



# Mineração de Dados Aplicada à Engenharia

Profº - Dr. Thales Levi Azevedo Valente

[thales.l.a.valente@gmail.com.br](mailto:thales.l.a.valente@gmail.com.br)

# Sejam Bem-vindos !



**Os celulares devem  
ficar no silencioso  
ou desligados**

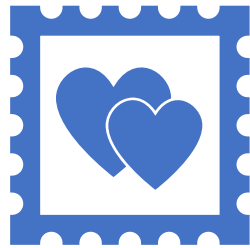
Pode ser utilizado  
apenas em caso  
de emergência



**Boa tarde/noite, por  
favor e com licença  
DEVEM ser usados**

Educação é  
essencial

# Na aula de anterior...



## **Realizar uma dinâmica para conhecer um ao outro**

Discutir sonhos e desejos

A importância de ter um objetivo definido



## **Discutir boas práticas de estudo**

Importância de um cronograma

Importância do foco

Importância de revisões periódicas

Alimentação e exercício

# Objetivos de hoje



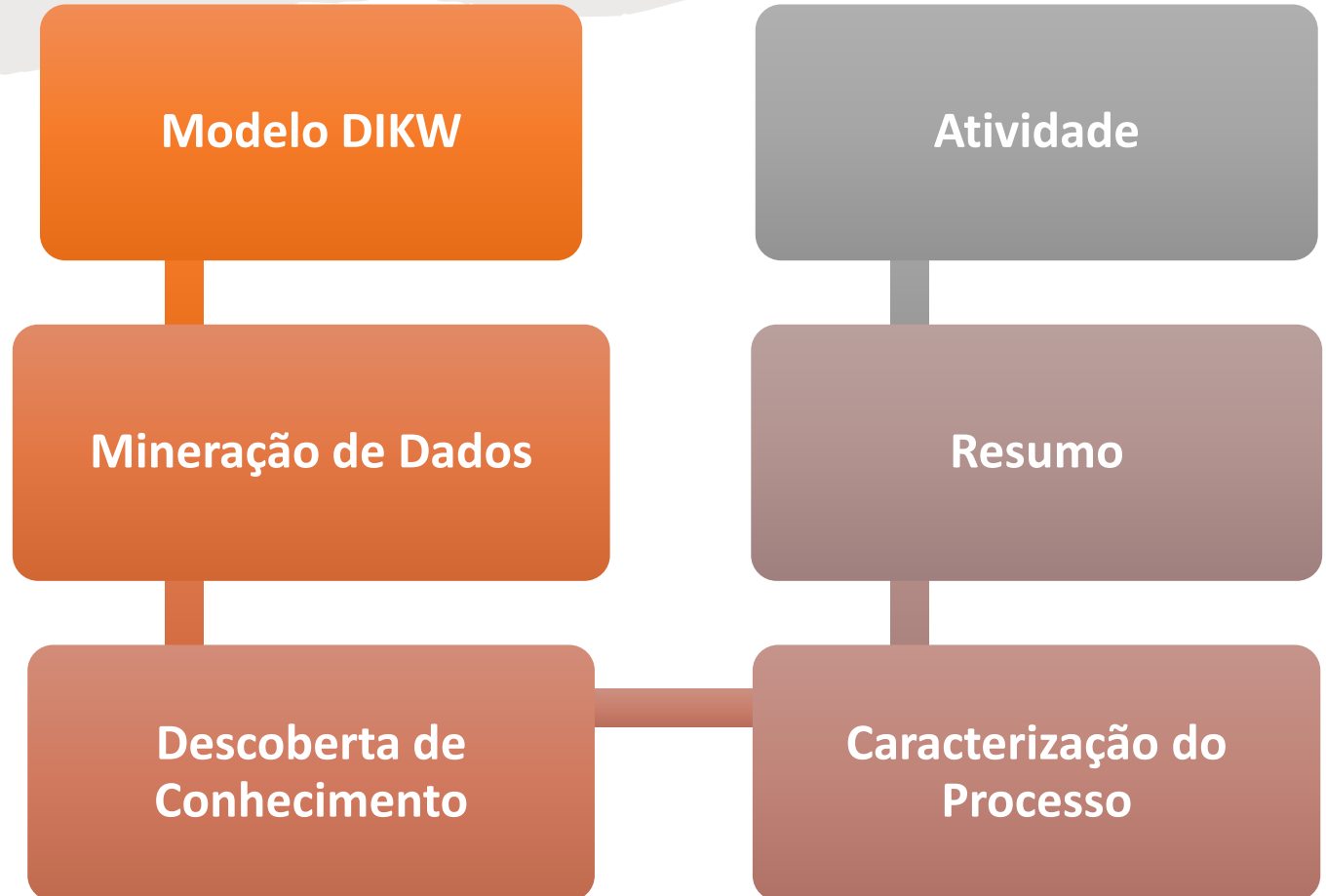
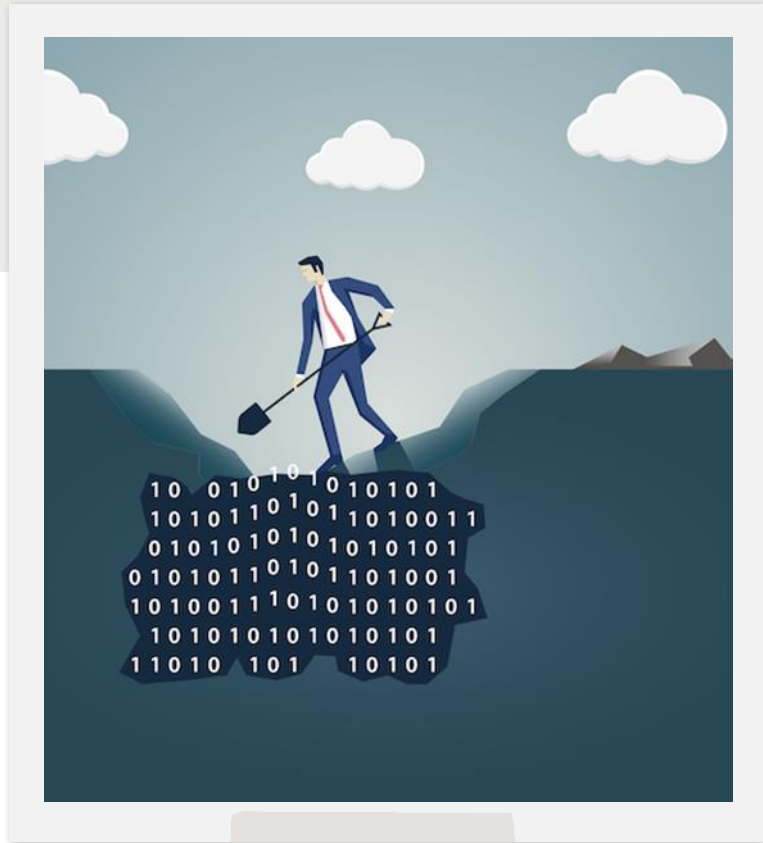
Mostrar conceitos introdutórios sobre mineração de dados, descoberta de conhecimento e aplicações para essas áreas



