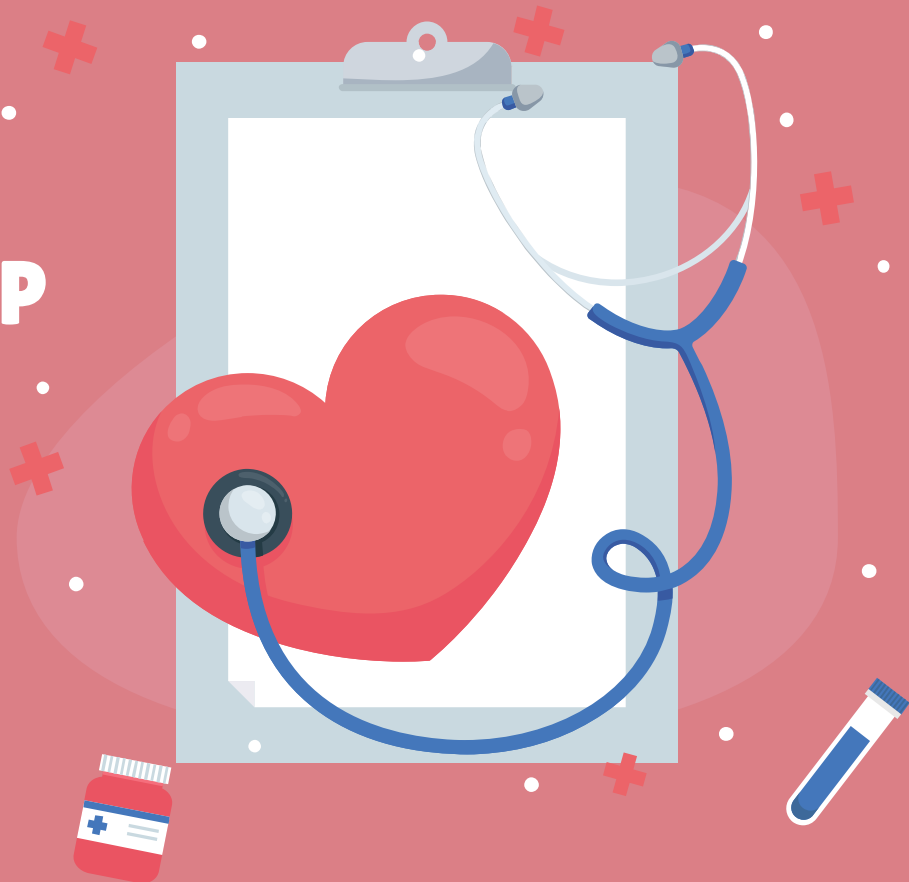


MINICURSO/SECOMP DATA SCIENCE

Professores: Márcio Henrique,
Michael Miller, Rodrigo Santos





Márcio Henrique

- Aluno de Ciência da Computação (8º período)
- Estagiário Desenvolvedor Web na DOITY.
- Secretário do DIACOM e voluntário da CS
- Pesquisador em I.A para educação



Michael Miller

- Aluno de Ciência da Computação (8º período)
- Desenvolvedor Web e Cientista de dados em projetos do NEES.
- Coordenador de comunicação do DIACOM e voluntário da CS
- Pesquisador em I.A para educação



Rodrigo Santos

- Aluno de Ciência da Computação (8º período)
- Vice-presidente do IEEE, Secretário da CS, membro e ex-presidente do DIACOM.
- Cientista de dados
- Pesquisador em I.A focada à cardiologia.

Pré-requisitos

- Conhecimento em Lógica de Programação;
- Conhecimento de alguns conceitos de Estatística;
- Conhecimento da linguagem de programação Python;
- Conhecimento de alguns pacotes como Pandas, Numpy, Matplotlib...

