



# Licap

## Formação Cientista de Dados



**PUC Minas**

# Formação do Cientista de Dados

## Definições Preliminares – Módulo Básico

Luis Enrique Zárate

## Definições básicas para Ciência de Dados

1. O que é dado
2. O que é idéia acerca do dado
3. O que é informação
4. O que é conhecimento
5. O que é aprendizado
6. O que é descoberta de conhecimento
7. Exemplo ilustrativo

# ➡ Definições preliminares para o futuro Licap cientista de dados

- O que é dado?
- O que é idéia?
- O que é informação?
- O que é conhecimento?
- O que é aprender ?
- O que é descobrir?
- O que é descoberta de conhecimento em bases de dados, KDD?

# Conceitos preliminares

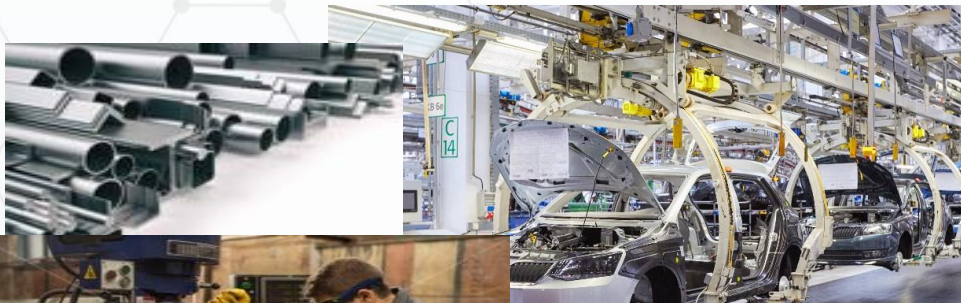


O que é Dado?  
O que é Idéia?  
O que é Informação?  
O que é Conhecimento?  
O que é Aprender ?  
O que é Descobrir?  
O que é KDD?

# ➡ O que é Dado?



- Do lat. *Datus*, elemento ou quantidade conhecida que serve de base à solução de algum problema.
- O dados é um estímulo sensorial: um símbolo, uma imagem, um som.
- O dado é qualquer representação de fatos ou idéias dos quais pode ser extraída informação para um processo de descoberta de conhecimento.





# ➡ O que é Idéia?

- Do lat. *Ideia*, é uma representação abstrata, elaborada pelo pensamento, acerca de um objeto.
- Ter uma idéia é ter uma noção acerca de um dado e do conceito que este representa num processo de descoberta de conhecimento.



# ➔ O que é Informação?



- Do lat. *Informatio*, significado de um Dado.
- A informação existe quando o cérebro humano recebe um conjunto de dados e os processa neuralmente com o conhecimento prévio.
- Informação é a compreensão acerca do significado de um dado. A informação é a que julgamos relevante para um processo de descoberta de conhecimento.

# ➔ Dado vs. Informação



Se a temperatura externa está em  $40^{\circ}\text{C}$  (dado)

**Temos informação acerca da sensação térmica que nos espera ao sair (existe uma informação)**

Se nos informam que a temperatura externa está em  $95^{\circ}\text{F}$  (dado)

**Não temos informação do que nos espera ao sair (não há informação). Haverá informação se conhecemos a conversão  $^{\circ}\text{F}$  para  $^{\circ}\text{C}$  (conhecimento prévio).**

# ➡ O que é Conhecimento?



- Do lat. *Cognoscere*, ato ou efeito de entender e compreender as propriedades e as características de alguma “coisa” ou problema.
- A construção do conhecimento envolve os processos de **Percepção** e **Reflexão** (processo cognitivo).

# ➡ O que é Conhecimento?



Num **processo de descoberta de conhecimento**, a **Percepção** correspondem à captura de dados que consideramos necessários para explicar um problema.

A **Reflexão** corresponde à busca por relações entre os dados buscando conhecimento novo e não óbvio, via algoritmos computacionais.

# ➔ O que é Descobrir?



- Do lat. *discooperire*, encontrar o que estava escondido e era desconhecido.
- Descobrir padrões a partir dos dados.
- Em Data mining e Machine Learning interessa a descoberta de conhecimento útil e não-óbvio.

# ➡ O que é Aprender?



- Do lat. *Apprehendere*, adquirir conhecimento de alguma coisa; ficar sabendo; instruir-se.
- Compreender como os dados, de forma individual ou em conjunto, explicam um problema
- O aprendizado a partir dos padrões descobertos sustentam a tomada de decisão.

# Exemplo ilustrativo



Consideremos que o objetivo é encontrar o perfil de pessoas Longevas e não-Longevas

**DADOS:**

Gênero	Relação com amigos	Educação formal
Níveis de ansiedade e estresse	Limitações físicas	Carreira profissional
Relação com crianças	Sintomas de depressão	Dores e suas intensidades
Frequência de atividade física	Sentimentos positivos	Resultados de testes de mobilidade
Consumo de álcool e tabaco	Relação com o parceiro(a)	Status econômico atual
Boa autoavaliação	Histórico médico	Uso de medicamento
Relação com família	Satisfação com a vida e perspectiva	Sentimentos negativos
Resultados de teste de memória	Casa e automóvel	.....



# Exemplo ilustrativo



Consideremos que o objetivo é encontrar um perfil de pessoas Longevas e não-Longevas

**IDÉIA:**  
**CONCEITO**  
**QUE O DADO**  
**REPRESENTA:**

Relação com família
Limitações físicas
Sintomas de depressão
Sentimentos positivos
Relação com o parceiro(a)
Boa autoavaliação
Satisfação com a vida e perspectiva



Uma boa relação com a família pode contribuir para diminuição da stress e do sentimento de abandono



O sentimento de satisfação com a vida traz tranquilidade para viver evitando a stress.

# Exemplo ilustrativo



Consideremos que o objetivo é encontrar um perfil de pessoas Longevas e não-Longevas

**INFORMAÇÃO  
CONTIDA NO  
DADO**

Gênero
Ansiedade e estresse
Carreira
Frequência de atividade física
Consumo de álcool e tabaco
Boa autoavaliação
Relação com família
Resultados de teste de memória



Dependendo da região, homens podem ter uma expectativa de vida menor que as mulheres.

Caso a região não possua diferenças sociais a expectativa seria igual para ambos, daí pode não haver informação relevante no atributo GÊNERO.

# Exemplo ilustrativo



Consideremos que o objetivo é encontrar um perfil de pessoas Longevas e não-Longevas

## DESCOBERTA DE CONHECIMENTO:

SE            Problemas sérios saúde < Nível 4    AND  
                 Mobilidade < Nível 3                    AND  
                 Estrutura social < 5

Então **não-Longevo**

SE            Freq. Atividade física > 7                    AND  
                 Histórico médico > 4

Então **Longevo**

.....



# Licap

## Formação Cientista de Dados

Obrigado!



**PUC Minas**