

# INTRODUÇÃO A CIÊNCIA DE DADOS NA ENGENHARIA DE PETRÓLEO

# Aula 1 - Introdução

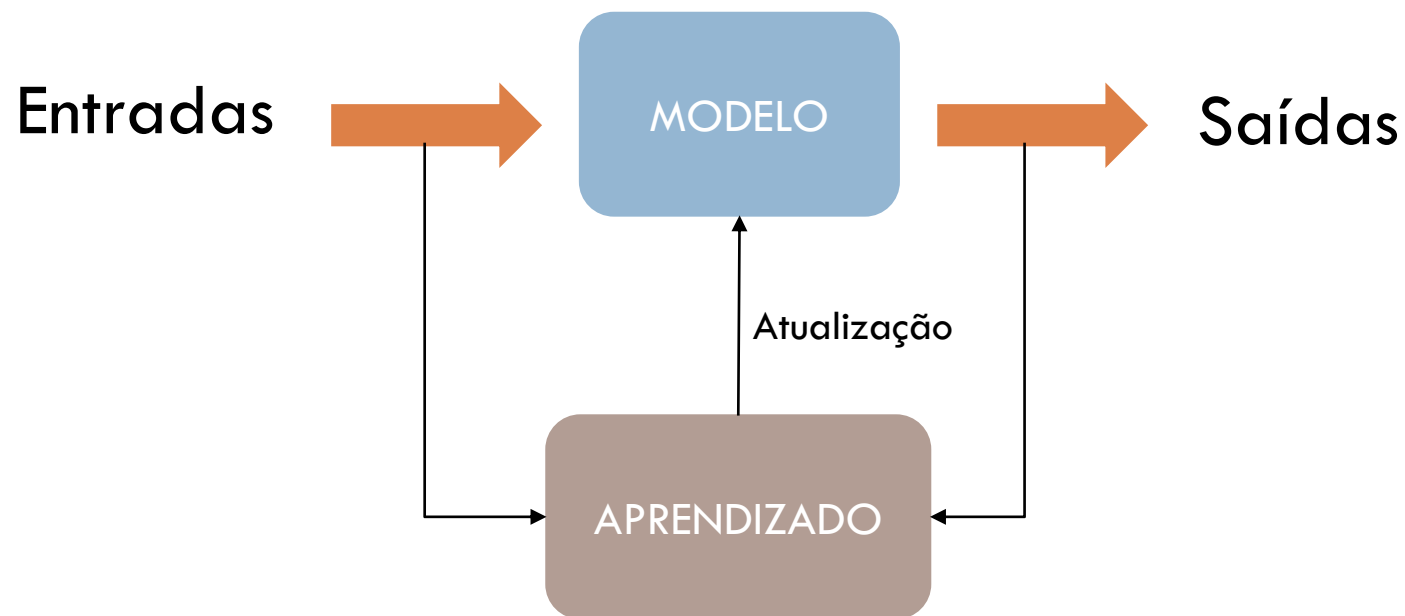
2

- Modelagem Matemática
- Modelos baseados em física x Modelos baseados em dados
- Ciência de Dados
- Aplicações na Indústria de Petróleo
- Sites Indicados

# Modelagem Matemática

3

## □ MODELAGEM

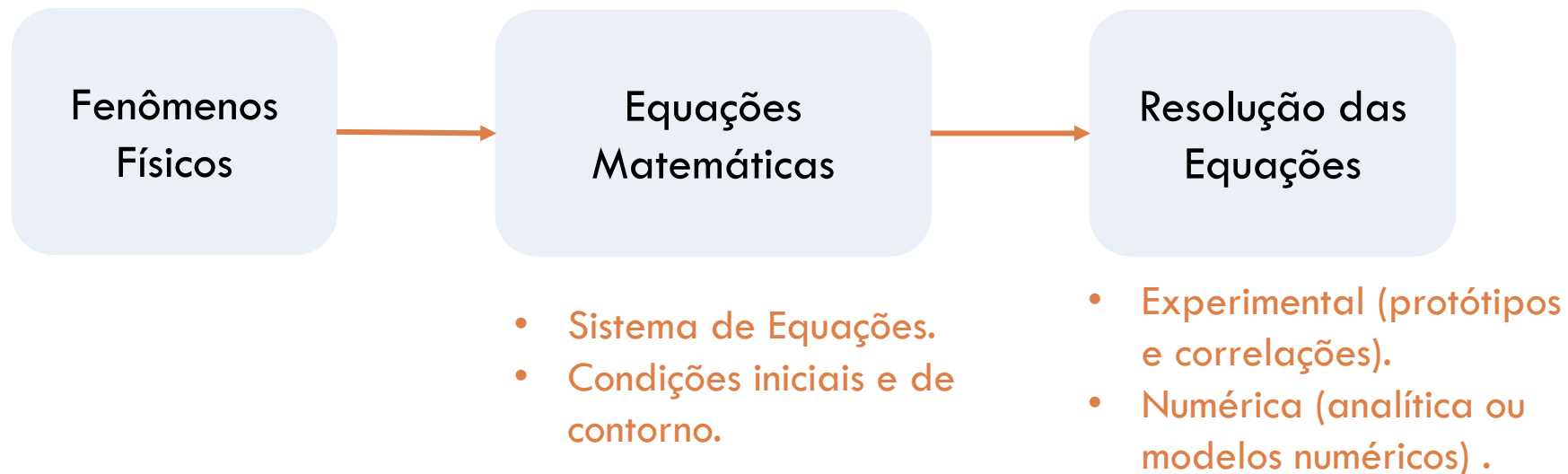


- **Modelo**: Conjunto de equações utilizadas para descrever/entender o comportamento de um sistema.
- **Equações** transformam parâmetros de **ENTRADA** conhecidos em respostas previstas de parâmetros de **SAÍDA**.
- **Aprendizado**: Identificação do melhor modelo (Ajuste de Histórico, calibração de modelos, treinamento).

