

Projeto Final ICD

Descrição

Neste projeto vamos usar as técnicas que aprendemos em sala de aula em alguns problemas de CD. Para tal, você vai gerar um relatório (Notebook ou Latex) analisando uma base de dados de seu interesse. Caso queira, temos bases que sugerimos abaixo também. O importante é que seja uma análise além do trivial (computar uma média) de uma base que necessita de ao menos exploração, testes e inferência. Você pode usar bases do seu IC/Trabalho.

Tarefas

Com base no cronograma abaixo, você vai ser avaliado nas seguintes tarefas:

1. Definir Tema e grupo. Preencher planilha [<https://docs.google.com/spreadsheets/d/1WEKOaaaU6fZyMAEmICbNTchdZErCysDxpVWJk7hVukg/edit?usp=sharing>]. Grupos de no máximo três pessoas! Quanto mais pessoas mais trabalho será esperado do relatório. **(2pts)**.
 - a. Ao definir o tema, defina também. Uma base de dados, temos idéias abaixo.
 - b. Quais são as perguntas de ciência de dados que vão ser abordadas
 - c. Descreva brevemente a base, colunas, linhas, tipos de dados e natureza da mesma.
2. **Discussão com Monitores:** Após definir o passo acima, você já pode iniciar o trabalho do projeto. Recomendamos que seja feito ao menos uma reunião intermediária com os monitores. Marque um horário no moodle e apresente a sua ideia. **(3 pts)**
3. **Relatório Final:** Entregue um relatório sobre os resultados que você levantou para responder às suas perguntas. Sua análise dos dados devem apresentar:
 - a. **Uma caracterização (análise exploratória) inicial dos dados (2pts)**
 - b. **Pelo menos, dois testes de hipótese/intervalos de confiança (2pts)**
 - i. **Os ICs podem ser apresentados nos resultados de regressão e classificação abaixo.**
 - ii. **Os testes de hipótese também podem ser utilizados abaixo para comparar modelos.**
 - c. **Pelo menos uma regressão (3pts)**
 - d. **Pelo menos um algoritmo de aprendizado/classificação (3pts)**

No seu relatório, você deve apresentar pelo menos os seguintes pontos:

- e. Introdução com Motivação e Pergunta de Pesquisa
- f. Metodologia

- i. Descreva sua base
 - ii. Quais métodos e modelos foram utilizados. Justifique os mesmos.
 - g. Resultados. Sugiro separar em
 - i. Caracterização (análise exploratória)
 - 1. Testes de hipótese podem vir aqui.
 - ii. Previsão (uma ou duas sub-seções dependendo dos modelos utilizados)
 - h. Conclusões
 - i. Responda sua pergunta pelo menos de forma evidencial.
- 4. Apresentação.** Poster com seus resultados valendo **5pts**. Pode imprimir slides ou apresentar em Tablet/Notebook.

Cronograma

	Atividade	Data
1	Definição de Tema e Grupos	Até o dia 29/04
2	Discussão com Monitores (marquem horário)	Até o dia 29/05
3	Entrega Final	Até o dia 19/06
4	Apresentações	24 e 26 de Junho

Possíveis Bases

Saúde: <https://www.kaggle.com/russellyates88/suicide-rates-overview-1985-to-2016>

Esporte: <https://www.kaggle.com/karangadiya/fifa19>

Jogos: <https://www.kaggle.com/chuckephron/leagueoflegends> OU

<https://www.kaggle.com/rounakbanik/pokemon>

Texto (ver com Túlio): eu posso passar parte da minha base de sequências de comentários (ID_tópico, timestamp, ID_usuario, conteúdo textual)