Ao final da aula, os alunos serão capazes de definir o modelo DIKW, ter uma idéia do processo de mineração de dados e identificar áreas em que podem aplicar



# Roteiro



# Algumas Perguntas

---

- Que livros um cliente da Amazon gostaria de comprar ?
- É seguro dar crédito em dinheiro a uma determinada pessoa ?
- É possível detectar o roubo de um cartão de crédito pelo seu uso ?
- É possível detectar falhas em sensores?
- É possível detectar falhas em componentes mecânicos?
- É possível extrair informações de documentos e fornecer insights?
- Como faço para juntar diferentes bases de dados referentes a um mesmo problema?
- É possível fazer análise de gráficos de criptomoedas e encontrar relacionamentos?

Mas....

- 
- O que são Dados?

# O modelo DIKW

---

- DIKW
  - Data (Dados)
  - Information (Informação)
  - Knowledge (Conhecimento)
  - Wisdom (Sabedoria)
- Hierarquia relacionada a conceitos sobre conhecimento
- Conceitos são encadeados definindo
  - Contexto
  - Compreensão



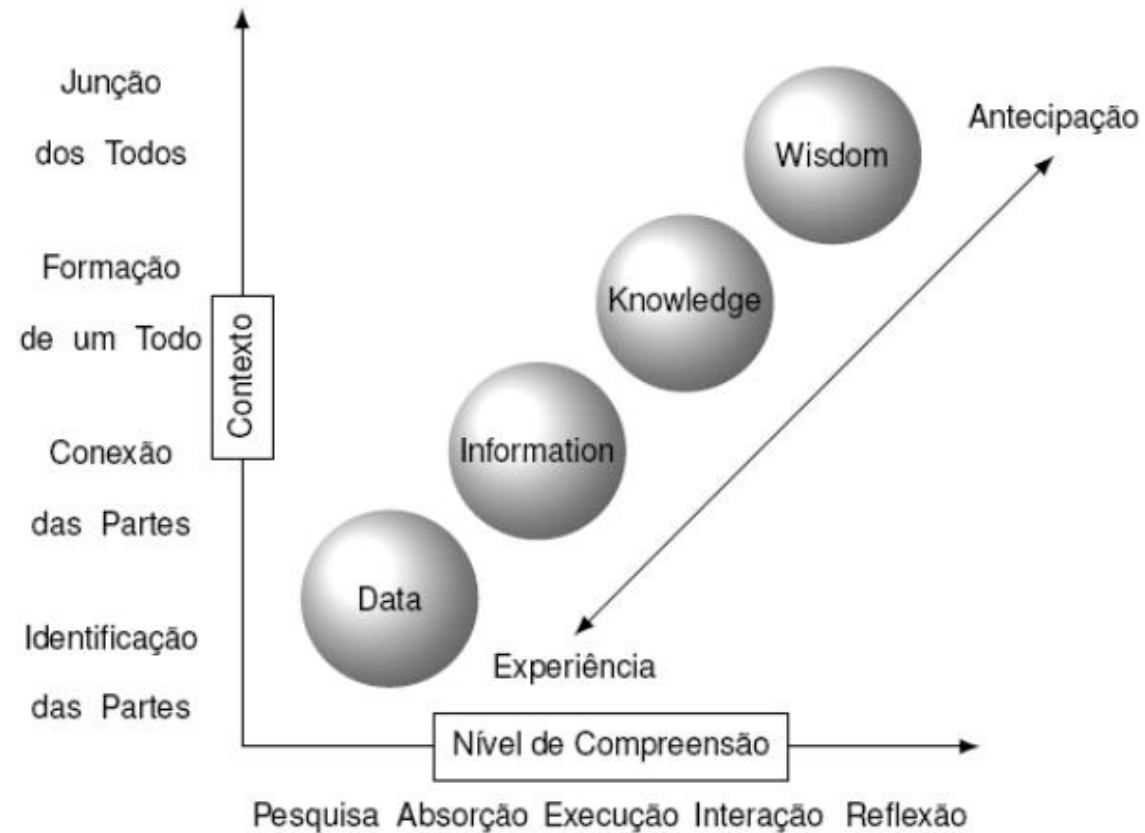
# O modelo DIKW

---

- DIKW
  - Os seus componentes em ordem crescente de importância:
    - **Dados** (Data) é o nível mais básico;
    - **Informação** (Information) acrescenta contexto e significado aos dados;
    - **Conhecimento** (Knowledge) acrescenta a forma como usar adequadamente a informação;
    - **Sabedoria** (Wisdom) acrescenta o entendimento de quando utilizá-los.
- Desta forma, a hierarquia DIKW é um modelo teórico que se mostra útil na análise e no entendimento da importância e limites das atividades dos ***trabalhadores do conhecimento***.

# O modelo DIKW

- Representação do Eixo



# Dados e Informação

---

- Dados: um símbolo, um fato ou evento sem relação com outras coisas.
  - Ex. Neva
- Informação: compreensão de uma relação ou contextualização de dados. Por exemplo, uma relação de causa e efeito.
  - Ex. A temperatura caiu 8 graus e então começou a nevar.

