**Uso de Aprendizado de Máquina para Predição de Mortes por Doença Cardíaca**

**Alunos**

* Fábio Andrade
* Gismar Barbosa
* João Amazonas
* Laís Gregório
* Thainnara Lima

**Resumo**

<Escrever no final>

**Introdução**

Doenças cardiovasculares afetam o sistema circulatório e são a principal causa de mortalidade global, com 17,9 milhões de mortes em 2016 (OPAS/OMS, 2024). Essas doenças, causadas principalmente por placas nas artérias, podem ser prevenidas por meio do controle de fatores de risco comportamentais, como tabagismo, dieta inadequada e sedentarismo (Bourbon et al., 2016).

Levando em consideração que o aprendizado de máquina contribui para otimizar recursos na saúde, priorizando atendimentos críticos e reduzindo custos, com base no dataset de Chicco e Jurman (2020), contendo os dados de 299 pacientes com 12 características clínicas, este artigo, aplicando técnicas de KDD (Souza, 2023) e, ferramentas de aprendizado de máquina (ML), tem como objetivo principal, analisar a eficiência, dos métodos supervisionados de ML (classificação de dados) para predizer morte por insuficiência cardíaca a partir dos dados clínicos deste dataset. Essa abordagem, em uma aplicação do mudo real, poderia aprimorar a detecção precoce e, gestão das doenças cardiovasculares, sendo uma ferramenta estratégica para diminuição da mortalidade.

**Metodologia**

*Dados*

Os dados, originais (dataset), podem ser visualizados a partir a partir do repositório “Heart Failure Clinical Records” (2020), sendo possível baixá-los (formato de arquivo CSV) na url “https://archive.ics.uci.edu/static/public/519/heart+failure+clinical+records.zip”. Os dados são rotulados pelos seguintes atributos (ver em tabela 1):

|  |  |  |  |
| --- | --- | --- | --- |
| Nome do Atributo | Tipo | Descrição | Unidade de Medida |
| age | Inteiro | idade do paciente | Anos |
| anaemia | Binário | diminuição de glóbulos vermelhos ou hemoglobina |  |
| creatinine\_phosphokinase | Inteiro | nível da enzina CPK no sangue | mcg/L |
| diabetes | Binário | se o paciente tem diabetes |  |
| ejection\_fraction | Inteiro | porcentagem de sangue saindo do coração a cada contração | % |
| high\_blood\_pressure | Binário | se o paciente tem hipertensão |  |
| platelets | Continuo | plaquetas no sangue | kiloplateIets/mL |
| serum creatinine | Continuo | nível de sódio no sangue | rnEqL |
| sex | Binário | mulher ou homem |  |
| smoking | Binário | se o paciente fuma ou não |  |
| time | Inteiro | período de acompanhamento do paciente | dias |
| death event | Binário | se o paciente faleceu durante o período de acompanhamento |  |

Esses dados, para facilitar acesso, assim como a reprodução e/ou aplicação deste processo metodológico, tais dados se encontram disponível, para consulta, no repositório (link do repositório) (referência do repositório).

*Weka*

<Explanar, 3 linhas sobre a ferramenta, com refeências>

‘’’’’’’’””””””’’’’’’’’’’’’’’’’””””””””””áááááááááááááááLaís

*KDD*

Este projeto está sendo construído como continuidade do projeto da disciplina de Mineração de Dados (CET-0611), onde os processos que envolvem o KDD foram explorados e aplicados ao conjunto de dados (dataset), visando a mineração de dados. Portanto, neste trabalho, como o foco são os processos de classificação de (predição) de dados, não iremos replicar os passos do projeto anterior. O processo de KDD, aqui, será realizado a partir de pré-processamentos necessários para os classificadores de dados (caso necessário) para normalização de dados.

Nesse ponto é interessante citar que, quando houver pré-processamento, ele será citado no processo intrínseco ao classificador utilizado (durante a tarefa de classificação).

Mas, existe um processo necessário, porém não obrigatório, relacionado ao processo de pré-processamento que a conversão do arquivo CSV para o padrão ARFF:

Processo de conversão CSV para ARFF <passo a passo>

*Classificadores*

Penso que, ainda que já saibamos o que cada classificador faça, é bom defini-lo, com uma referência para cada um.

Naïve Bayes

<Passo a Passo>

k-NN

<Passo a Passo>

MLP

<Passo a Passo>

O MLP ;é ... confomr Fulano (www..)...

Passo a passo:

Ensembles Bagging

<Passo a Passo>

Ensembles Vote

<Passo a Passo>

Lembrar sempre de descrever se o modelo está normalizando ou não, se houve ajustes e quais foram feitos, citando os campos dos ajustes (para cada execução do classificador).

**Resultados de Discussão**

Aqui a proposta é apresentar as tabelas resultantes das matrizes de confusão, criar uma tabela, comparativa com todos os modelos executados, mostrando a eficiência de cada modelo, para apontar qual é os modelos que mais performou. Pode-se criar um gráfico, para deixar visual a diferença entre esses – para essa tarefa podemos utilizar o F-Score.

Nesse ponto, é interessante fazer as análises dos dados apresentados e, baseados nos dados, qual o melhor resultado e por que esse resultado.

**Conclusão**

**Referências**

https://www.mdsaude.com/exames-complementares/valor-de-referencia/

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, D., & Jurman, G. (2020). *Heart Failure Clinical Records*.

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.

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.

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.

**Apêndices**