# Modelos baseados na física x Modelos baseados em Dados

4

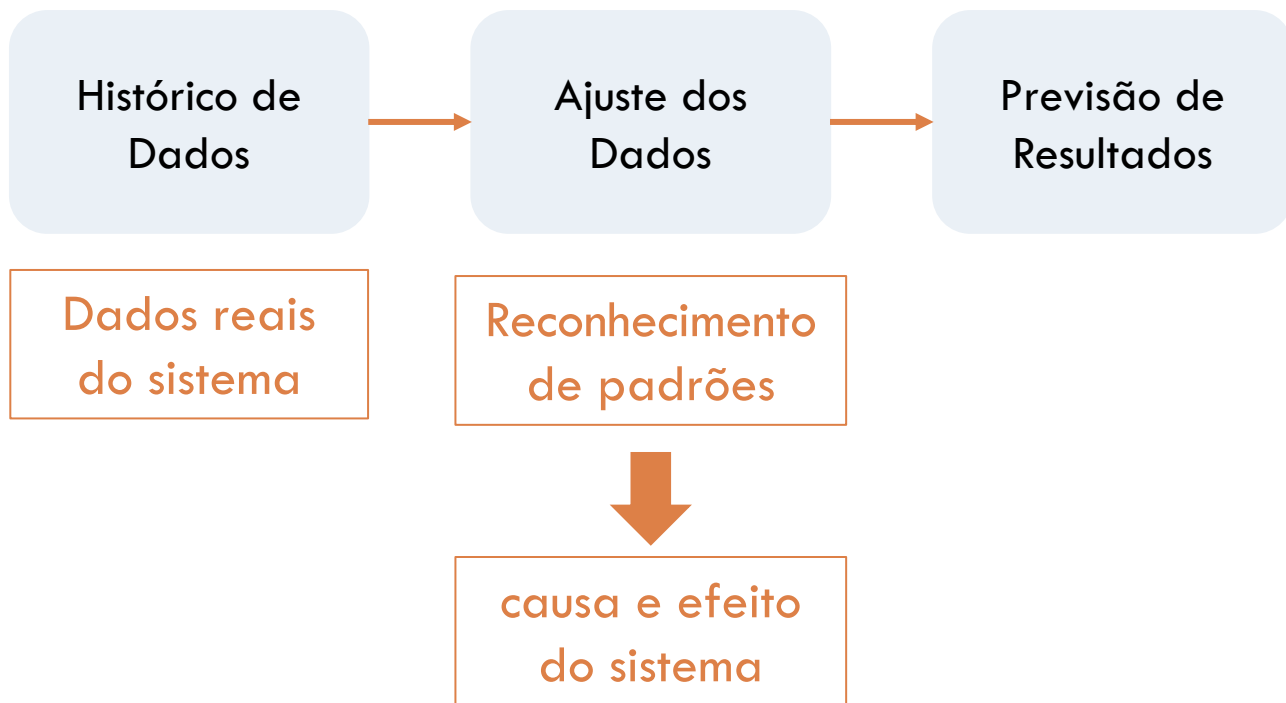
- Modelos baseados na física: Fundamentados nas leis físicas que regem o sistema.



# Modelos baseados na física x Modelos baseados em Dados

5

- Modelos baseados em Dados: Informações dos processos são derivadas diretamente dos dados, não requerem uma compreensão precisa dos processos físicos do sistema.



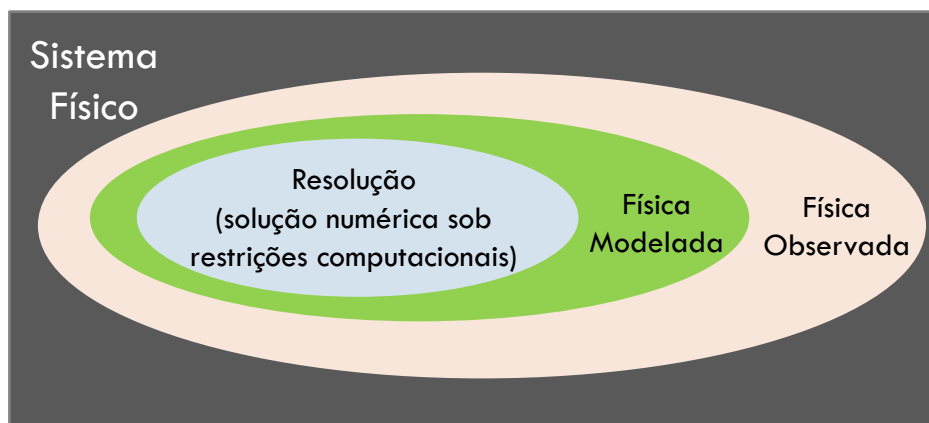
Date	Qliq	Qoil	Qwater	Qg	WC	TPT	Status
12/03/2015	2428	394	2035	232907	84	93	Yes
23/04/2015	2329	747	1582	260277	68	96	No
05/05/2015	2416	391	2025	260557	84	95	Yes
12/06/2015	2323	633	1690	370250	73	101	Yes
31/07/2015	2304	1123	1181	434015	51	109	No
08/08/2015	2351	615	1736	372962	74	102	Yes
20/08/2015	2363	321	2042	356196	86	102	No
02/09/2015	2374	311	2063	325247	87	100	No
16/10/2015	2406	344	2062	231806	86	93	Yes
16/05/2016	2418	433	1985	265612	82	97	?

- ✓ Quais variáveis se correlacionam?
- ✓ Como elas se comportam ao longo do tempo?
- ✓ Posso prever alguma dessas variáveis a partir das outras? Se sim, quais?