# Conhecimento e Sabedoria

---

- Compreensão de um padrão que normalmente permite deduzir o que acontecerá ou identificar algo descrito por um conjunto de fatos ou símbolos.
  - Ex. Se a umidade do ar está muito elevada e a temperatura cai a 40C, então a atmosfera muito provavelmente não consegue segurar as gotículas de água e então estas gotículas congelam durante a queda e neva.
- Agrega mais que um entendimento de princípios fundamentais incorporados ao conhecimento que são essencialmente a base para o conhecimento que faz o que ele é.

# A Origem da “Data Information Knowledge Wisdom”

---

- Frank Zappa fez alusão a hierarquia em 1979 ["Packard Goose" in album Joe's Garage: Act II & III (Tower Records, 1979)]:
  - Information is not knowledge,
  - Knowledge is not wisdom,
  - Wisdom is not truth,
  - Truth is not beauty,
  - Beauty is not love,
  - Love is not music,
  - and Music is THE BEST

# Hierarquia de Ackoff: Data-Information-Knowledge-Understanding & Wisdom

---

- Compreensão exige diagnóstico e prescrição, que considera serem mais que "conhecimento" menos que sabedoria.
- Enquanto as informações agem rapidamente, o conhecimento tem uma vida mais longa e compreensão tem apenas uma aura de permanência.
- Sabedoria é considerada como "permanente" no verdadeiro sentido.

# DIKW

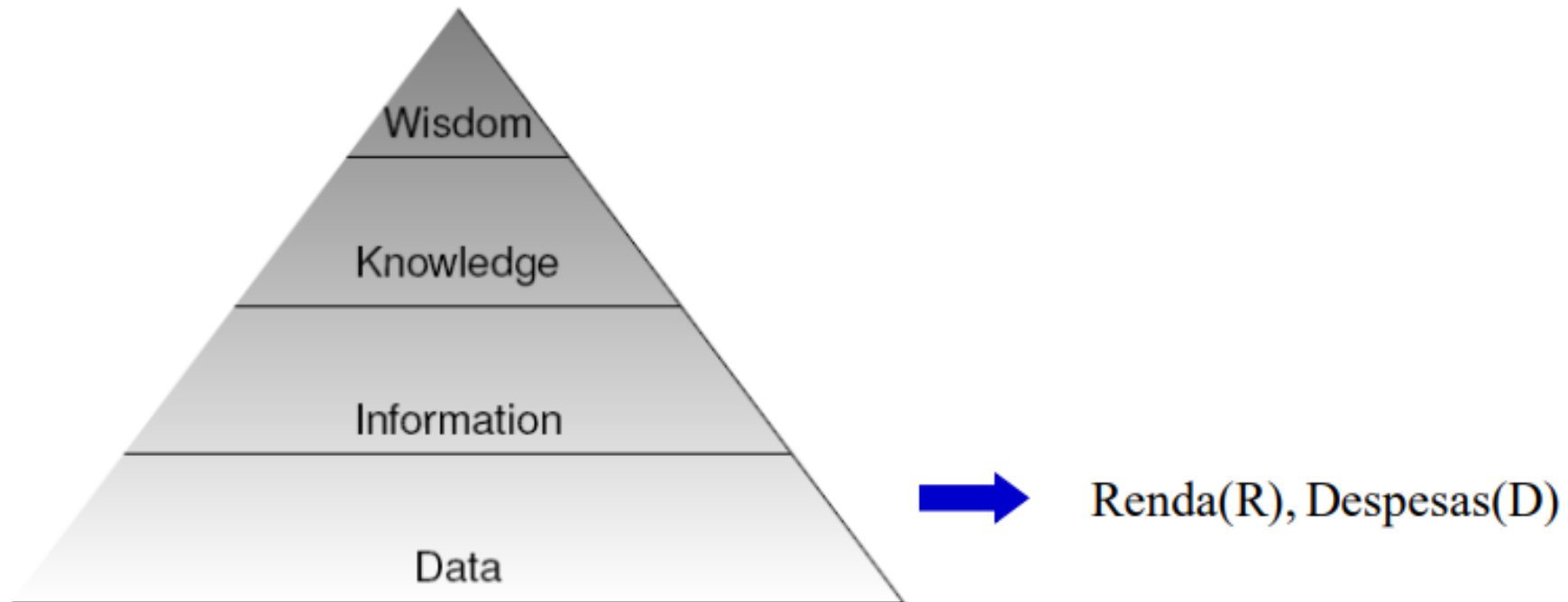
TOM CHALKLEY



December 1982 issue of THE FUTURIST.

