



Ciência de Dados

Atualmente, a área de Ciência de Dados tem ganhado muita popularidade, devido à viabilidade concreta de podermos extrair valor agregado dos dados. Esse valor se materializa no uso de tecnologia que nos permitem implementar algoritmos, com o conhecimento de especialistas, em negócios que vão dos empresariais até os de caráter científico. Nesse sentido, vamos abordar brevemente os seguintes tópicos, que são muito importantes para a área de Ciência de Dados:

- Aspectos Essenciais sobre Dados;
- Fundamentos para Tomada de Decisão;
- Estatística e Ciência de Dados;
- Linguagem de Programação Python.

Aspectos Essenciais sobre Dados

Os dados são os pilares para qualquer aplicação de Ciência de Dados. Eles precisam tratar explicitamente os seguintes aspectos:



Os tipos: podem ser numéricos, datas e alfanuméricos. Por exemplo, o nome de uma empresa é dado do tipo alfanumérico, já a quantidade de livros que

uma pessoa lê por ano é um dado do tipo inteiro.



Como são organizados: se os dados são estruturados, não-estruturados ou semiestruturados. Podemos encontrar exemplos de dados estruturados nas tabelas de bancos de dados. Em relação aos dados não-estruturados, temos exemplos dos arquivos de vídeo. Por fim, na situação em que trabalhamos com dados semiestruturados, temos como exemplos os arquivos nos formatos XML e JSON.

Além da questão do tipo e organização dos dados, também precisamos tratar outros aspectos, que levam em consideração as propriedades de volume, variedade, veracidade, velocidade e valor. Essas propriedades, inclusive, constituem-se na base da área conhecida como BigData. Portanto, precisamos levar em consideração todos esses aspectos, quando tratamos com dados.

Agora, vem uma questão muito importante: para que servem esses dados? Vamos responder a essa pergunta no próximo tópico: fundamentos para tomada de decisão.

Fundamentos para Tomada de Decisão

Vimos que os aspectos dos dados podem ser bastante complexos. Portanto, precisamos de uma boa justificativa para realizar todo esse trabalho. Podemos encontrar essa justificativa exatamente no valor agregado dos dados, que formam a base para a tomada de decisão.

Constantemente, gestores das mais diversas áreas precisam realizar escolhas que vão causar impactos sobre os negócios, portanto é natural

que eles utilizem de ferramentas que lhes dêem informações importantes sobre os impactos das decisões que precisam tomar. Essas ferramentas são concretizadas através do uso de algoritmos que empregam conhecimento do negócio e técnicas de estatística, matemática e aprendizado de máquina. E como é que essas áreas se relacionam para apoiar os gestores na tomada de decisão?


Vamos tratar dessa questão através da explicação do relacionamento entre estatística e os demais aspectos de ciência de dados.

Estatística e Ciência de Dados

A estatística trata, explicitamente, do uso de técnicas para estudarmos os dados. Ela é um dos elementos básicos da Ciência de Dados, pois nos permite buscar por relações não-triviais que sejam valiosas. Por exemplo, podemos empregar as técnicas estatísticas de séries temporais para estudar o comportamento do índice de pluviosidade de uma determinada região. E qual seria a utilidade prática desse tipo de estudo?

Bem, através do entendimento de como o índice de pluviosidade se comporta ao longo do tempo, podemos traçar estratégias para proteger a população contra possíveis desastres. Ao entendermos as relações entre os dados, podemos gerar os elementos fundamentais para o suporte da tomada de decisão, através de técnicas bem consolidadas que nos dão precisão no processo de análise. Mas como podemos implementar essas técnicas?

Para implementarmos as técnicas de estatísticas, precisamos conhecer ferramentas de software específicas, como as linguagens de programação, que nos oferecem bibliotecas com a capacidade de

escalar soluções. Uma das linguagens mais populares atualmente, para aplicações de Ciência de Dados, é o Python. No próximo tópico, vamos abordar alguns aspectos dele. 