# Modelos baseados na física x Modelos baseados em Dados

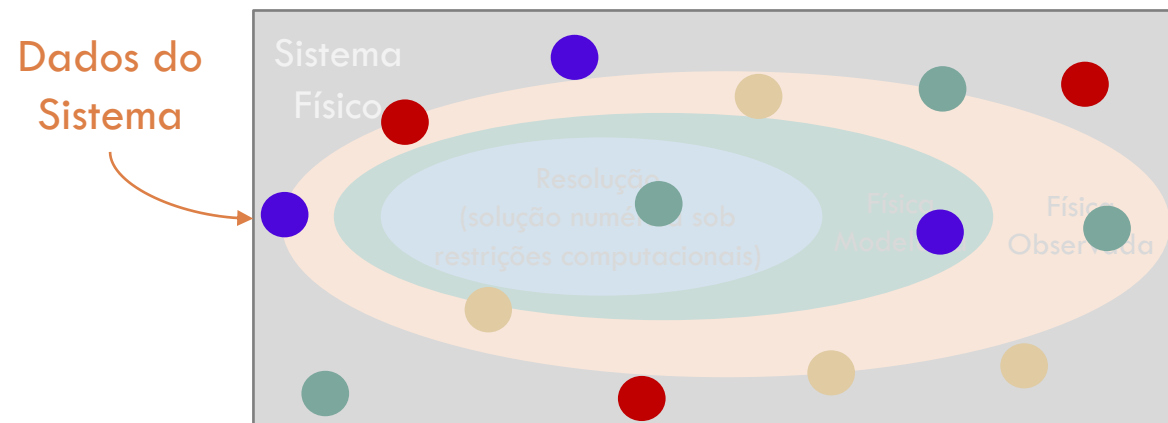
6

## □ Modelos baseados na física:



Modificado de Rasheed et al. (2019).

## □ Modelos baseados em Dados:



Como os dados são uma manifestação da física conhecida e desconhecida, ao desenvolver um modelo orientado a dados, é possível considerar a física completa.

# Modelos baseados na física x Modelos baseados em Dados

7

<b><u>Modelos baseados na Física</u></b>		<b><u>Modelos baseados em Dados</u></b>	
	Base sólida em função da física do problema.		Funciona como caixa-preta.
	Difícil de integrar dados históricos de longa data nos modelos computacionais.		Leva em consideração dados e experiências históricas de longo prazo.
	Sensível à instabilidade numérica devido a uma série de condições (iniciais, contorno, incertezas nos parâmetros de entrada).		Depois do modelo treinado, é consideravelmente estável para fazer previsões/inferências.
	Menos suscetíveis a viés.		O viés nos dados é refletido na previsão do modelo.
	Boa generalização para novos problemas com física semelhante.		Generalização ruim para problemas não vistos.

# Modelos baseados na física x Modelos baseados em Dados

8

## □ Quando usar o modelo baseados em dados ?

- Modelos físicos muito complexos;
- Grande tempo computacional dos simuladores;
- Grande quantidade de dados (“informações”) disponíveis;
- Melhoria dos modelos físicos ...

## □ Modelo de dados irá substituir os modelos físicos?

- Tendência na formação de **modelos híbridos** (dados + físicos) para resolução dos problemas de engenharia para tomada de decisão.



- **Ciência de dados** (do inglês, *data science*) é o processo que utiliza algoritmos, métodos e sistemas para extrair conhecimento e obter *insights* de dados estruturados e não estruturados. Área multidisciplinar que aplica inteligência analítica e aprendizado de máquinas para colaborar com a realização de previsões, aprimoramento da otimização, e melhoria das operações e da tomada de decisões.

