



Orientações para Aula EAD e Atividade em Grupo

Classificação

Regressão Logística e Knn

Esta aula é composta dos materiais a serem assistidos e estudados:

- a) Vídeo com Objetivo da Aula e Orientações: assistir ao vídeo da aula contendo o objetivo da aula e as orientações.
- b) IA_EAD_avaliacaoDosModelosPreditivos_20251.pdf: material para estudo contendo a Avaliação de Modelos Preditivos.
- c) IA_modulosPythonAnaliseDeDados_Partellell.pdf: material complementar a respeito de módulos do Python para a análise de dados, como: matplotlib, seaborn, numpy e pandas.
- d) IA_modulosPythonAnaliseDeDados_Partelll_sciKitLearn.pdf: material complementar a respeito do módulo scikit-learn contendo os modelos trabalhados na aula, regressão logística e KNN.

Tendo por base a leitura e navegação dos materiais das aulas de classificação envolvendo as técnicas: Regressão Logística e Knn e os itens anexo a atividade:

- i) IA_problemaNegocioEPreparacao_COVID19_20242.pdf: apresentação do problema de negócio da atividade, ou seja, "prever se uma pessoa pode vir a óbito ou não por COVID-19";
- ii) articleClassification_COVID19.pdf: artigo Yavuz e Dudak (2020) contendo detalhes sobre o problema de negócio.
- iii) df_covid.csv: *dataset* utilizado pelo artigo Yavuz e Dudak (2020).
- iv) df_covid_preparados.csv: *dataset* df_covid.csv modificado, conforme descrito no PDF do item 'i' acima.
- v) df_covid_preparados_normalizados.csv: *dataset* df_covid_preparados.csv com os dados normalizados.
- vi) df_covid_preparados_normalizados_semAtrCorr.csv: *dataset a ser gerado pelo grupo* contendo os atributos normalizados e sem os atributos com alta correlação positiva/negativa.
- vii) IA_EAD_COVID19_AprendizadoOtimizacao_Classificacao_KNN_RegressaoLogistica.ipynb: Notebook contendo a atividade a ser realizada pelo grupo.

A atividade a ser realizada no Notebook deve ser:

- 1 Executar a análise exploratória que consta no notebook no *dataset* "df_covid_preparados.csv".
- 2 Executar a normalização do atributo idade e gerar o *dataset* "df_covid_preparados_normalizados.csv".
- 3 A partir da correlação obtida com o mapa de calor no *dataset* "df_covid_preparados_normalizados.csv", gerar um novo *dataset* contendo os atributos que não possuem alta correlação positiva/negativa e que terá o nome df_covid_preparados_normalizados_semAtrCorr.csv.
- 4 Realizar a execução do último dataset (df_covid_preparados_normalizados_semAtrCorr.csv) pelos algoritmos Knn e Regressão Logística, estudados na teoria, com o atributo idade normalizado com e sem os atributos de alta correlação positiva/negativa



UNIVERSIDADE PRESBITERIANA MACKENZIE

Faculdade de Computação e Informática

Prof. Dr. Ivan Carlos Alcântara de Oliveira

Inteligência Artificial – 7ºN CC - Noite



- 5 Escolher algum dos algoritmos executados com as melhores métricas (precisão, acurácia, f1-score, recall, matriz de confusão), ou seja, cujos resultados são mais satisfatórios (explique o motivo da escolha) e o *dataset* correspondente e fazer os ajustes dos seus hiperparâmetros na otimização.
- 6 Salvar o modelo preditivo gerado.

Ao finalizar, encaminhar os *datasets* utilizados e gerados, o modelo salvo e o notebook “IA_EAD_COVID19_AprendizadoOtimizacao_Classificacao_KNN_RegressaoLogistica.ipynb” contendo a execução do solicitado via Ambiente Virtual e o nome dos integrantes do grupo até a data estipulada.

Referências

ÜNAL, Yavuz; DUDAK, Muhammed Nuri. Classification of covid-19 dataset with some machine learning methods. **Journal of Amasya university the institute of sciences and technology**, v. 1, n. 1, p. 30-37, 2020.