



Orientações para Aula EAD

Estudo e Atividade em Grupo

Aprendizado de Máquina Supervisionado

Classificação: Estudo dos algoritmos Naïve Bayes, SVM, Random Forest e Adaboost, Experimento e Atividade

Os objetivos e orientações da aula são encontrados no vídeo desta aula EAD.

Em resumo, esta aula EAD contempla o estudo/experimentos individual a respeito de mais algoritmos de classificação e a realização de uma atividade envolvendo todos os algoritmos de Aprendizado de Máquina estudados. Ou seja, nesta aula, você deve:

- 1) Realizar o estudo e testes dos materiais abaixo contidos em “EADestudoEExperimentos.zip”:
 - a. IA_aprendizadoCompletoClassificacao_20251_vfinal_resumida.pdf: estudar a respeito dos algoritmos de classificação clássicos: Naïve Bayes e Máquinas de Vetores de Suporte (*Support Vector Machines* - SVMs), além dos algoritmos do tipo Ensemble Random Forest e Adaboost. Ao final desse material também são encontrados “Fatores a Considerar para a Escolha do Algoritmo Ideal” de Aprendizado de Máquina.
 - b. IA_AprendizadoClass_Todos_v20251.ipynb: Notebook a respeito da Detecção de Câncer de Mama envolvendo o dataset “df_cancer_mama_preparados.csv” (atributos: bi_rads, age, shapes, margin, density e severity - preditor) e contendo: uma breve Análise Exploratória; o estudo de técnicas de Seleção de Atributos, tais como: seleção univariada, eliminação recursiva de atributos e métodos ensemble para seleção de variáveis, com isso obtendo o dataset “dfdadosMama_atrib_selec.csv” com os atributos selecionados (age, shapes, margin e severity - preditor); a aplicação e avaliação de Vários Algoritmos de Classificação, sendo eles: KNN, Árvore de Decisão, Regressão Logística, Naïve Bayes, SVMs, Random Forest e Adaboost; a Otimização do Modelo - Ajuste de Hyperparâmetros com Grid Search Parameter Tuning e Random Search Parameter Tuning; e, por fim, a partir do algoritmo com a otimização dos hiperparâmetros, o modelo preditivo (modelo_rf_class_cancermama.sav) é salvo.
- 2) Fazer a atividade descrita abaixo e com materiais encontrados no arquivo “EADAtividade.zip”:
 - a. A atividade a ser realizada nesta aula EAD deve ser executar o *notebook* “IA_EAD_Projeto_COVID19_SelecaoAprendizadoOtimizacao_Classificacao_v20251_final.ipynb” entregue e realizar as ALTERAÇÕES internas solicitadas. Uma síntese das atividades a fazer seria: realizar a seleção de atributos de df_covid_preparados.csv pelo método ensemble para seleção de variáveis gerando o dataset “dfcovid_atrib_selec.csv”; separar 10 dados para validação a partir do *dataset* com os atributos selecionados e gravar no dataset “dfcovid_atrib_selec_valida.csv”; fazer os experimentos com os sete métodos de classificação estudados (se necessário reestude-os); avaliar e selecionar o melhor modelo; otimizar o melhor modelo, segundo alguns dos seus hiperparâmetros; gerar e salvar o melhor modelo otimizado; e realizar testes de validação com os 10 dados (separados) não utilizados na etapa de treino e testes e encontrados em “dfcovid_atrib_selec_valida.csv”.
 - b. A data de entrega está configurada na área da nossa disciplina correspondente a atividade dentro do Ambiente Virtual.
 - c. Devem ser entregues: o *notebook* com a sua execução completa e o nome dos integrantes do grupo; o *dataset* com os atributos selecionados sem os 10 dados de



UNIVERSIDADE PRESBITERIANA MACKENZIE
Faculdade de Computação e Informática

Prof. Dr. Ivan Carlos Alcântara de Oliveira

Inteligência Artificial – 7ºN CC - Noite

validação; o *dataset* contendo somente os 10 dados para validação; e o modelo preditivo gerado.



Obs.: Considerando a densidade desta atividade, ela deverá ser realizada em duas semanas e equivale a 2 atividades EAD.

REFERÊNCIAS

COVID-19. *Mexico Patient Health Dataset*. (2020, 05 19). Disponível em Kaggle.com: <https://www.kaggle.com/riteshahlawat/covid19-mexico-patient-health-dataset> . Data da consulta: 15 de Março de 2025.

YAVUZ, Ü. N. A. L.; 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. Disponível em: <https://dergipark.org.tr/en/pub/jauist/issue/55760/748667>. Data da Consulta: 15 de Março de 2025.