# CIÊNCIA DE DADOS

10

Aumento  
considerável dos  
dados e da  
tecnologia

Novos modelos  
matemáticos e  
estatísticos

Avanços nas previsões  
em tempo real,  
Big data

ESTATÍSTICA

MINERAÇÃO  
DE DADOS

ANALÍTICA  
PREDITIVA

CIÊNCIA DE  
DADOS

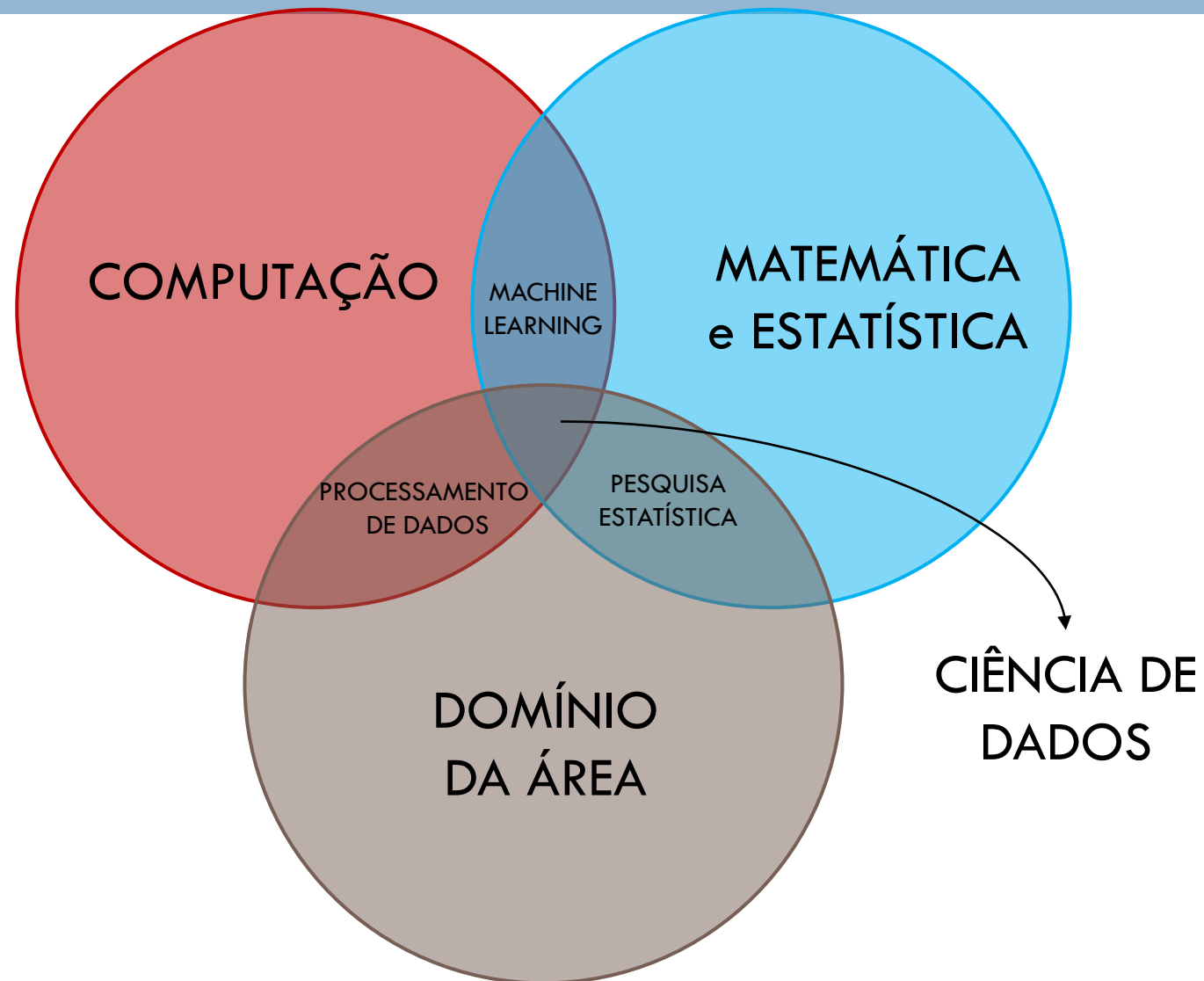
Extrair padrões  
dos dados

Previsões mais  
acuradas

- Compilação e limpeza de dados.
- Aplicação de métodos estatísticos.