Linguagem de Programação Python

A linguagem de programação Python é muito popular para o desenvolvimento de aplicações de Ciência de Dados. Essa popularidade foi alcançada devido a diversos fatores, tais como:

- Facilidade de sintaxe;
- Facilidade de encontrar ambientes de desenvolvimento;
- Disponibilidade de muita documentação de domínio público com diversos exemplos;
- Comunidade fortemente engajada.

Um exemplo de ambiente de desenvolvimento para programarmos em Python é o Google Colab. Para podermos utilizar essa ferramenta, precisamos ter uma conta no Gmail e acessar o endereço: <https://colab.research.google.com/> (acesso em 24/02/2023).

O Python possui diversas bibliotecas que facilitam o nosso trabalho de desenvolvimento. Abaixo, temos um exemplo de um programa em Python implementado no Google Colab:



```
a = int(input("Digite um numero: "))
```

```
b = int(input("Digite um numero: "))
```

```
maior = a
```

```
if(b>maior):
```

```
    maior = b
```

```
print(f'O maior numero eh: {maior}')
```

O objetivo desse programa é comparar dois números e encontrar o maior deles. Vamos fazer um programa mais interessante em Python, agora. O objetivo do código abaixo é criar uma lista com cinco dados aleatórios do tipo inteiro e com uma variação de até 100 unidades entre eles:

```
import numpy as np
```

```
from scipy import stats
```

```
variacao=100
```

```
quantidade=5
```

```
dados = np.random.randint(variacao,size=quantidade)
```

```
print("dados:{}".format(dados))
```

Agora, vamos estudar as características estatísticas básicas desses dados através do código abaixo:



```
media = dados.mean()

desvio_padrao=dados.std()

variancia=dados.var()

minimo = dados.min()

maximo = dados.max()

moda = stats.mode(dados)

print("media:{}".format(media))

print("desvio padrão:{}".format(desvio_padrao))

print("variância:{}".format(variancia))

print("mínimo:{}".format(minimo))

print("máximo:{}".format(maximo))

print("moda: {}; quantidade: {}".format(moda.mode,moda.count))
```

É fundamental que você tente executar os códigos para adquirir segurança e tenha mais facilidade nos exemplos que vamos estudar mais adiante.

Bem, agora que você adquiriu esses conhecimentos chegou a hora de realizarmos exercícios que nos ajudarão a consolidar o nosso

aprendizado. Bons estudos!



Atividade Extra

Uma dica bastante interessante é que você assista o filme “Moneyball – o homem que mudou o jogo”. Além de ser um bom filme, você terá a possibilidade de ver como a utilização adequada dos dados pode gerar resultados surpreendentes.

Referência Bibliográfica

- FORBELLONE, A. L. V.; EBERSPACHER, H. F. **Lógica de programação: a construção de algoritmos e estruturas de dados**. 3.ed. Pearson: 2005.
- LEAL, G. C. L. **Linguagem, programação e banco de dados: guia prático de aprendizagem**. Intersaberes: 2015.
- VICCI, C. (Org.). **Banco de dados**. Pearson: 2014.
-

PUGA, S.; RISSETTI, G. **Lógica de programação e estruturas de dados com aplicações em java**. Pearson: 2016.



- ELMASRI, R.; NAVATHE, S. B. **Sistemas de banco de dados**. 7.ed. Pearson: 2018.

- ASCENCIO, A. F. G.; ARAUJO, G. S. de. **Estrutura de dados: algoritmos, análise da complexidade e implementações em java e C/C++**. Pearson: 2010.

- PUGA, S.; FRANÇA, E.; GOYA, M. **Banco de dados: implementação em SQL, PL/SQL e Oracle 11g**. Pearson: 2013.

- MEDEIROS, L. F. de. **Banco de dados: princípios e prática**. Intersaberes: 2013.

Ir para exercício