



Especificação: Avaliação 02

A segunda avaliação da disciplina consiste na aplicação e avaliação de algoritmos de classificação com e sem a seleção de atributos, no contexto de dados financeiros. A base de dados que será utilizada é proveniente de experimentos associados detecção de clientes bons e ruins dadas características como idade e movimentações financeiras em um banco. Para tal avaliação, as implementações dos algoritmos vistos em aula e disponíveis no pacote *scikit-learn* da linguagem de programação Python devem ser utilizados.

1. Classificadores

Os seguintes algoritmos de classificação devem ser considerados para aplicação e avaliação dos resultados:

- Regressão Logística (Logistic Regression)
- Árvore de Decisão (*Decision Tree*)
- *Random Forest*

Considere a acurácia e a medida F-measure na avaliação dos resultados. Por fim, compare os algoritmos, comente e apresente sua interpretação para os resultados obtidos.

2. Base de dados

German Credit Dataset¹ (credit-g). O conjunto de dados credit-g representa a qualidade de um determinado cliente dadas algumas informações sobre sua movimentação financeira. Originalmente, possui 21 observações/exemplos. Porém, as classes dessas observações estão desbalanceadas (com 700 observações de bons clientes e 300 de clientes ruins). Recomenda-se a aplicação de técnicas de pré-processamento aos dados (normalização, discretização, entre outras) e o comentário sobre seu impacto nos resultados obtidos. Caso seja de interesse, podem realizar o balanceamento dos dados e comentar os resultados obtidos.

3. Avaliação

Este trabalho será avaliado em três partes:

- **Avaliação do relatório:** Notas de 0 a 100% - **Peso 45%**
 - Utilize gráficos e tabelas a fim de facilitar a comparação e obtenção de conclusões. A análise e conclusões do grupo devem ser apresentados em forma de um relatório com no máximo **5 páginas**. Relatórios contendo somente gráficos e/ou tabelas sem análises profundas e fundamentadas receberão nota **zero**.

¹ Maiores detalhes podem ser encontrados no link:
<https://openml.org/search?type=data&status=active&id=31>



UFOP

Universidade Federal de Ouro Preto – UFOP
Instituto de Ciências Exatas e Biológicas – ICEB
Departamento de Computação – DECOM
Disciplina: BCC325– Inteligência Computacional
Professor: Jadson Castro Gertrudes

- **Precisão do melhor modelo** (porcentagem de predições corretas feitas pelo modelo): Notas de 0% a 100% - **Peso 10%**
- **Apresentação da metodologia implementada**: Notas de 0% a 100% - **Peso 45%**
- As apresentações serão realizadas nos dias **13 e 15/06/2022**.
 - Cada apresentação deve ser de, no máximo, **30 minutos**. Note que o tempo inclui a demonstração da implementação.
 - A apresentação deverá ser feita por **todos** os integrantes do grupo. Todos os integrantes do grupo devem ser capazes de explicar sobre o trabalho. Os critérios de avaliação da apresentação serão:
 - I. Clareza e organização da apresentação.
 - II. Capacidade de justificar cada passo metodologia implementada.
- O trabalho deve ser feito em grupos de 5 pessoas.
- Os grupos devem ser definidos pelos alunos até o dia **12/05/2022**. Caso os alunos não atendam o prazo, os grupos serão sorteados aleatoriamente via software.
- A ordem e a data de apresentação serão definidas aleatoriamente via software.
- As apresentações deverão ser assistidas por **todos** os alunos da disciplina. As datas de apresentações serão utilizadas para contabilização da presença naquela semana.
- Todos os grupos devem utilizar a plataforma **Google Colab** para apresentação dos experimentos.
- Casos excepcionais serão tratados pelo professor.