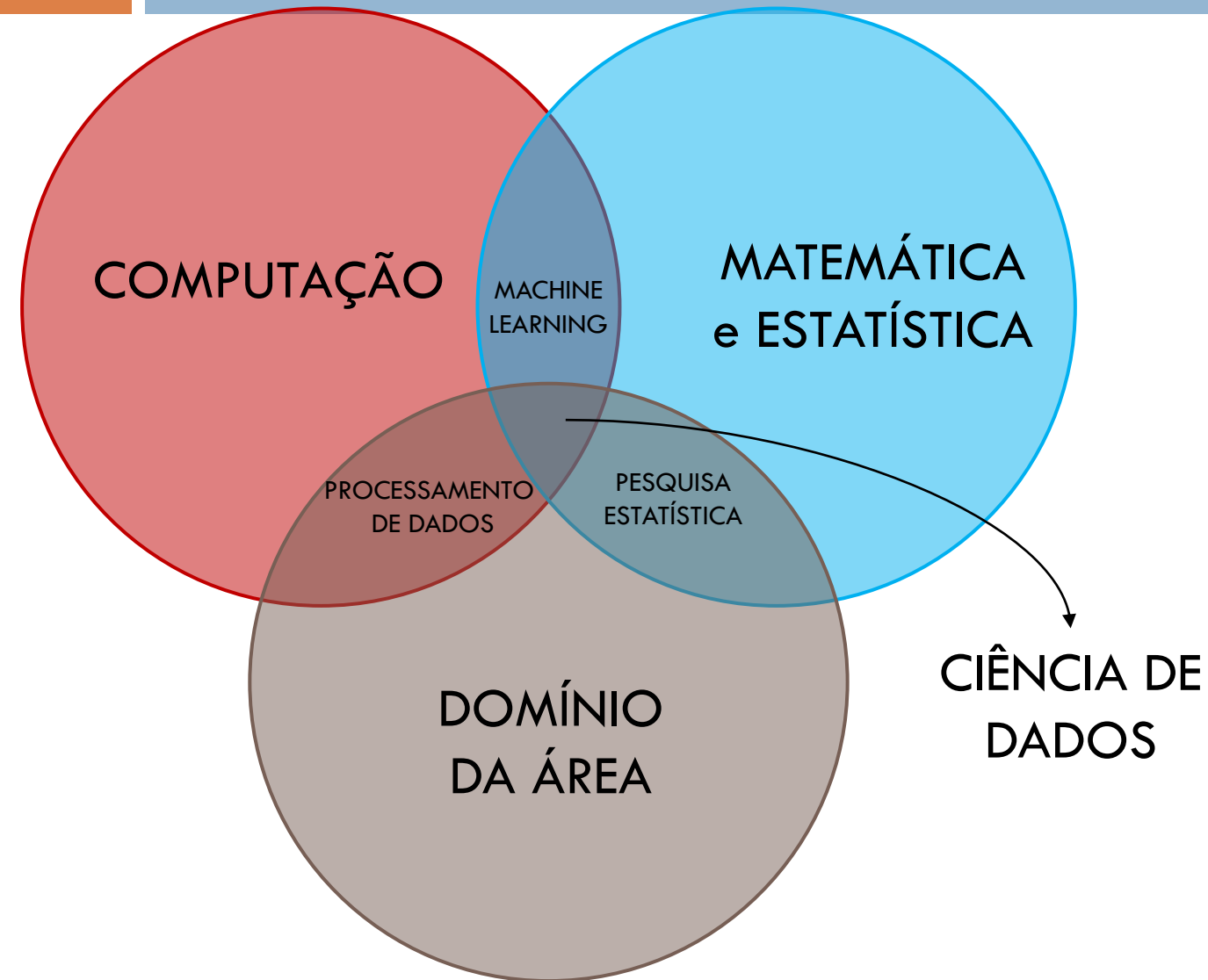
# CIÊNCIA DE DADOS

11



# CIÊNCIA DE DADOS

12



## COMPUTAÇÃO:

### ❑ Implementação

#### ▪ Linguagem de Programação

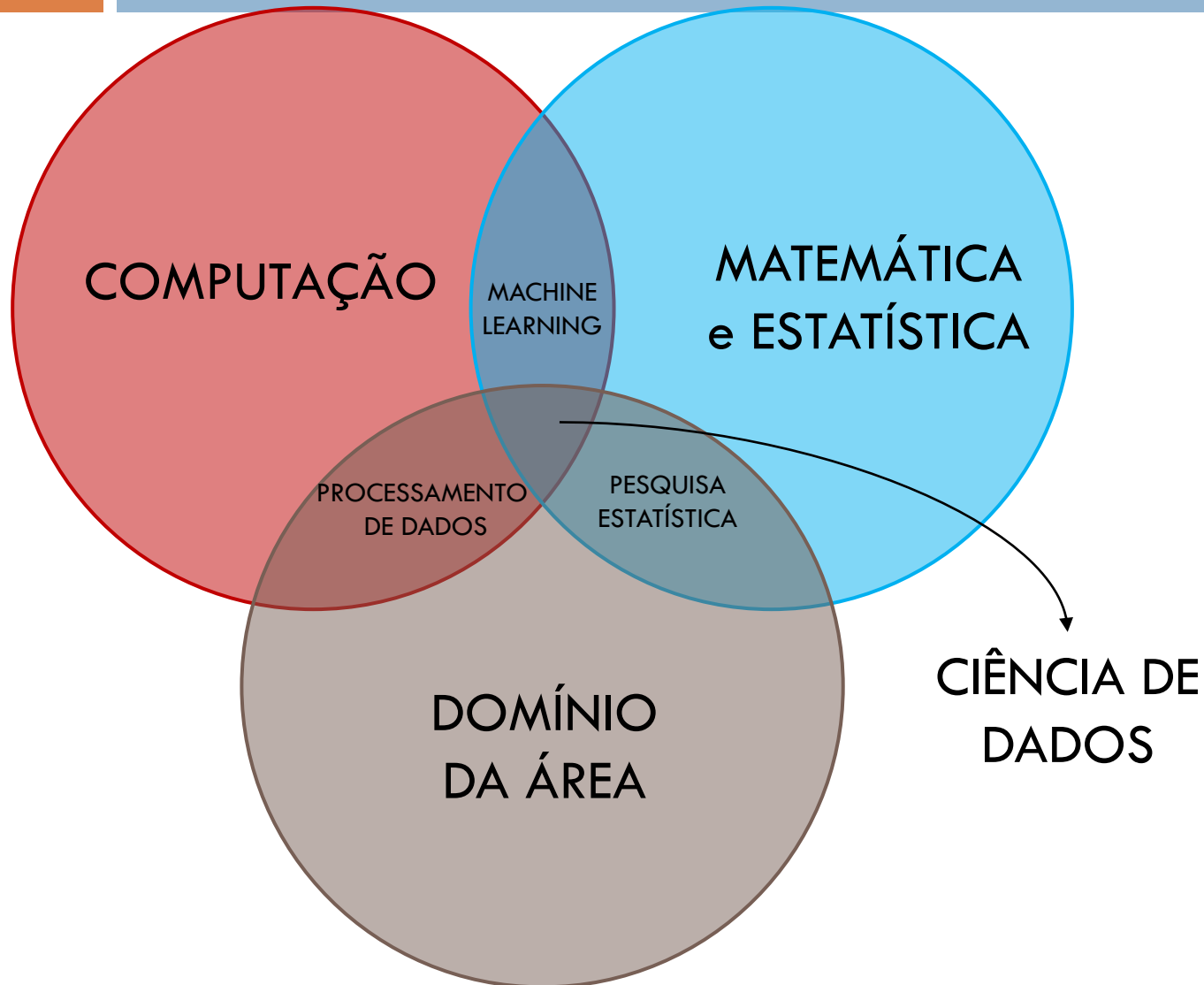
- Estatística: **Python**, R, C,...
- Base de dados: SQL.

