

 		Escola SENAI “Armando de Arruda Pereira” UFP - 1.23 Situação Problema		Aprovado por:	NOTA:
Componente Curricular: Inteligência Artificial & Big Data					
Alunos:			Turma:	Número:	
Professor: Miguel Bozer da Silva				Data:	
Curso: Tecnólogo em Análise e Desenvolvimento de Sistemas			Duração:		

Descrição da Situação Problema

Nesta avaliação, cada aluno receberá **um dos cinco conjuntos de dados industriais** preparados pelo professor. Cada conjunto possui **7 colunas de entrada (atributos)** e **1 coluna de saída (classe)**, que deve ser prevista por meio de técnicas de **aprendizado de máquina**.

1. Contexto dos conjuntos de dados

Os dados simulam cenários comuns em ambientes industriais e estão organizados da seguinte forma:

- **Produção de Máquinas** ([dados_producao.csv](#))
 - **Coluna de saída:** [falha](#) (valores possíveis: "[falha](#)" ou "[sem falha](#)")
- **Qualidade de Produto** ([dados_qualidade.csv](#))
 - **Coluna de saída:** [qualidade](#) (valores possíveis: "[aprovado](#)" ou "[reprovado](#)")
- **Manutenção Preditiva** ([dados_manutencao.csv](#))
 - **Coluna de saída:** [manutencao](#) (valores possíveis: "[precisa](#)" ou "[não precisa](#)")
- **Consumo de Energia em Fábrica** ([dados_energia.csv](#))
 - **Coluna de saída:** [eficiencia](#) (valores possíveis: "[eficiente](#)" ou "[ineficiente](#)")
- **Segurança Operacional** ([dados_seguranca.csv](#))
 - **Coluna de saída:** [risco](#) (valores possíveis: "[alto](#)" ou "[baixo](#)")

2. Tarefas a serem realizadas

- **Preparação dos dados**
 - Tratar valores presentes nas colunas de entrada para possibilitar o treinamento de Machine Learning
 - Converter colunas **categóricas em texto** em variáveis numéricas utilizando **One-Hot Encoding e/ou Label Encoding**.
 - Normalizar ou padronizar variáveis numéricas, caso necessário.
 - **Desáfio:** Tratar os valores nulos substituindo-os pela mediana (colunas contínuas) ou moda (coluna categórica)
- **Treinamento de modelos de Machine Learning**
 - Utilizar pelo menos **5 algoritmos diferentes** vistos em aula. Exemplos: Regressão Logística, Árvore de Decisão, Random Forest, KNN, SVM, Naive Bayes, entre outros.
- **Avaliação dos modelos**
 - Dividir os dados em **conjunto de treino e teste**.
 - Treinar os modelos no conjunto de treino e avaliar no conjunto de teste.
 - Obter as matrizes de confusão e comparar a **acurácia** de cada modelo para informar o melhor entre eles;
 - Plotar a curva ROC obtenha o ROC-AUC para verificar o resultado dos modelos treinados
- **Seleção e Salvamento do Melhor Modelo**
 - O modelo com maior acurácia deverá ser salvo em arquivo (**.pkl**) utilizando **joblib** ou **pickle**.
 - O **pré-processamento de escalonamento dos dados** (ex.: **StandardScaler** ou **MinMaxScaler**) também deve ser salvo.
- **Previsão com Dados de Entrada via **input()** do python**

2. Requisitos da Entrega

- Realizar a análise exploratória e o tratamento dos dados;
- Treinar e comparar ao menos 5 modelos diferentes de Machine Learning;
- Comparar os modelos com as diferentes métricas de desempenho.

Os itens avaliados estão na tabela a seguir. **ORGANIZAR O SEU RELATÓRIO DE ACORDO COM ESSA TABELA**

Item	Critério	Pontos	Realizou?
Exploração dos Dados			
1	Verificação de valores nulos	5	
2	Geração de estatísticas descritivas (média, desvio, mínimo/máximo).	5	
3	Visualizações simples dos dados com gráficos	5	
Tratamento dos dados			
4	Converter textos em números – Label Encoding / One Hot Encoding e remover alguma coluna que PODE ser desnecessária	5	
5A*	Remover as linhas com os valores nulos	5	
5B*	Substituir os valores nulos pela mediana ou moda	20	
6	Normalização ou Padronização + Justificativa da escolha entre as duas opções	5	
Treinamento dos modelos e métricas de desempenho			
7	Treinamento de 5 modelos diferentes de Machine Learning para problema.	5	
8	Obter as métricas de desempenho (matriz de confusão, precisão, revocação, f1-score, acurácia) para todos os modelos treinados.	5	
9	Elaboração da curva ROC e obtenção do ROC-AUC. Avaliar o desempenho do seu modelo a partir dessa métrica	10	
Conclusões da atividade e Qualidade do trabalho			
10	Conclusão final da atividade (Os modelos obtidos poderiam ajudar a empresa em que sentido? O que seria pior para o seu modelo, os falsos negativos ou falsos positivos?).	10	
11	Organização do relatório entregue pelo grupo	10	
Utilização do modelo salvo			
12	Salvar o pré-processamento e o melhor modelo e fazer uma previsão recebendo os dados via input()	15	
TOTAL		85/100	

* Escolher ou a opção A ou B.