

# CSI462 - Sistemas de Apoio a Decisão

## TP001 – 30 pontos

*Prof<sup>a</sup>. Janniele Aparecida Soares Araujo – janniele@ufop.edu.br*  
*<https://professor.ufop.br/janniele>*

21 de setembro de 2020

## 1 ORIENTAÇÕES

- Data de Entrega: **14 de Outubro de 2020 até às 12h**
- O código deverá ser disponibilizado no Google Colab ou no GitHub.
- O programa pode ser implementado em qualquer linguagem.
- A gravação da apresentação deverá ser anexados (.zip ou .rar.) na tarefa do moodle, ou então disponibilize o link para acesso.
  - Na apresentação deverão ser mostrados: o código, o seu funcionamento, as características da base de dados e um relatório com os resultados obtidos. Indique no relatório, além das informações encontradas, o link para acesso ao código. O relatório também precisa ser anexado.
  - **Receberá nota zero** o membro do grupo que não apresentar e casos de **plágio ou cópia**.

## 2 IMPLEMENTAÇÃO

O Trabalho Prático consiste em **implementar** algoritmos para as tarefas de mineração de dados (classificação, associação, agrupamento e detecção de anomalias), elaborar um **relatório** para análise da base dos dados demonstrando as suas características e resultados obtidos através da execução dos algoritmos.

- **Base de dados:** CASOS DE SÍNDROME RESPIRATÓRIA AGUDA GRAVE HOSPITALIZADOS.
  - Os alunos devem trabalhar sobre a base de dados de casos de SRAG disponibilizada pelo governo em anexo na tarefa do moodle.
  - A base de dados é grande e possui vários campos que não são obrigatórios de preenchimento ou que não são essenciais. Vocês podem visualizar o dicionário de dados disponibilizado pelo governo em anexo na tarefa do moodle.
  - Para realizar o trabalho vocês **podem** utilizar uma amostra de mil linhas da base de dados, não precisa ser a base completa, pois pode demorar para carregar.
  - Identifique quais atributos podem ser utilizados para encontrar alguma informação útil para a sua tarefa de mineração de dados.
  - Faça um pré-processamento da base excluindo os atributos irrelevantes. Faça a normalização caso os atributos selecionados não estejam normalizados.

**2.1 Grupo I - Classificação:** identifique rótulos que podem ser trabalhados para classificar os exemplos (transações) da base de dados. Por exemplo, a idade pode ser categorizada em faixa etária, tem também raça e outros. Demonstre se é possível classificar com uma boa precisão os elementos considerando alguns atributos dessa base.

**2.2 Grupo II - Associação:** identifique relações interessantes entre os atributos dos exemplos (transações) da base de dados. Por exemplo, quais fatores de risco aparecem juntos, quais fatores de risco possuem relação com determinados estados, entre outros. Demonstre o grau de confiança das regras.

- 2.3 **Grupo III - Agrupamento:** identifique grupos na base de dados. Por exemplo, quais grupos de estados mais afetados, quais os grupos de pacientes, entre outros. Mostre as características principais que dividem esses grupos e o erro do agrupamento.
- 2.4 **Grupo IV - Detecção de Anomalias:** identifique 10 exemplos (transações) mais anômalos (diferentes das demais) da base de dados. Por exemplo, o município ou estado que se destaca, um paciente diferenciado, entre outros. Mostre o grau de estranheza desses exemplos.