### ❑ Importante para gerenciamento dos dados:

#### ❑ Compilação e Preparação dos dados.

# CIÊNCIA DE DADOS

13

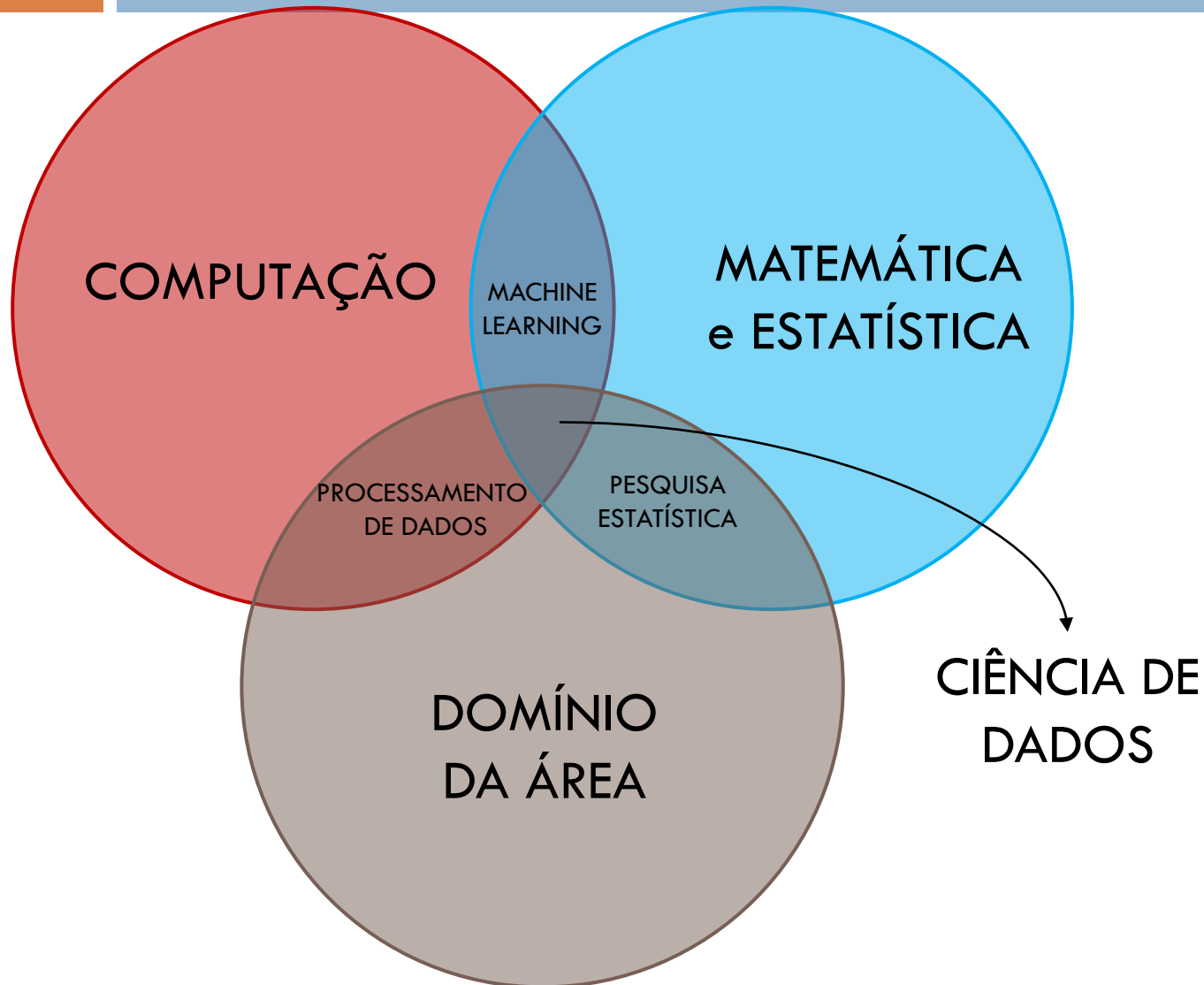


## MATEMÁTICA E ESTATÍSTICA:

- ❑ Álgebra, cálculos, probabilidade, teorema de Bayes.
- ❑ Análise Exploratória dos dados, estatística descritiva, inferência, seleção de variáveis, validação de modelos, estimadores.
- ❑ Modelos de regressão.
- ❑ Qual método usar?

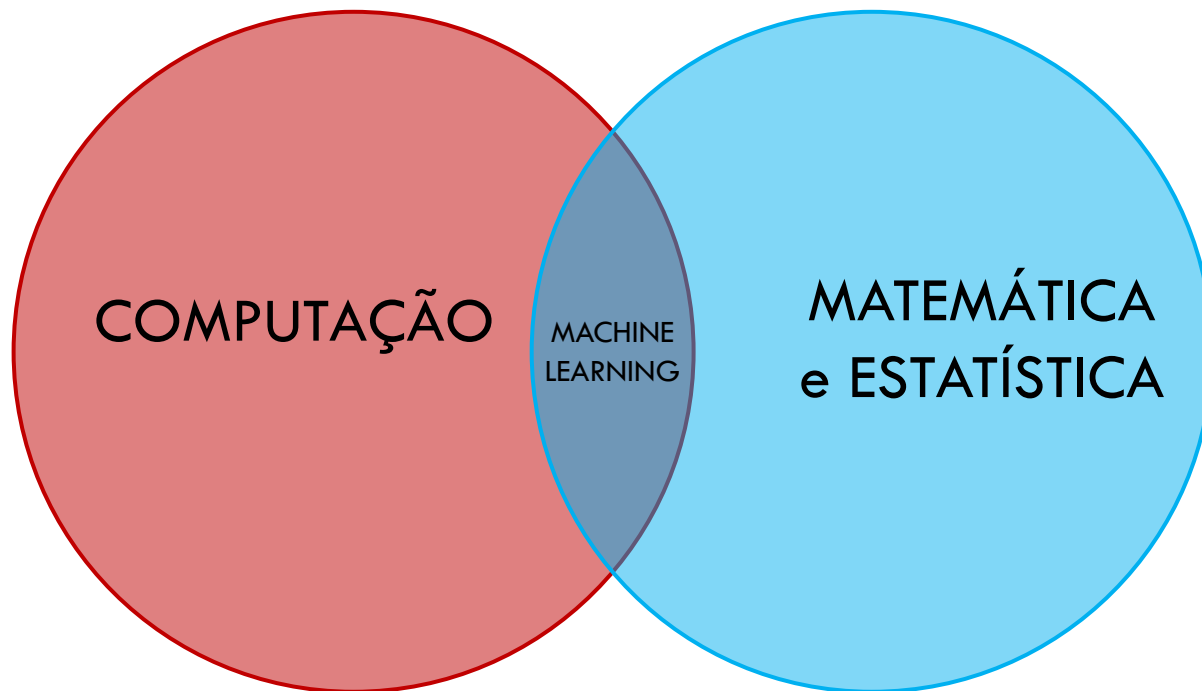
# CIÊNCIA DE DADOS

14



## DOMÍNIO DA ÁREA:

- ☐ Entender os objetivos.
- ☐ Definir as métricas.
- ☐ Fazer as perguntas corretas.
- ☐ Traduzir os resultados para a equipe sem o conhecimento técnico.



