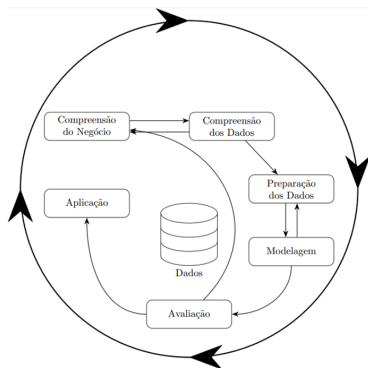


Aluno(a): _____ Matrícula: _____

- 1 ponto 1. É comum encontrar dados faltantes quando estamos analisando um dataset. Cite pelo menos duas maneiras que permitem tratar esse tipo de situação.
- 1 ponto 2. Um dataset pode conter dois tipos de atributos: categóricos e numéricos. Diferencie-os.
- 1 ponto 3. Cite duas boas práticas que devemos adotar ao realizar scraping de uma página web.
- 1 ponto 4. Diferencie as três abordagens para modelar um Data Warehouse: estrela, floco de neve e constelação.
- 1 ponto 5. O analista de dados João está interessado em criar um modelo que consiga prever o resultado de jogos do Campeonato Brasileiro. Ele decidiu seguir o processo CRISP-DM nesse projeto. Nesse momento do projeto, ele está realizando a coleta de dados dos campeonatos de 2003 até 2024 em sites esportivos e fazendo uma análise estatística desses dados. Em que fase do CRISP-DM João se encontra?



- A. Compreensão do Negócio
B. Compreensão dos Dados
C. Preparação dos Dados
D. Modelagem
E. Avaliação
F. Aplicação

- 1 ponto 6. Considere as afirmativas sobre Modelagem Multidimensional:
- I. Na modelagem dimensional, a chave da tabela de fatos corresponde à associação das chaves primárias das tabelas de dimensões.
- II. Data Warehouse é baseado em um modelo de dados multidimensional em que dados formam uma estrutura chamada data cube
- III. A modelagem em Data Cubes não permite que sejam feitas operações de agregação nos dados
- Quais as afirmações corretas?
- A. Apenas I
B. Apenas II
C. I e II
D. II e III
E. I, II e III

- 1 ponto 7. Associe os processos de Pré-Processamento à esquerda com suas respectivas descrições:
- | | | | |
|-------|------------------------------|-----|---|
| (1) | Agregação | () | Realizar a padronização e normalização dos dados |
| (2) | Amostragem | () | Identificação e exclusão de valores nulos no dataset |
| (3) | Discretização | () | Técnica utilizada para realizar seleção de um conjunto menor de registros de um dataset |
| (4) | Remoção de Valores Faltantes | () | Criação de novas colunas no dataset, pela modificação de alguma coluna existente |
| (5) | Seleção de Features | () | Combinação de dois ou mais registros em um único |
| (6) | Transformação de Atributos | () | Reduz a dimensionalidade dos dados removendo colunas irrelevantes |

- 1 ponto 8. Considere as afirmativas sobre os processos de pré-processamento:
- I. Outliers são registros que possuem valores consideravelmente diferentes do que o resto dos outros objetos do dataset
 - II. Os valores ruidosos referem-se a modificações dos valores originais e que, portanto, consistem em erros de medidas ou em valores consideravelmente diferentes da maioria dos outros valores do conjunto de dados, os chamados outliers. Há duas abordagens para o tratamento destes dados: inspeção e correção manual; e identificação e limpeza automática.
 - III. Procedimentos para integração de dados consistem em realizar ações que permitam integrar, adequadamente, dados provenientes de diversas fontes de dados. Geralmente, quando os dados provêm de diferentes fontes, os valores assumidos por atributos não possuem o mesmo domínio ou não estão formatados sob o mesmo tipo de dado, ainda que digam respeito à mesma descrição de uma entidade do mundo real. As principais motivações para a aplicação de procedimentos de integração de dados são, portanto, a presença de valores inconsistentes e a presença de valores redundantes.
- Sobre as alternativas acima, é CORRETO afirmar que:
- A. Apenas a sentença I é verdadeira.
 - B. As sentenças I e II são verdadeiras.
 - C. As sentenças II e III são verdadeiras.
 - D. Todas as sentenças são verdadeiras.
 - E. Todas as sentenças são falsas.

- 1 ponto 9. Segundo a definição de Data Warehouse, qual das características a seguir não se aplica a esse tipo de sistema?
- A. Utilizada para registrar operações diárias e transacionais
 - B. Suporte a Tomada de Decisão
 - C. Orientada a Assunto
 - D. Integrada
 - E. Temporal

- 1 ponto 10. É uma técnica baseada na extração de dados a partir da leitura do código-fonte de páginas na internet. Esse é o conceito de:
- A. Mineração de Dados
 - B. Recuperação de Informações
 - C. Web Scraping
 - D. Sistema de Gerência de Bancos de Dados
 - E. Beautiful Soup