

Atividade 1 - Lista de exercícios 1

▼ E-mail

joaovitorfd2000@gmail.com

▼ Como você definiria o que é ciência de dados? (tenta responder sem consultar, não precisa estar certo, queremos sua perspectiva)

Eu definiria ciência de dados como a área que estuda técnicas e metodologias para se lidar com os dados. Ao longo dos anos diversas ferramentas matemáticas foram desenvolvidas para que se pudesse lidar com os dados e extrair deles informações relevantes para um determinado fim.

▼ Apresente brevemente as etapas do processo de ciência de dados típico em um projeto de dados

Temos as seguintes etapas: (1) importação: trata de selecionar uma base de dados preexistente e "adicioná-la" de forma programática no código a fim de ser analisada posteriormente (2) organização/estruturação/limpeza/tidy: considerando que a base de dados pode estar inconsistente, ou seja, apresente dados com tipos diferentes, ou então tenham dados faltantes, não esteja tão bem estruturada, enfim... No geral para deixar um conjunto de dados "tidy" é necessário que: "as variáveis estejam em colunas, as observações em linhas e os valores em células." (3) entendimento: O entendimento é constituído de três partes que serão descritas em seguida. "Em linhas gerais" essa etapa se refere ao processo do cientista em conseguir manipular o conjunto de dados em prol de extrair dele diversas informações de seu interesse. Sejam elas uma composição matemática de diversas variáveis, um belo gráfico ou a exposição explícita algum tipo de padrão percebido. (3.1) transformação: A primeira das sub etapas consiste em utilizar os dados já existentes nas tabelas com o propósito de gerar outros dados que possam vir a ser mais elucidativos. (3.2) visualização: Aqui os dados analisados são representados de uma forma visual para que fique mais agradável e claro ao observador a forma com que os dados estão se comportando. E auxiliar na questão da descoberta de relações

implícitas entre os dados analisados. (3.3) modelagem: Trata da busca por alguma propriedade matemática emergente das relações existentes entre os dados. (4) comunicação: Embora os dados já tenham sido limpos, estruturados, transformados, etc. isso acaba não importando aos outros caso eles não sejam bem expostos. E é do que se trata essa etapa. Nela ocorre a busca por uma boa, com "boa" me refiro a eficiente e explicativa, explicação dos achados em relação aos dados estudados.

▼ Trapalho final

Lembrando que agora é só um rascunho, vamos perguntar novamente mais perto da entrega do projeto.

▼ Qual data set você selecionou para analisar e por que? (pode ser sincero)

O dataset escolhido será um que foi sugerido por Tang durante a disciplina de heurísticas em relação a organização de grades de matérias.

Provavelmente relacionado a este link: <https://www.unitime.org/>

▼ Qual pergunta você gostaria de responder sobre esse data set?

Perguntas: qual a heurística mais eficiente para encontrar o resultado? Qual a melhor forma de mostrar a evolução das heurísticas? Como podemos apresentar o resultado final?

▼ Como pretende responder a pergunta usando o dataset escolhido?

Seguirei de trás pra frente. Primeiro buscarei como apresentar um resultado final. Depois tentarei fazer um gráfico dinâmico que represente a evolução que a heurística terá. Com isso, após N interações, podemos salvar o resultado final encontrado e tentar achar um novo resultado com uma nova heurística.