# Pirâmide do Conhecimento

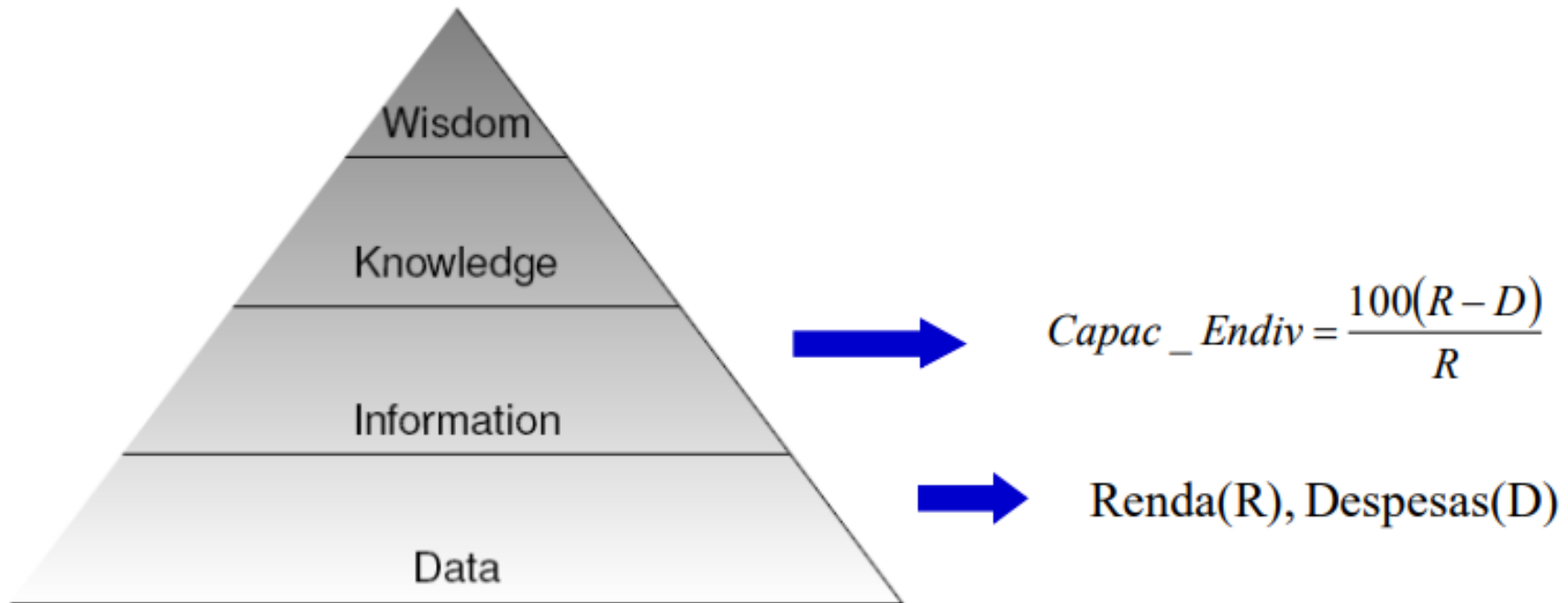
---





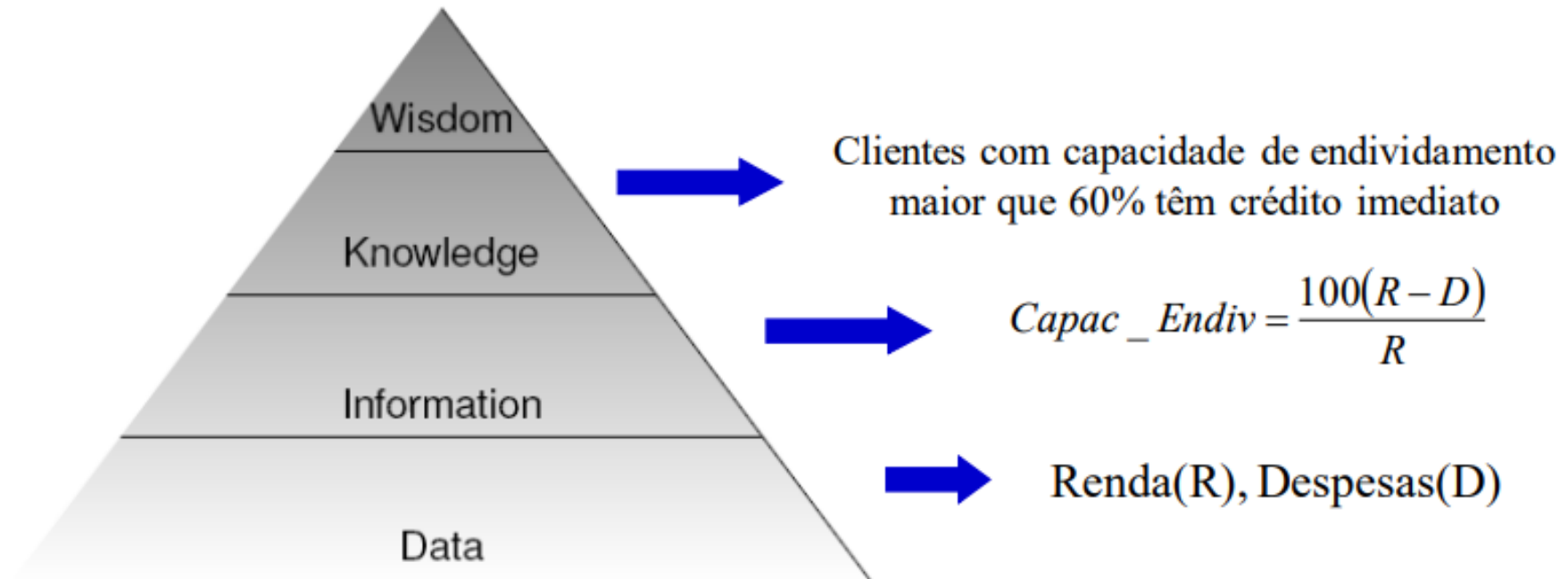
# Pirâmide do Conhecimento

---



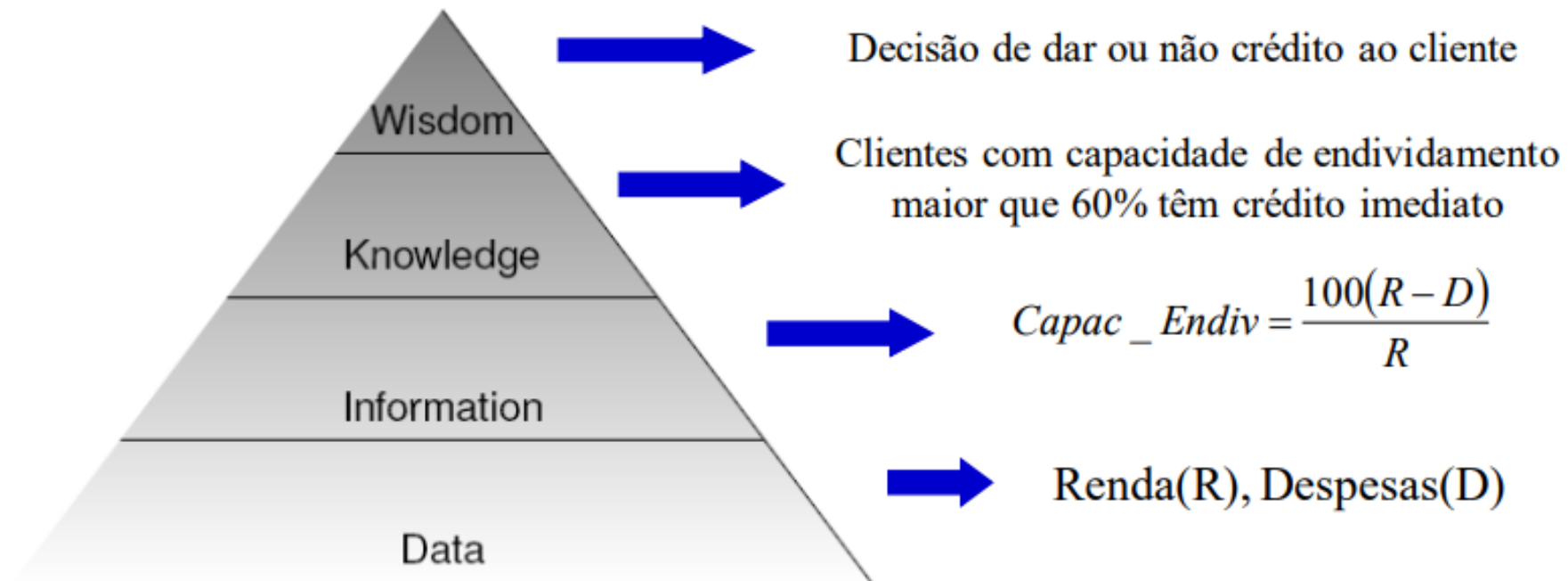
# Pirâmide do Conhecimento

---



# Pirâmide do Conhecimento

---



# Consider:

---

- I have a box.
- The box is 3' wide, 3' deep, and 6' high.
- The box is very heavy.
- The box has a door on the front of it.
- When I open the box it has food in it.
- It is colder inside the box than it is outside.
- You usually find the box in the kitchen.
- There is a smaller compartment inside the box with ice in it.
- When you open the door the light comes on.
- When you move this box you usually find lots of dirt underneath it.
- Junk has a real habit of collecting on top of this box.
- What is it?

# Pirâmide do Conhecimento

---

- Os **dados** são os elementos através dos quais é possível se obter **informação**
- Com **informação** é possível construir **conhecimento**

## Mineração de Dados

Através da investigação dos dados  
podemos chegar a descoberta do  
conhecimento

# Mineração de Dados

---

- Não há consenso sobre definição