## MACHINE LEARNING:

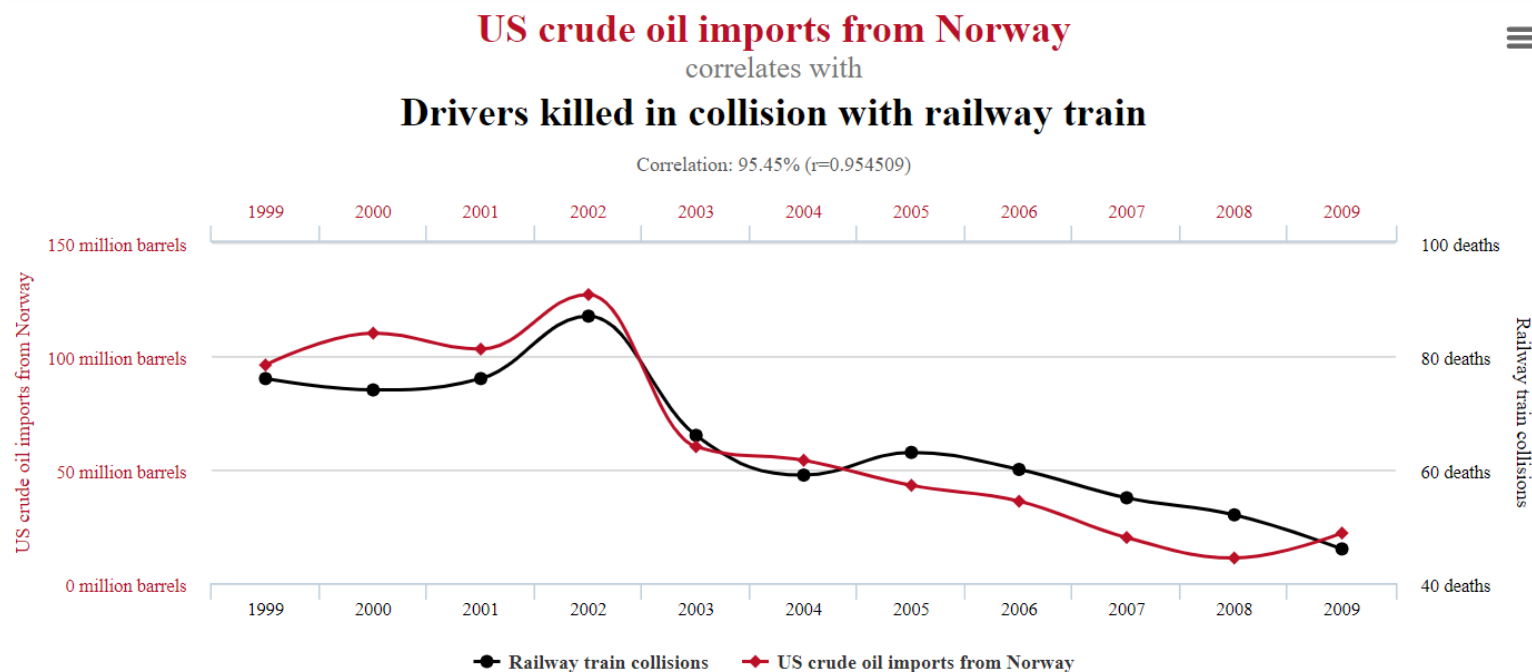
- ❑ Utiliza uma variedade de algoritmos que aprendem iterativamente a partir dos dados de treinamento para melhorar, descrever dados e prever resultados.

# CIÊNCIA DE DADOS

16

## Importância do domínio da área:

- ❑ O modelo de ML geralmente procura padrões/correlações entre variáveis. Mas devemos ter um cuidado em interpretar estas correlações, pois nem sempre elas são coerentes.



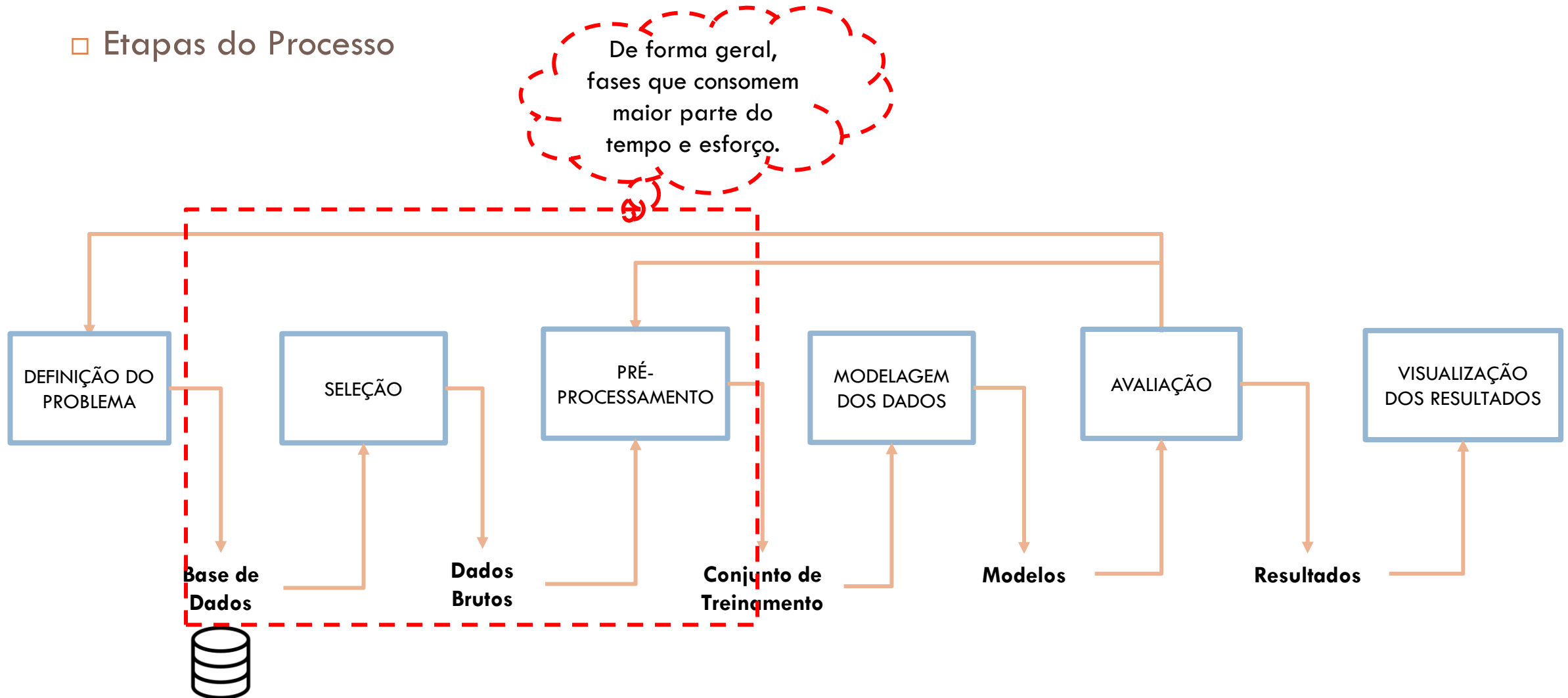
<https://www.tylervigen.com/spurious-correlations>



