

Aula 13 - Assíncrona - Atividade Prática 02

Análise de Desempenho do Modelo Através de Engenharia de *Features*

Esta atividade corresponde à nossa segunda atividade prática **valorizada** do curso de *Machine Learning*.

Instruções:

- 1 - Reunam-se em grupos de **no máximo 3 pessoas**;
- 2 - A entrega deve ser realizada via CANVAS através de um artigo no formato PDF gerado a partir do *notebook*. Para tal fim, os grupos poderão utilizar o *Jupyter Notebook* ou o *Google Drive* para implementação do que foi solicitado.
- 3 - O período de disponibilidade para recebimento da atividade será de **11/09/23 às 17h, até 21/09/23, às 17h**.
- 4 - No cabeçalho do documento devem constar o **nome completo e o RA de TODOS os membros do grupo**.

Práticas a serem desenvolvidas

1. Leitura de *dataset* presente em diretório do computador
2. Divisão de dados em grupos de treinamento e teste
3. Criação e aplicação de modelo de *Machine Learning*
4. Seleção de *features* por critério de limiar de variância
5. Extração de *features* por Análise de Componentes Principais (PCA)

Questão 1) Nesta atividade, utilize o *k-Nearest Neighbor* para verificar o desempenho do algoritmo de acordo com a engenharia de features empregada. Para tal fim, siga os seguintes passos:

1. Utilize o *dataset ionosphere* presente na pasta *datasets* para realizar uma tarefa de **classificação binária**.
2. Separe os dados de entrada e saída considerando que a coluna de índice 34 corresponde ao *label (target)* dos dados de cada linha.
3. Separe os dados em dados de treinamento e teste utilizando a proporção 70% e 30%, respectivamente, e `random state = 0`.
4. Estabeleça arbitrariamente o número de vizinhos $k = 5$, inicialmente.
5. Realize a classificação e obtenha os seguintes resultados:
 - a) Acurácia
 - b) Precisão
 - c) *Recall*
 - d) F1-Score
 - e) Plote a Matriz de Confusão
1. Repita todo o processo dos itens 1 até 5 utilizando o método de **Seleção de Features por limiar de variância**, considerando o limiar de 35%.
2. Repita todo o processo dos itens 1 até 5 utilizando o método de **Extração de Features por Análise de Componentes Principais (PCA)**, considerando a porcentagem de informação a ser mantida no prolema de 85%.

Questão 2) Compare os resultados obtidos em todos os processos do item anterior e justifique as diferenças obtidas através dos conceitos de engenharia de *features* estudados durante as aulas síncronas.