Relatório Competição 1

André Freitas Amaral

Dezembro 2022

1. **Análise exploratória dos dados**

Foi feita uma análise de cada coluna do dataset sobre como a mesma se comportava entre diversos pacientes, bem como, poderia influenciar ou ser influenciada por outras colunas, tendo ou não dados, ou dados discrepantes/inesperados.

1. **Pré-processamentos realizados**

As colunas: DS\_INDICACAO\_ACIDENTE, DS\_TIPO\_INTERNACAO, DS\_TIPO\_ATENDIMENTO, DS\_UNIDADE\_TEMPO\_DOENCA, QT\_TEMPO\_DOENCA e CD\_GUIA\_REFERENCIA foram ignoradas e zeradas pois eram irrelevantes no processo avaliativo de autorização de procedimentos.

1. **Configuração experimental**

Se utilizou Python, na versão 3, com as bibliotecas numpy, pandas, sklearn e itertools.

1. **Algoritmos utilizados**

KNN, Árvores de decisão, Naive Bayes, SVM e Florestas aleatórias.

1. **Resultados**

Entre 49 e 68% de acurácia.

1. **Referências bibliográficas**

<https://realpython.com/knn-python/>

<https://www.w3schools.com/python/python_ml_decision_tree.asp>

<https://www.kaggle.com/code/prashant111/naive-bayes-classifier-in-python>

<https://scikit-learn.org/stable/modules/svm.html>

https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html