Introdução

- O que é Data Science?
- Pode estar em tudo e, na verdade, já está em tudo
- Uma área de muitas áreas (Física, Matemática, Estatística, Computação etc)
- KDD - Knowledge Discovery in Databases

KDD – Knowledge Discovery in Databases

1

Seleção

2

Pré-processamento

3

Transformação

4

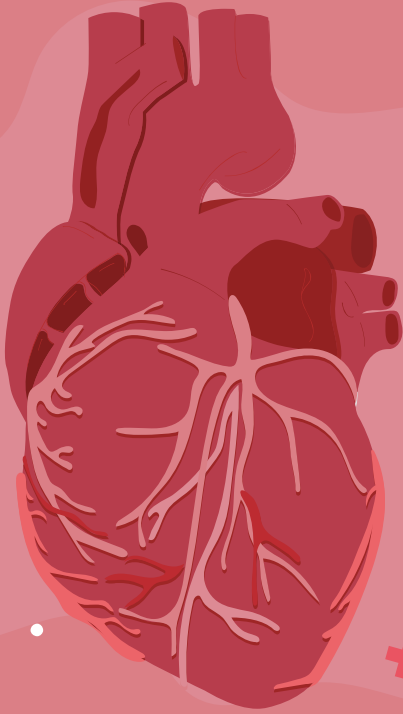
Mineração de
Dados

5

Interpretação/Avaliação

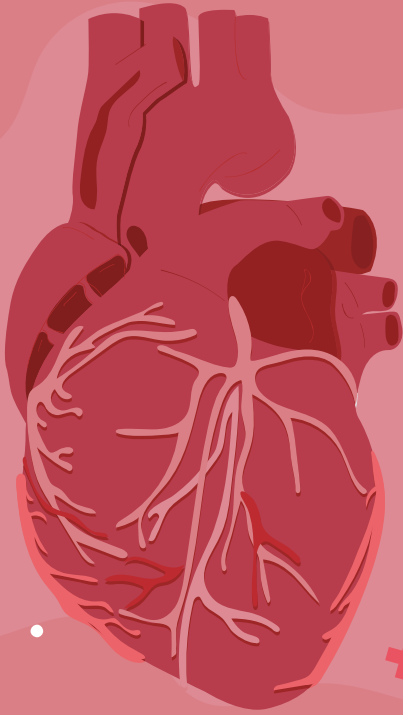
6

Conhecimento
(resultado final)



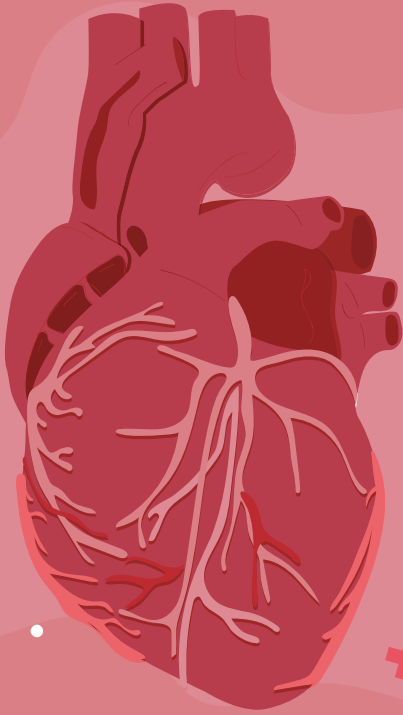
Seleção

- Selecionar um conjunto ou subconjunto de dados que fará parte da análise;
- Os dados podem estar em diferentes formatos (estruturados ou não).



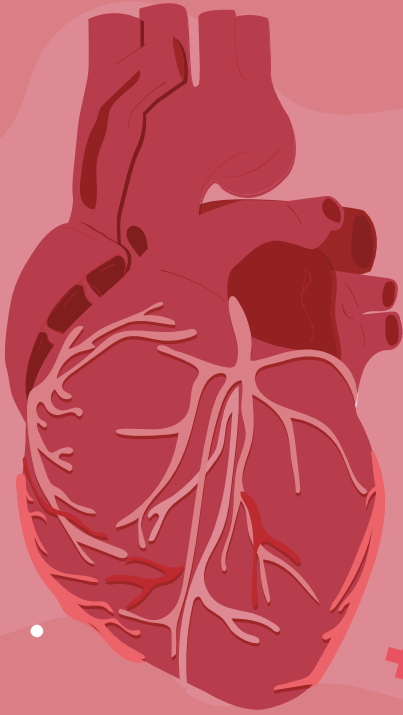
Pré – Processamento

- A base passa por um processo de limpar, corrigir ou remover dados inconsistentes;
- É feita a verificação de dados ausentes ou incompletos, além de identificar anomalias (outliers).



