Modus Operandi do Projeto

(palavras separadas por barras são sinônimas)

1 - Selecionar base de dados corporativa ou pública, a qual possua atributos diversos e um atributo especial que possa ser usado como Label/Classe/Rótulo/Alvo (exemplo: cliente adimplente ou inadimplente);

2 – Apresentar um relatório descritivo da base, usando a biblioteca PANDAS. Também estará liberado apresentar este relatório em Power BI ou Apache Superset ou Kibana; Informações mínimas descritivas: Valor máximo, valor mínimo, média, mediana, outliers e totalizações;

3 – Selecionar dois algoritmos de classificação (Se for baseado na literatura, melhor ainda. Por exemplo, podem existir artigos que evidenciam que determinado algoritmo funciona melhor para a base escolhida);

4 – Fazer experimento, no Python, com cross-validation de 10, para os dois algoritmos, apresentando: Acurácia, Precisão, Sensibilidade/Cobertura Positiva/Recall e Medida-F1/F1-Measure/F-Measure.

5 – Com o algoritmo vencedor (baseie-se na Medida-F1), demonstre, no Python, uma chamada ao algoritmo e a classificação que ele faria (Um oráculo).

6 – Use sua criatividade e apresente um relatório da classificação numa ferramenta de BI. O uso de templates está liberado.