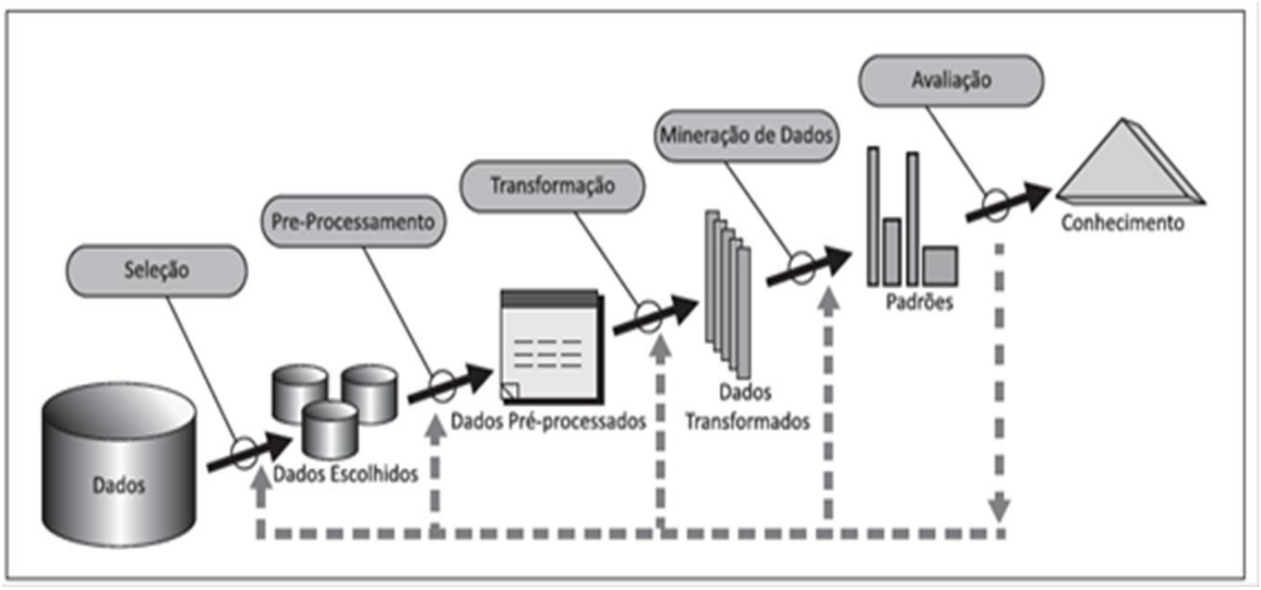
IFSULDEMINAS - Campus Passos
Banco de Dados II

Prof. Hiran Nonato
Atividade Avaliativa – Mod II – 24/06/2020
Valor: 1,0 ponto

Conteúdo: Processamento de descoberta de conhecimento – Seleção de Atributos e transformação dos Dados.

Exercícios

P1) Neste exercício continuaremos a trabalhar com o Processo de Descoberta de Conhecimento em Bases de Dados. Nesta semana iremos aprofundar um pouco mais nesse processo e analisar mais algumas etapas que utilizaremos em propostas reais de análise e mineração de dados. Aqui, a sua missão será ajustar um conjunto de dados em SQL para que possa ser melhor trabalhado no Python ou em alguma outra linguagem que permita a manipulação e análise dos dados.



Subtarefa 1: Importe o arquivo "dados.sql", em um banco de dados MySQL. Este banco de dados contém o mesmo conjunto de dados utilizado na última atividade avaliativa (pizzaria), só que agora em formato SQL.

Subtarefa 2: Realize a seleção dos atributos, ou seja, a partir desse banco de dados importado você deverá retirar aqueles atributos que não são essenciais para o seu processo de descoberta de conhecimento.

Neste caso, você não precisará dos seguintes atributos:

- Número do pedido;
- Cliente;
- Endereço;
- Telefone;
- Valor pizza;
- Valor entrega;
- Hora da entrega;

Subtarefa 3: Transformação dos dados. Nesta etapa você deverá realizar algumas transformações para que os atributos fiquem mais adequados para que o atributo a-priori possa ser utilizado.

Por exemplo, o campo data_pedido está com uma data no formato '2011-10-07'. Neste caso, dificilmente o algoritmo conseguirá encontrar padrões para essa data. Seria muito mais interessante se utilizasse o dia da semana. Neste caso, você conseguirá saber se o pedido foi realizado no meio da semana ou no final da semana.

Sendo assim, você deverá transformar os seguintes dados:

a) data_pedido
transformar essas datas para um padrão de dia da semana (segunda, terça, quarta, ...)
Exemplo: 24/06/2020 -> quarta

Sugiro que você utilize uma função para isso.

b) hora_pedido
transformar as horas de pedido em intervalos de tempo (Início, Pico e Final)
se (hora < 20:00) = Início
se (hora >= 20>00 e hora < 22:00) = Pico
se (hora >= 22:00) = Final

Sugiro que você utilize uma função para isso.

c) **valor_borda** e **valor_refrigerante**

transformar esses dados para que informe somente se houve ou não o adicional de borda ou a compra ou não de refrigerante.

se (valor_borda > 0) = borda sim
se (valor_borda <= 0) = borda não

se (valor_refrigerante > 0) = refrigerante sim
se (valor_refrigerante <= 0) = refrigerante não

d) valor_total e tempo_decorrido
a transformação do valor total e do tempo decorrido será diferente. Aqui você deverá utilizar uma distribuição de frequência para identificar quais são os melhores intervalos para alocação dos dados. Neste caso, você deverá utilizar a distribuição de "scott". Note que o arquivo "*pizzaria.csv*" utilizou outra distribuição.

O código abaixo te auxiliará a entender essa distribuição. Execute-o no Python.

```
dados = genfromtxt('tempo_decorrido.csv')
histograma = plt.hist(dados, bins="sturges")
#histograma = plt.hist(dados, bins="scott")
#histograma = plt.hist(dados, bins="fd")
#histograma = plt.hist(dados, bins=4)
#histograma = plt.hist(dados)
```

- *O arquivo tempo_decorrido.csv de exemplo também está disponível no GoogleClassroom.*

e) **Finalização.** Após todo esse processo, você deverá ter um arquivo semelhante ao utilizado na última aula (pizzaria.csv). Verifique se está ou não semelhante. Qual a diferença?

Entrega

Você deverá entregar um resumo (máximo duas páginas) sobre as etapas realizadas de seleção e transformação de dados para o processo descoberta de conhecimento. Neste resumo você deverá explicar, com suas palavras, o que você entendeu sobre esse processo e a importância dessas etapas para a mineração de dados. Além disso, você deverá apresentar todos os SQLs utilizados para a transformação dos dados (as functions).

Entrega até o dia 07/07/2020 às 23:59.
Trabalho individual.