  - Termo recente aplicado a confluência de idéias de estatística e ciência da computação
  - Terminologia também não é padronizada
- Definições podem ser restritas ou abrangentes
  - Estatística em grandes bases de dados
  - Reconhecimento de padrões
  - Descoberta de conhecimento

# Definição do Termo

---

“Mineração de Dados é o processo de **descoberta** de **novas** e **significativas** correlações, padrões e tendências em **grandes volumes de dados**, através do uso de técnicas de reconhecimento de padrões, estatística e outras ferramentas matemáticas.”

Gartner Group

# Definição do Termo

---

- Multidisciplinaridade
  - Estatística
  - Aprendizado de Máquina e Inteligência Computacional
  - Banco de Dados
  - Reconhecimento de Padrões



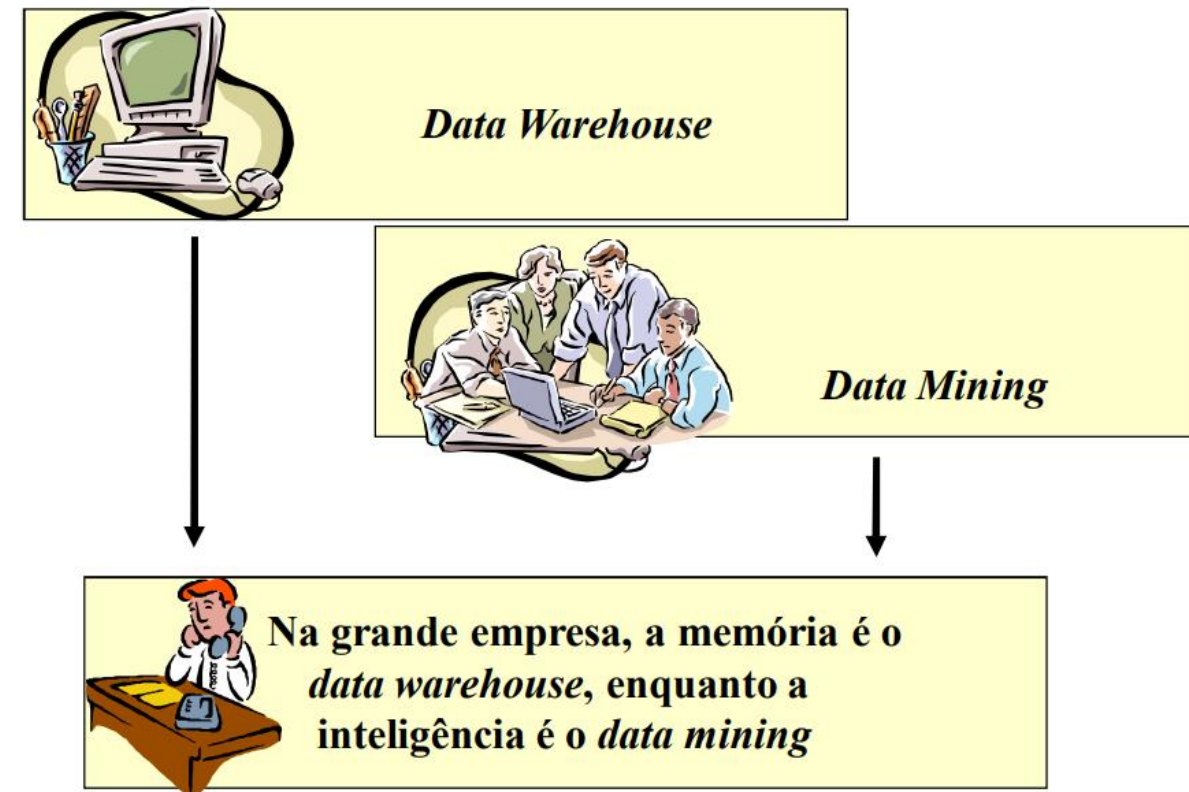
# Outra Definição

---

- Descoberta de novos padrões em bancos de dados
  - Padrões devem ser úteis (novos e válidos)
  - Padrões podem ser inesperados
- Pode envolver um conjunto de técnicas auxiliares
  - Limpeza de dados
  - Visualização
  - Warehousing
- Podemos agregar essas técnicas auxiliares em uma definição mais abrangente ?

