar padrões climáticos e entender os fatores que influenciam as chuvas.  
- Construir um modelo que auxilie na tomada de decisão em setores como agricultura e turismo.  
- Estudo prático de técnicas de aprendizado de máquina aplicadas a problemas reais.  
  
---  
  
## Etapas do Projeto  
  
### 1. Preparação do Ambiente  
#### O que foi feito:  
- As bibliotecas principais para análise e modelagem de dados foram importadas: `numpy`, `pandas`, `matplotlib`, `seaborn`, `sweetviz`, `scikit-learn`, entre outras.  
- O dataset utilizado, `weatherAUS.csv`, foi carregado e configurado no ambiente de trabalho.  
  
#### Justificativa:  
Essas bibliotecas fornecem ferramentas robustas para análise de dados, visualização e modelagem, garantindo eficiência no desenvolvimento do projeto.  
  
---  
  
### 2. Análise Exploratória de Dados (EDA)  
#### O que foi feito:  
- Leitura inicial dos dados para identificar informações básicas como número de instâncias, tipos de variáveis e presença de valores ausentes.  
- Exploração visual com gráficos (`seaborn`) para identificar relações entre variáveis como temperatura, umidade, e ocorrência de chuva.  
- Relatório automatizado gerado com `sweetviz` para detectar padrões e insights rapidamente.  
  
#### Justificativa:  
- A análise inicial é essencial para compreender os dados e detectar problemas que possam impactar a modelagem, como valores ausentes e outliers.  
- Ferramentas visuais ajudam a identificar correlações e tendências.  
  
---  
  
### 3. Pré-processamento dos Dados  
#### O que foi feito:  
- \*\*Tratamento de valores ausentes:\*\* Remoção ou preenchimento de valores nulos com base em critérios como média ou moda.  
- \*\*Codificação de variáveis categóricas:\*\* Uso de `LabelEncoder` para converter categorias em valores numéricos.  
- \*\*Padronização dos dados:\*\* Aplicação de `StandardScaler` para normalizar variáveis contínuas, garantindo que todas tenham a mesma escala.  
  
#### Justificativa:  
- Variáveis categóricas precisam ser transformadas para serem interpretadas pelos algoritmos de machine learning.  
- Padronização melhora o desempenho de algoritmos que são sensíveis à escala dos dados, como Regressão Logística.  
  
---  
  
### 4. Modelagem  
#### O que foi feito:  
- \*\*Modelo escolhido:\*\* Regressão Logística como ponto de partida.  
- \*\*Divisão dos dados:\*\* Separação em conjuntos de treino (80%) e teste (20%) para avaliar a performance.  
- \*\*Treinamento do modelo:\*\* Ajuste dos parâmetros do modelo nos dados de treino.  
- \*\*Avaliação do modelo:\*\* Métricas como acurácia, matriz de confusão e relatório de classificação foram usadas para medir o desempenho.  
  
#### Justificativa:  
- A Regressão Logística é um modelo simples e eficaz para tarefas de classificação binária, fornecendo um baseline confiável.  
- A separação dos dados evita overfitting e permite testar a generalização do modelo.  
  
---  
  
### 5. Resultados e Avaliação  
#### O que foi feito:  
- \*\*Métricas de desempenho:\*\*  
 - Acurácia do modelo: \*inserir valor do notebook\*.  
 - Matriz de confusão para visualizar falsos positivos e negativos.  
 - Relatório de classificação para medir precisão, recall e F1-score.  
  
#### Justificativa:  
- Métricas variadas fornecem uma visão completa sobre o desempenho do modelo, permitindo identificar suas limitações e pontos fortes.  
  
---  
  
## Conclusão  
O projeto demonstrou que é possível prever a ocorrência de chuvas na Austrália com um nível razoável de precisão usando Regressão Logística.   
  
### Possíveis Melhorias:  
- Experimentar algoritmos mais complexos, como Random Forest ou Gradient Boosting.  
- Realizar tuning de hiperparâmetros para melhorar a performance do modelo.  
- Incorporar mais variáveis relevantes ao problema.  
  
---  
  
## Autores  
- Eduardo Selber  
- Henrique Badin  
- Luca Caruso