| **Programa Residência em TIC 20 Capacita Brasil / C-JOVEM** | | |
| --- | --- | --- |
| Nome do curso: CIÊNCIAS DE DADOS | Aula 01 – Introdução à Ciência de Dados e Inteligência Artificial | Responsável: Prof.Bonfim Amaro Junior |

| **Plano de Aula** |
| --- |

| **Objetivo da Aula**: Apresentar aos(as) alunos(as) o ciclo básico de um projeto de Ciência de Dados, abordando desde a coleta e preparação dos dados até a modelagem e avaliação de um modelo preditivo usando Python no Jupyter Notebook. |
| --- |
| **Duração Total: 80-90 minutos** |
| **0. Introdução, Contextualização e Ambiente de Desenvolvimento (~15 minutos)**   * **Objetivo:** Explicar a importância da Ciência de Dados e IA no cenário atual e introduzir o problema que será abordado durante a aula.   + Conteúdo: Discutir brevemente como os dados estão sendo usados para tomar decisões em várias indústrias (e-commerce, saúde, educação, etc.). Explicar o problema do *Iris Dataset* — prever a espécie de uma flor com base em suas características.   + Abordar a ferramenta Jupyter/google collab para desenvolvimento prático. |
| **Etapa 1: Coleta e Visualização Inicial dos Dados (~10 minutos)**   * **Objetivo:** Mostrar como coletar e entender dados utilizando bibliotecas em Python como Pandas (Somente o básico).   + Conteúdo:     - Apresentar o *Iris Dataset* e mostrar como carregar dados usando o load\_iris() do sklearn.     - Mostrar a estrutura do DataFrame usando head() e discutir as características e o rótulo do problema. (SUPERVISIONADO)   + Demonstração no Notebook: Carregar os dados e exibir as primeiras linhas do DataFrame. (Pode usar outras funções para apresentação) |
| **Etapa 2: Preparação e Limpeza dos Dados (~5 minutos)**   * **Objetivo:** Explicar como preparar os dados para a modelagem, discutindo verificação de dados ausentes e normalização.   + Conteúdo:     - Verificar se há dados ausentes no dataset e confirmar que os dados estão prontos para análise.     - Explicar a importância de normalizar, remover valores faltantes e garantir a consistência dos dados.     - Mostrar o uso de funções como isnull() e describe() para explorar os dados.   + Demonstração no Notebook: Verificação de dados ausentes e geração de um resumo estatístico com describe(). |
| **Etapa 3: Análise Exploratória dos Dados (~20 minutos)**   * **Objetivo:** Ensinar como usar gráficos simples para entender a relação entre as variáveis do dataset. (MOTIVAÇÃO)   + Conteúdo:     - Visualizar a distribuição das espécies de flores com *countplot*.     - Explicar a importância de entender a relação entre características numéricas, como sépalas e pétalas, usando gráficos como o scatterplot e boxplot.     - Interpretar os gráficos e discutir como características específicas podem influenciar na classificação das espécies.   + Demonstração no Notebook: Gerar gráficos como *countplot* e *scatterplot* para comparar as espécies. |
| **Etapa 4: Modelagem dos Dados (~15 minutos)**   * **Objetivo:** Introduzir a criação de modelos de *Machine Learning* (IA) com Python, focando na modelagem preditiva.   + Conteúdo:     - Explicar o que são modelos de aprendizado supervisionado e como o Random Forest pode ser usado para resolver o problema de classificação.     - Dividir os dados em conjuntos de treino e teste usando train\_test\_split.     - Treinar o modelo de Random Forest e realizar previsões com o conjunto de teste.   + Demonstração no Notebook: Treinamento e avaliação do modelo com *RandomForestClassifier* e cálculo da acurácia. |
| **Etapa 5: Avaliação e Validação do Modelo (~10 minutos)**   * **Objetivo:** Avaliar o desempenho do modelo usando métricas de classificação.   + Conteúdo:     - Mostrar como calcular a acurácia do modelo usando accuracy\_score.     - Explicar o que são métricas como **precisão**, **recall** e **F1-score**, e gerar um relatório de classificação usando classification\_report.   + Demonstração no Notebook: Avaliar o desempenho do modelo e discutir os resultados. |
| **Etapa 6: Conclusão e Discussão (~10 minutos)**   * **Objetivo:** Revisar o que foi aprendido e discutir possíveis melhorias no processo.   + Conteúdo:     - Revisar cada etapa do ciclo de Ciência de Dados: coleta, limpeza, análise exploratória, modelagem e avaliação.     - Abrir para perguntas e discutir como melhorar o modelo ou aplicar o processo em outros datasets. |
| **Obs Final:** É importante que o(a) professor(a) atribua para cada etapa da aula **a** constante utilização da lógica de programação, bem como a modularização do algoritmo. Pois, a aula 02 tem como foco algoritmos e programação em Python. |