

CET0621 – Aprendizado de Máquina na Análise de Dados

Projeto Final: Proposta (versão 01 - inicial)

Título: Aplicação de técnicas de classificação de dados (Aprendizado de máquina) para prever mortes por doenças cardíacas.

Autores:

Fabio Andrade
Gismar Barbosa
João Amazonas
Laís Gregório
Thainnara Lima

RA:

ex182485
ex182527
ex182429
ex182382
ex182385

Descrição do problema: iremos utilizar, como base, o dataset do trabalho desenvolvido na disciplina “Mineração de Dados” (CET0611). Na disciplina anterior aplicamos as técnicas de KDD e agora, nossa ideia é, aplicar estratégias de aprendizado de máquina, utilizando tarefas de classificação para predição.

Objetivos do trabalho:

- Entender o domínio do dataset (médio/doenças/cardíacas);
- Aplicar técnicas, necessárias, de KDD;
- Aplicar métodos supervisionados de aprendizado de máquina (classificação dos dados);
- Analisar resultados sobre classificação das mortes por doença cardíaca (nível de eficiência do classificado e desafios enfrentados).

Justificativa para sua realização: levando em consideração (como demonstram algumas referências abaixo), doenças cardíacas ainda são as principais causas de morte no mundo. Desenvolver trabalhos que possam, a partir de um conjunto de dados, classificar os eventos de possíveis mortes, vai de encontro com o objetivo de atuar de forma preventiva (utilizando métodos

preditivos) para minimizar o problema em si. Uma vez que os dados já encontram ou, possam ser catalogados, faz sentido aplicar técnicas de aprendizado de máquina para a classificação destes dados – gerando informação e conhecimento sobre o assunto.

Observações:

- Inicialmente, os dados e planejamento são esses, mas, como estamos trabalhando e um processo de aprendizado contínuo (ao decorrer da matéria), possíveis alterações no escopo principal poderão ocorrer. Sempre que isso acontecer, serão consultas, com antecedências, ao professor;
- O mesmo poderá ocorrer com outros pontos do trabalho, como as referências utilizadas (inicialmente, estamos trabalhando com as da matéria anterior).

Referências:

Bourbon, Mafalda, et al. Doenças Cardiovasculares. fevereiro de 2016, p. 1–24.

Chicco, Davide, e Giuseppe Jurman. Heart Failure Clinical Records. UCI Machine Learning Repository, 2020. DOI.org (Datacite), <https://doi.org/10.24432/C5Z89R>.

Chicco, Davide, e Giuseppe Jurman. “Machine Learning Can Predict Survival of Patients with Heart Failure from Serum Creatinine and Ejection Fraction Alone”. BMC Medical Informatics and Decision Making, vol. 20, n°1, dezembro de 2020, p. 16. DOI.org (Crossref), <https://doi.org/10.1186/s12911-020-1023-5>.

“GitHub: Let’s Build from Here”. GitHub, 2024, <https://github.com/>.

Khanna, Nilima. “J48 Classification (C4.5 Algorithm) in a Nutshell”. Medium, 18 de agosto de 2021, <https://medium.com/@nilimakhanna1/j48-classification-c4-5-algorithm-in-a-nutshell-24c50d20658e>.

Marcela. “Classificação de dados: Abordagens e modelos em mineração de dados”. Awari, 31 de julho de 2023, <https://awari.com.br/classificacao-de-dados-abordagense-modelos-em-mineracao-de-dados/>.

OPAS/OMS. “Doenças cardiovasculares - OPAS/OMS | Organização Pan-Americana da Saúde”. OPAS - Organização Pan-Americana de Saúde, 31 de janeiro de 2024, <https://www.paho.org/pt/topicos/doencas-cardiovasculares>.

“Software de planilha online gratuito: Excel | Microsoft 365”. Microsoft Excel, <https://www.microsoft.com/pt-br/microsoft-365/excel>. Acesso em 5 de junho de 2024.

Souza, Alex. “Knowledge Discovery in Databases (KDD)”. Blog Do Zouza, 26 de julho de 2023, <https://medium.com/blog-do-zouza/knowledge-discovery-in-databases-kdd462ea2775715>.

“Weka 3 - Data Mining with Open Source Machine Learning Software in Java”. WEKA The Workbench for Machine Learning, <https://waikato.github.io/wekasite/index.html>. Acesso em 31 de maio de 2024.