Machine Learning (P02)

Artificial Intelligence, 2022-23

**João Apresentação (21152), Pedro Simões (21140), Gonçalo Cunha (21145)**

**Conteúdo**

[1. Introdução 3](#_Toc124699909)

[1.1. Contexto 3](#_Toc124699910)

[1.2. Motivação e Objetivos 3](#_Toc124699911)

[1.3. Estrutura do documento 3](#_Toc124699912)

[1.4. Data set 3](#_Toc124699913)

[1.4.1. Descrição 3](#_Toc124699914)

[1.4.2. Meta data 3](#_Toc124699915)

[1.4.2.1. Colaboradores 3](#_Toc124699916)

[1.4.2.2. Licença 3](#_Toc124699917)

[1.4.2.3. Frequência de atualização esperado 3](#_Toc124699918)

[2. Automatic classification 4](#_Toc124699919)

[2.1. Objetivos de negócio a alcançar 4](#_Toc124699920)

[2.2. Algoritmos e parâmetros selecionados 4](#_Toc124699921)

[2.3. Critérios de seleção de dados 4](#_Toc124699922)

[2.4. Preparação dos dados 4](#_Toc124699923)

[2.5. Avaliação da aplicação dos algoritmos ML 4](#_Toc124699924)

[2.6. Resultados Finais 4](#_Toc124699925)

[3. Clustering 5](#_Toc124699926)

[3.1. Objetivos de negócio a alcançar 5](#_Toc124699927)

[3.2. Critérios de seleção de dados 5](#_Toc124699928)

[3.3. Preparação dos dados 5](#_Toc124699929)

[3.4. Avaliação da aplicação do algoritmo K-Means 5](#_Toc124699930)

[3.5. Resultados finais 5](#_Toc124699931)

[4. Association rules 6](#_Toc124699932)

[4.1. Objetivos de negócio a alcançar 6](#_Toc124699933)

[4.2. Critérios de seleção de dados 6](#_Toc124699934)

[4.3. Preparação dos dados 6](#_Toc124699935)

[4.4. Avaliação da aplicação do algoritmo Apriori 6](#_Toc124699936)

[4.5. Resultados finais 6](#_Toc124699937)

[Results Analysis 7](#_Toc124699938)

[Conclusion 7](#_Toc124699939)

# Introdução

## Contexto

Este trabalho prático, relativo à unidade curricular de **Inteligência Artificial,** propende a melhorar a desenvolver um programa que recebe uma data set e desenvolve a classificação, clustering e regras de associação da própria.

Para a realização desta foi utilizado o Orange e Knime.

## Motivação e Objetivos

* Implementar e analisar diferentes abordagens de Machine Learning;
* Métodos para resolver um problema específico usando um conjunto de dados aberto/público.

## Estrutura do documento

O documento está estruturado de forma que seja de simples leitura. Existe recurso a referências de material fornecido pelo professor Joaquim Silva e/ou referências a excertos de Web grafia.

## Data set

## Descrição

O Data set selecionado para este projeto é de uma determinação da espécie da flor Iris, dividida em 3 espécies (Iris-septosa, Iris-versicolor, Iris-virginica).

A data set divide-se em 6 colunas:

* Id;
* SepalLenghtCm (comprimento da sépala em cm);
* SepalWidthCm (largura da sépala em cm);
* PetalLengthCm (comprimento da pétala em cm);
* PetalWidthCm (largura da pétala em cm);
* Species (espécie) -> Categórico.

## Meta data

## Colaboradores

UCI Machine Learning (Owner)

## Licença

[CC0: Public Domain](https://creativecommons.org/publicdomain/zero/1.0/)

## Frequência de atualização esperado

Não especificado (Atualizado 6 anos atrás)

# Classificação automática

<<The following tasks should be accomplished:

* Define the business goals to be achieved;
* Select two algorithms and the parameters to be used;
* Present the data selection criteria and explain how the data was prepared;
* Apply the ML algorithms and evaluate the generate models;
* Optimize the selected algorithm.

Document the intermediate and final results>>

## Objetivos de negócio a alcançar

## Algoritmos e parâmetros selecionados

## Critérios de seleção de dados

## Preparação dos dados

## Avaliação da aplicação dos algoritmos ML

## Resultados

# Clustering

<<The following tasks should be accomplished:

* Define the business goals to be achieved;
* Present the data selection criteria and explain how the data was prepared;
* Apply and evaluate the K-Means algorithm;
* Optimize the algorithm parameters.

Document the intermediate and final results>>

## Objetivos de negócio a alcançar

## Critérios de seleção de dados

## Preparação dos dados

## Avaliação da aplicação do algoritmo K-Means

## Resultados

# Regras de Associação

<<The following tasks should be accomplished:

* Define the business goals to be achieved;
* Present the data selection criteria and the data preparation steps;
* Apply and evaluate the Apriori algorithm, adjusting the algorithm parameters.

Document the intermediate and final results>>

## Objetivos de negócio a alcançar

## Critérios de seleção de dados

## Preparação dos dados

## Avaliação da aplicação do algoritmo A priori

## Resultados

# Análise de resultados

<< Present an analysis of the results obtained in the previous sections based on performance metrics. Include a **link to a Git repository** with the code developed under the project >>

# Conclusão

<<Include the lessons learned from the execution of the project.

The structure of the report should be adapted according to each project characteristics. Don’t forget to **remove these comments** (help text)!>>

# Bibliografia

Iris Species Data set: https://www.kaggle.com/datasets/uciml/iris