# Podemos Estender essa Definição ?

- A descoberta do conhecimento se resume apenas a análise dos dados ?
  - No mundo real os dados não estão prontos para serem prontamente analisados
  - Ser humano necessita de formas intuitivas para visualizar resultados



# Descoberta de Conhecimento em BD

---

- KDD: Knowledge Discovery in Database
- Existem nomes tais como: knowledge discovery in database, data mining, knowledge extraction, information discovery, data archaeology, information harvesting e ainda data pattern processing

“KDD é um processo, de várias etapas, não trivial, iterativo e iterativo, para identificação de padrões compreensíveis válidos, novos e potencialmente úteis a partir de grandes conjuntos de dados.”

U. M. Fayyad et al., “The KDD Process for Extracting Useful Knowledge from Volumes of Data”, 1996

# Descoberta de Conhecimento em BD

---

“KDD é um processo, de **várias etapas**, não trivial, iterativo e iterativo, para identificação de padrões compreensíveis válidos, novos e potencialmente úteis a partir de grandes conjuntos de dados.”



# Descoberta de Conhecimento em BD

---

“KDD é um processo, de **várias etapas**, não trivial, iterativo e iterativo, para identificação de padrões compreensíveis válidos, novos e potencialmente úteis a partir de grandes conjuntos de dados.”



# Descoberta de Conhecimento em BD

---

“KDD é um processo, de **várias etapas**, não trivial, iterativo e iterativo, para identificação de padrões compreensíveis válidos, novos e potencialmente úteis a partir de grandes conjuntos de dados.”



# Descoberta de Conhecimento em BD

---

“KDD é um processo, de várias etapas, **não trivial**, iterativo e iterativo, para identificação de padrões compreensíveis válidos, novos e potencialmente úteis a partir de grandes conjuntos de dados.”

## Fatores Operacionais

- Volumes de dados grandes e heterogêneos
- Tratamento de resultados em diferentes formatos
- Integração de algoritmos específicos

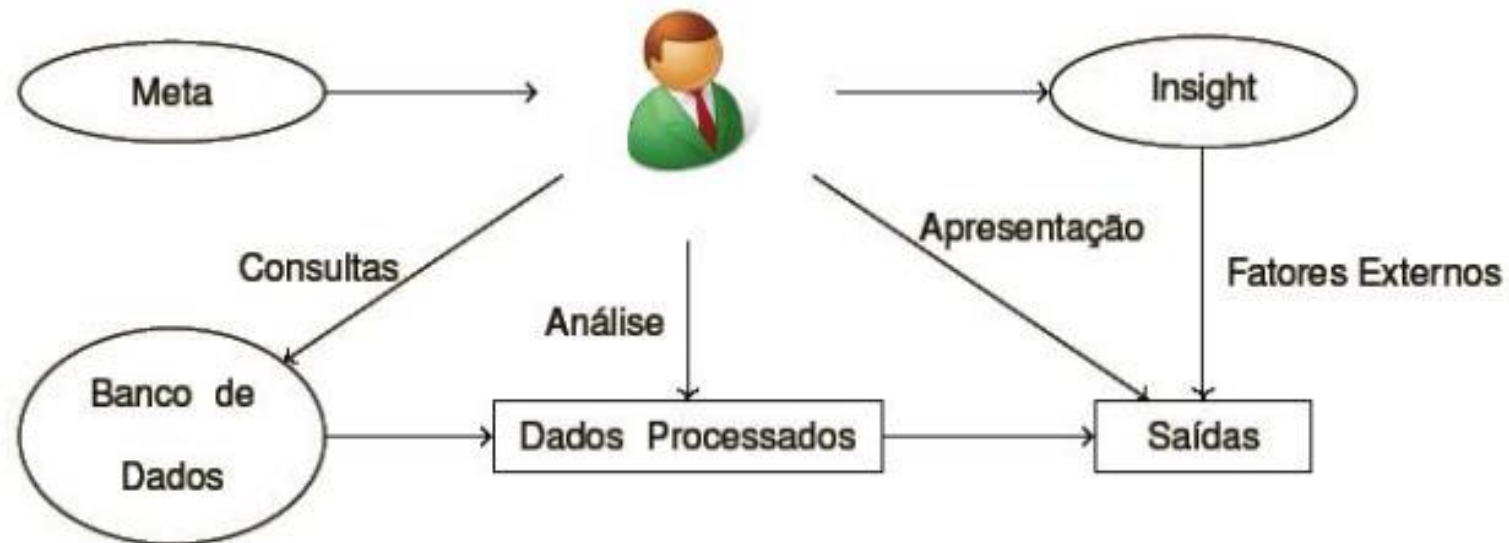
## Fatores de Controle

- Formulação dos objetivos
- Escolha do algoritmo
- Interpretação dos resultados



# Descoberta de Conhecimento em BD

“KDD é um processo, de várias etapas, não trivial, **interativo** e iterativo, para identificação de padrões compreensíveis válidos, novos e potencialmente úteis a partir de grandes conjuntos de dados.”





# Descoberta de Conhecimento em BD

---

“KDD é um processo, de várias etapas, não trivial, interativo e **iterativo**, para identificação de padrões compreensíveis válidos, novos e potencialmente úteis a partir de grandes conjuntos de dados.”



# Descoberta de Conhecimento em BD

---

“KDD é um processo, de várias etapas, não trivial, interativo e iterativo, para identificação de padrões compreensíveis **válidos**, novos e potencialmente úteis a partir de grandes conjuntos de dados.”

- Conhecimento deve ser verdadeiro
- Adequado ao contexto da aplicação (Diagnóstico de Doenças x Livros Amazon)

# Descoberta de Conhecimento em BD

---

“KDD é um processo, de várias etapas, não trivial, interativo e iterativo, para identificação de padrões compreensíveis válidos, **novos** e potencialmente úteis a partir de grandes conjuntos de dados.”

