

Trabalho de Ciência de Dados - FIS01085

marco.idiart@gmail.com

March 2023

1 Objetivo

O objetivo do trabalho é usar métodos de Ciência de Dados para resolver um problema ou verificar uma hipótese usando para isto um banco de dados. Formato O trabalho tem que ser entregue na forma de um relatório em pdf e os scripts usados para análise em arquivos separado. O relatório tem no mínimo as seguintes seções:

- **Introdução:**

Contém uma motivação sucinta para o estudo, o tipo de sistema estudado a característica de suas variáveis, as perguntas que se quer responder e as hipóteses usadas. Uma hipótese pode ser, por exemplo, a aposta que a variável y que se quer modelar depende de um conjunto menor de preditores do que aqueles presentes no banco de dados. Por exemplo, o conjunto de dados tem 10 variáveis e você acha que só duas delas são relevantes. Neste caso o estudo visa confirmar isto. Claro que nada impede que se conclua que as variáveis inicialmente escolhidas não eram boas e que outras variáveis do banco são mais importantes. Entenda as hipóteses como pontos de partida e não como limitadores do estudo. Mas claro que uma hipótese pode ser bem mais cautelosa e simplesmente afirmar que y depende de algumas variáveis do banco sem necessariamente termos que nos comprometer. (Não menos que 1/2 página e não mais que 1 página)

- **Métodos:**

Nesta seção você descreve os métodos que você escolheu para resolver o problema, e porque escolheu eles. Se espera que no mínimo sejam usados dois métodos diferentes de ciência de dados. O ideal seria mais. Você também tem que dar informações sobre detalhes específicos de cada método como suas escolhas de hiperparâmetros e outras características. Por exemplo, se você escolheu uma rede neural, dizer quantas camadas e de que tipo serão usadas. Também tem que mencionar as técnicas de validação usadas, e quais as métricas que serão usadas para indicar a qualidade dos modelos gerados. Aqui você também descreve o conjunto de dados usado e algum tipo de manipulação feita nos dados antes de usar as técnicas de análise. (Não menos que 1/2 página e não mais que 1 página)

- **Resultados:**

Aqui estão os resultados da aplicação das técnicas no conjunto de dados. Tabelas e gráficos aparecem nesta seção, assim como descrição dos resultados, comparações, e conclusões parciais. (Aqui o número de páginas depende das figuras e tabelas)

- **Conclusões:**

Resumo dos resultados com discussão sobre as limitações dos métodos usados. (Não menos que 1/2 página e não mais que 1 página).

2 Avaliação

A avaliação terá 3 contribuições para a nota final.

1. **Apresentação** (30% da nota)

Neste item importa como o trabalho é apresentado, se está bem escrito, se tabelas e gráficos estão bem referenciados, com escalas e etc, se o relatório descreve bem o que foi feito, se é possível a partir do relatório reproduzir a análise, etc.

2. **Abordagens** (50% da nota)

Neste item será avaliado as escolhas dos métodos e a sua implementação. Por exemplo, quem escolheu métodos muito parecidos ou fez uma análise muito superficial pontua menos que quem escolheu métodos mais distintos e aprofundou.

3. **Conclusões** (20% da nota)

Neste item será avaliado se as conclusões realmente são derivadas das aplicações dos métodos. Ou seja, se avaliará o domínio teórico dos métodos.