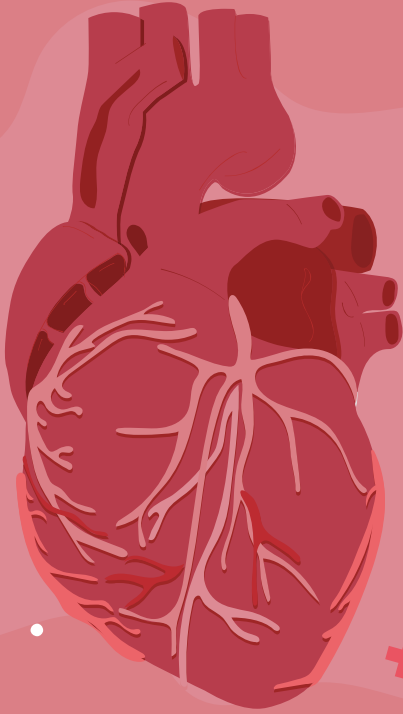
Transformação

- É Feita a normalização, agregação, criação de novos atributos, redução e sintetização dos dados
- Entre essa etapa de transformação e a próxima (mineração), os dados são agrupados para aplicação dos modelos de análise.



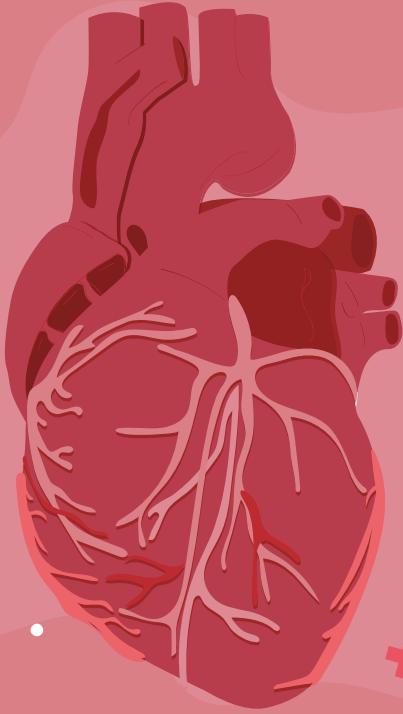
Mineração de Dados

- Construção e aplicação de modelos de mineração de dados (ML, DL);
- A descoberta do conhecimento pode ser feita através de classificação ou agrupamento, a depender do tipo de aprendizagem.



Interpretação/Avaliação

- Esta etapa consiste em avaliar o desempenho do modelo, normalmente com dados de teste;
- A validação pode ser feita através de diversas medidas estatísticas.



Conhecimento

- Informações importantes obtidas a partir dos resultados da interpretação e avaliação, de forma a resolver o problema dado.

Nosso Contexto

- Base de Dados do Kaggle: Heart Failure Prediction Dataset
- 12 colunas no dataset, onde teremos 11 colunas referente a elementos de um indivíduo e uma coluna alvo, representando se ele desenvolveu a doença ou não.
- <https://www.kaggle.com/datasets/fedesoriano/heart-failure-prediction>

Tipos de aprendizagem

Supervisionada

- Modelos preditivos com base em dados de entrada e saída
- Classificação e Regressão
- Algoritmos como KNN, SVM e árvores de decisão

Não supervisionada

- Agrupa e interpreta dados baseado apenas em dados de entrada
- Clusterização
- Algoritmos como K-means, K-medoids e DBSCAN

Dados

- 0 Age -> Idade
- 1 Sex -> Sexo
- 2 ChestPainType -> Tipo de dor no peito
- 3 RestingBP -> Pressão sanguínea em repouso
- 4 Cholesterol -> Colesterol
- 5 FastingBS -> Açúcar no sangue em jejum
- 6 RestingECG -> Resultados de eletrocardiograma em repouso
- 7 MaxHR -> Frequência cardíaca máxima alcançada
- 8 ExerciseAngina -> Angina induzida por exercício
- 9 Oldpeak -> Exercício em relação ao descanso
- 10 ST_Slope -> Inclinação do segmento ST de pico do exercício
- 11 HeartDisease -> Diagnóstico



OBRIGADO!

mhvo@ic.ufal.br, mmrc@ic.ufal.br, rss3@ic.ufal.br