- Descobertas óbvias não interessam
- Conhecimento gerado deve ser novo

# Descoberta de Conhecimento em BD

---

“KDD é um processo, de várias etapas, não trivial, interativo e iterativo, para identificação de padrões compreensíveis válidos, novos e potencialmente **úteis** a partir de grandes conjuntos de dados.”

- Conhecimento deve proporcionar benefícios

# Caracterização do Processo

---

- Aplicação de KDD é dividida em 3 componentes
  - Problema a ser submetido
  - Recursos disponíveis para a solução do problema
  - Resultados obtidos através do uso dos recursos

# Caracterização do Processo

---

- Problema
  - Conjunto de Dados
    - Aspecto Intensional – estrutura dos dados
    - Aspecto Extensional – Casos ou registros
  - Especialista no Domínio
    - Representa pessoa que conhece o assunto
  - Objetivos da Aplicação
    - Características esperadas do modelo
    - Exemplo: precisão mínima de 85% ao conceder crédito
    - Podem não estar muito claros no início do processo

# Caracterização do Processo

---

- Recursos Disponíveis
  - Especialista em KDD
  - Ferramenta de KDD
    - Ambiente de Mineração de Dados
    - Algoritmos Isolados
  - Plataforma Computacional
    - Hardware
    - Capacidade de Processamento
    - Memória

## •Dev

- **Identificar e utilizar conhecimento *a priori* sobre o problema**
- **Escolher ferramentas e métodos**
- **Direcionar as ações do processo**
- **Conduzir a avaliação dos resultados**

# Caracterização do Processo

---

- Resultados Obtidos
  - Modelo de Conhecimento
    - Deve ser avaliado com relação ao cumprimento das expectativas definidas nos objetivos
    - Usado para comparações
  - Históricos
    - Como os modelos de conhecimento foram gerados
    - Melhor controle do processo
    - Permitem análise e revisão das ações realizadas



# Macroobjetivos e Orientações

---

- Aplicação de KDD pode ser classificada em duas dimensões
  - Orientação das Ações
    - Validação de Hipóteses Postuladas
    - Descoberta de Conhecimento
  - Macroobjetivos
    - Predição – permite fazer previsão a partir de históricos
    - Descrição – permite descrever o conhecimento existente na base

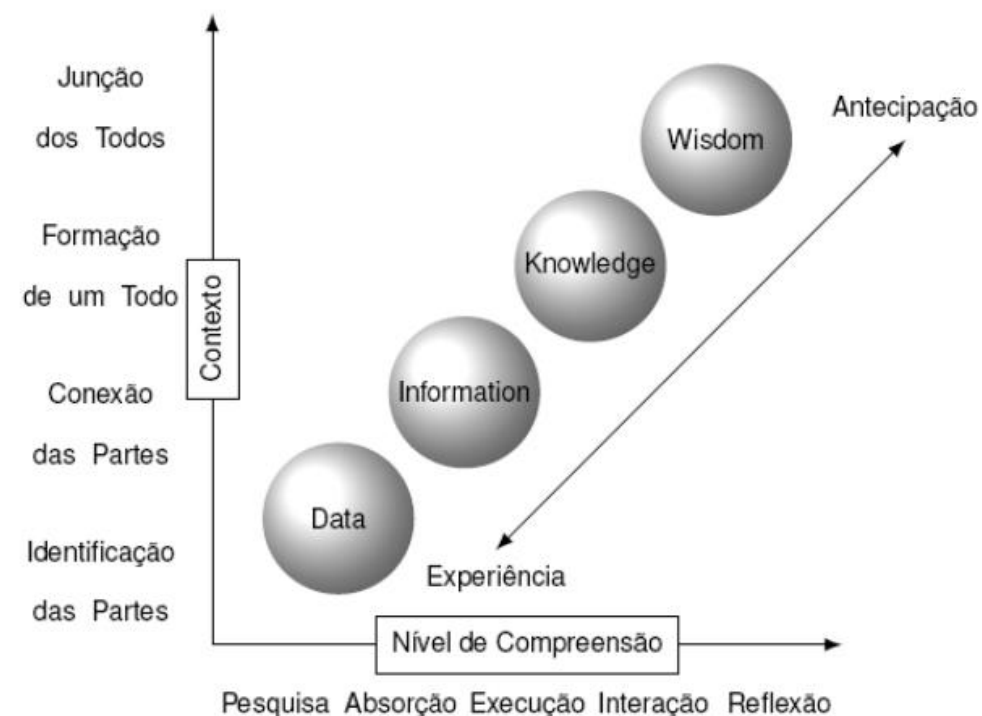
# Aplicações

---

- Bancária (aprovação de crédito),
- Ciências e medicina (descoberta de hipóteses, diagnóstico, classificação, predição),
- Comerciais (segmentação, localização de consumidores, identificação de hábitos de consumo),
- Engenharia (simulação e análise, reconhecimento de padrões, processamento de sinais e planejamento),
- Financeira (apoio para investimentos, controle de carteira de ações),
- Gerencial (tomadas de decisão, gerenciamento de documentos),
- Internet(ferramentas de busca, navegação, extração de dados),
- Manufatura(modelagem e controle de processos, controle de qualidade, alocação de recursos),
- Segurança(detecção de bombas, icebergs e fraudes) etc. Análise de Churn

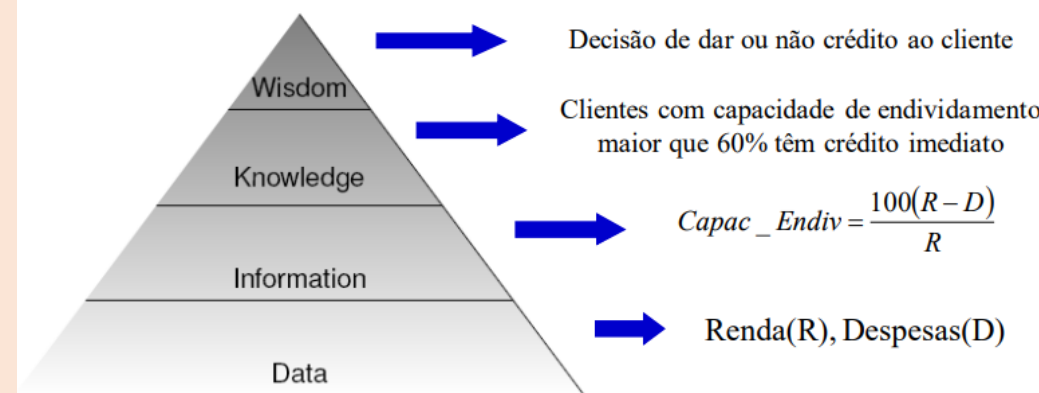
# Resumo

“KDD é um processo, de **várias etapas**, não trivial, interativo e iterativo, para identificação de padrões compreensíveis válidos, novos e potencialmente úteis a partir de grandes conjuntos de dados.”



