



UNIVERSIDADE FEDERAL DE UBERLÂNDIA

Disciplina: **Mineração de Dados**

Curso: **Gestão da Informação**

Professora: **Elaine Ribeiro Faria**

Aula Prática no 7

Tema: Comparação de Classificadores

Instruções:

- 1- Crie um relatório a partir das solicitações descritas a seguir.
- 2- Adicione um cabeçalho com a seguinte informação
Nome e nro de matrícula dos(a) alunos(a)
- 3- Gere um arquivo pdf
- 4- Envie pelo Microsoft Teams

Data da entrega: 08/12/22 às 23:59

Avaliativa: **SIM - 10 pontos**

Enunciado

- 1- Para esta atividade avaliativa deverão ser escolhidos:
 - a. Três algoritmos de classificação, sendo que pelo menos um deles deve ser um algoritmo que não foi visto em sala de aula.
 - b. Uma estratégia de divisão da base de dados em treino e teste.
 - c. Pelo menos duas medidas de avaliação.
 - d. Pelo menos duas bases de dados públicas
- 2- A tarefa consiste em avaliar o desempenho dos três algoritmos de classificação em cada uma das bases de dados usando as medidas de avaliação escolhidas.
- 3- Para a comparação dos algoritmos, o aluno poderá utilizar uma ferramenta de mineração de dados (como por exemplo, o Weka), poderá utilizar implementações prontas dos algoritmos, estratégias e medidas de avaliação disponíveis em pacotes nas diferentes linguagens de programação (ex: scikit learn no Python, pacotes da linguagem R, etc.), buscar por implementações dos algoritmos e medidas na internet e em repositórios de código, implementar os seus próprios algoritmos, como por exemplo já foi feito com o algoritmo KNN.
- 4- As bases de dados deverão ser pré-processadas adequadamente de forma a permitir a execução de cada algoritmo.
- 5- O aluno deve testar diferentes parametrizações dos algoritmos. Pelos menos duas parametrizações para cada algoritmos devem ser testadas. Ex: dois valores diferentes de K para o KNN.
- 6- Um relatório deverá ser produzido contendo:
 - a. O nome da ferramenta usada para os experimentos, ou o nome da linguagem e pacote de dado usado ou o link da implementação obtida na internet
 - b. O nome dos algoritmos utilizados e uma descrição do algoritmo, que não foi visto em sala de aula. Tente descrever os passos desse algoritmo
 - c. O nome da estratégia de divisão da base de dados em treino e teste, que foi adotada
 - d. O nome e a fórmula das medidas de avaliação utilizadas.

- e. Os parâmetros usados na configuração do algoritmo.
- f. O nome da base de dados usada, o link de onde a base foi obtida, uma descrição da base que inclui: nro de instâncias, nro de atributos, tipo de dado de cada atributo e se possui valores ausentes.
- g. Os pré-processamentos que foram aplicados na base antes de executar cada um dos algoritmos de classificação
- h. Uma discussão dos resultados obtidos realçando qual foi o melhor algoritmo para cada base de dados. É importante ter uma tabela que sumarie os resultados obtidos.