“Mineração de Dados é o processo de **descoberta** de **novas** e **significativas** correlações, padrões e tendências em **grandes volumes de dados**, através do uso de técnicas e reconhecimento de padrões, estatística e outras ferramentas matemáticas.”

Gartner Group



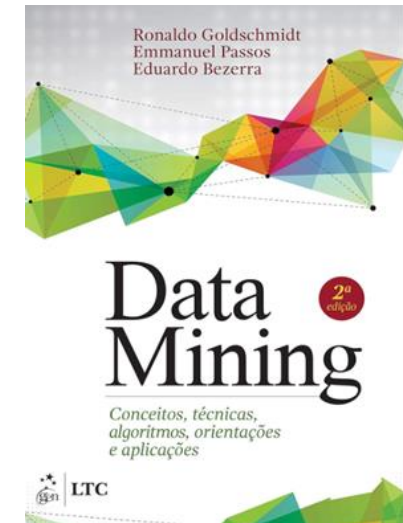
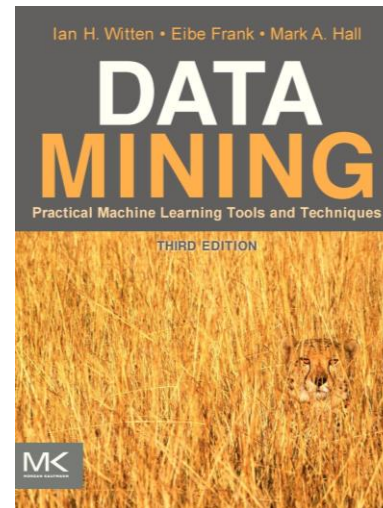
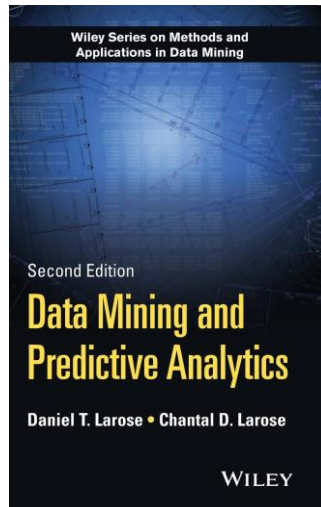
# Atividade

---

- Listar 5 problemas dentro da engenharia em que a mineração de dados pode ser aplicada e com qual objetivo
- Buscar bases de dados públicas para 3 problemas distintos
- Listar 5 bibliotecas para trabalhar com dados descrevendo o objetivo de cada uma delas
- Definir 3 temas para trabalho final

# Bibliografia

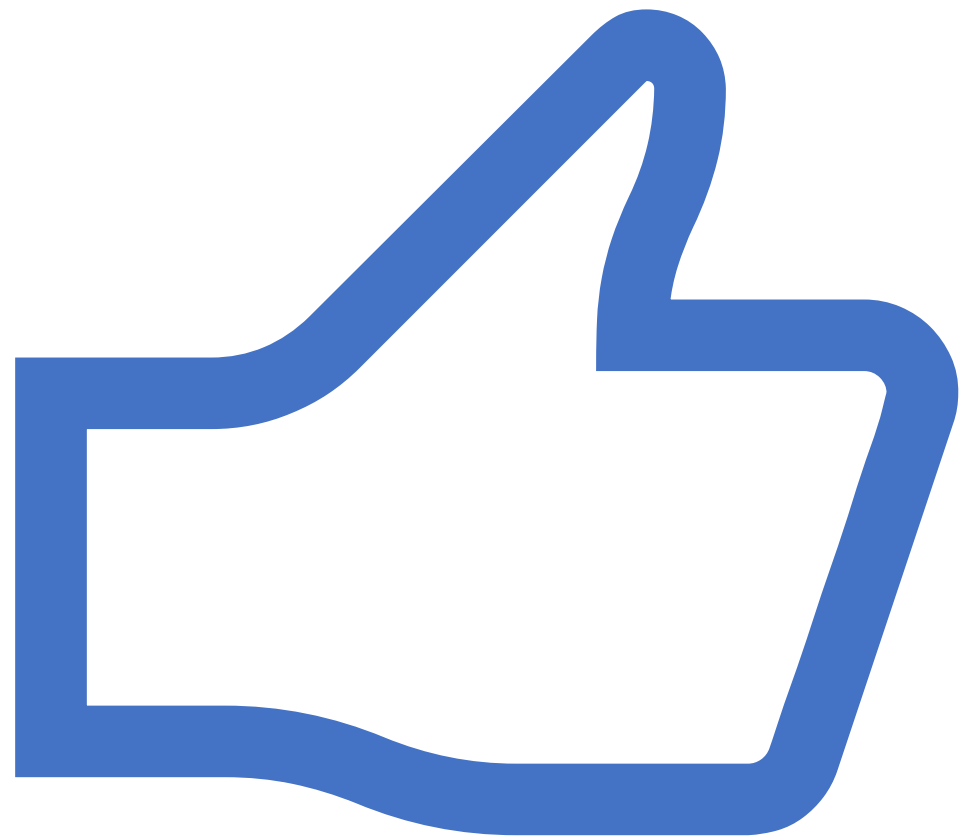
---



# Dúvidas?



Obrigado !





Apresentador

Thales Levi Azevedo Valente

E-mail:

thales.l.a.valente@gmail.com