# CIÊNCIA DE DADOS

17

## □ Etapas do Processo

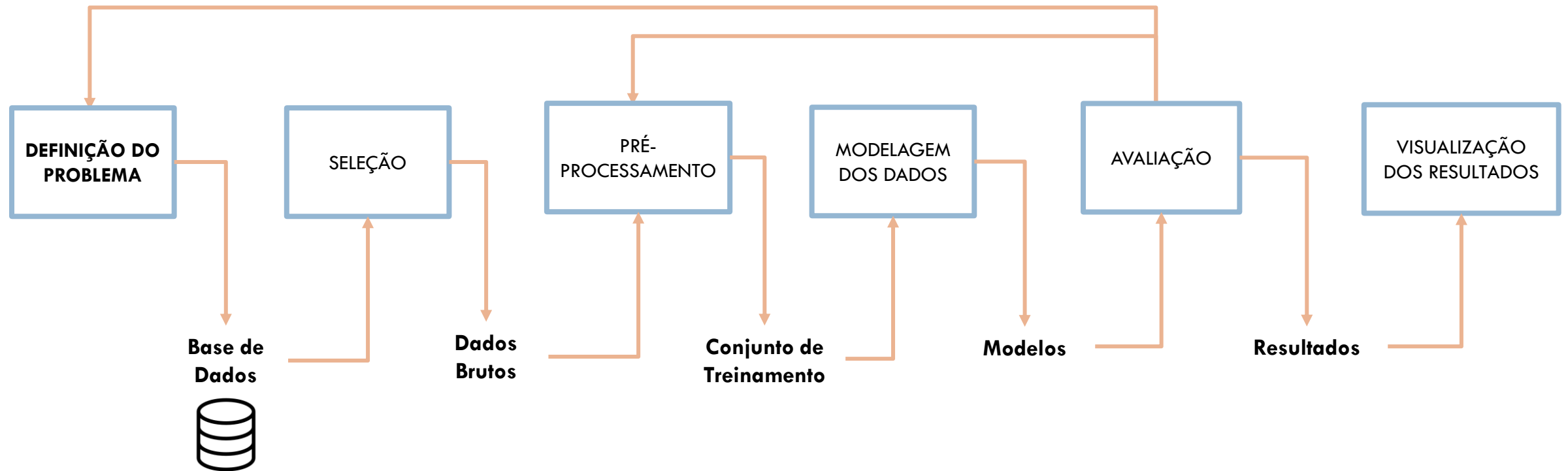


# CIÊNCIA DE DADOS

18

## □ Etapas do Processo

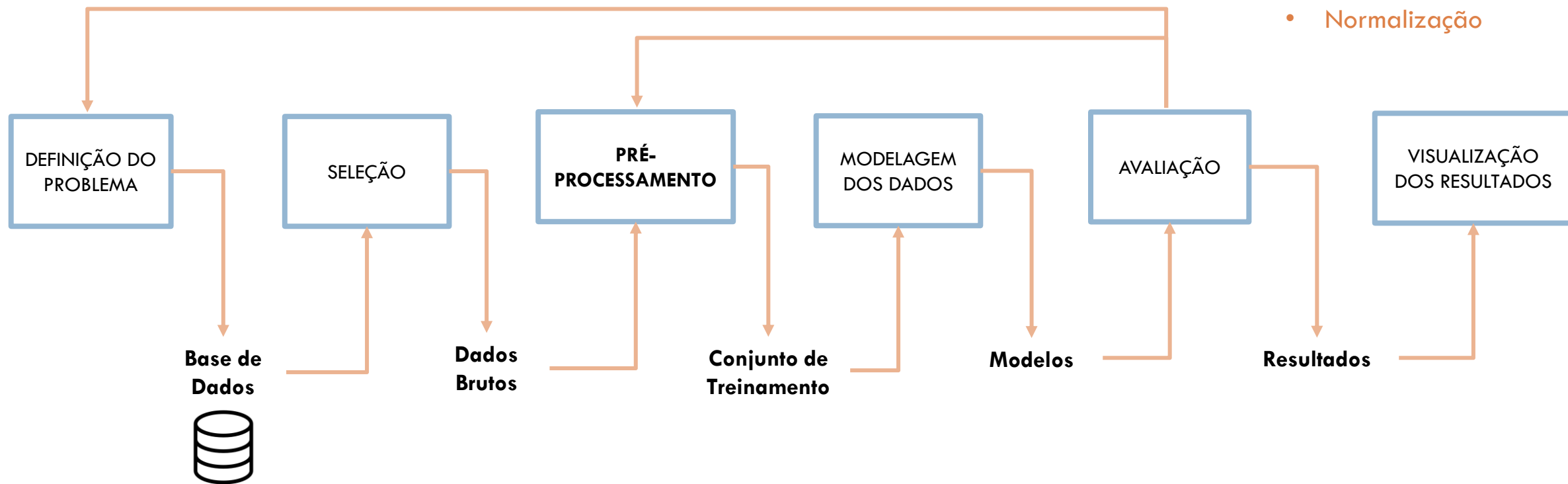
**Etapa 1 - Definição do problema:** Definir objetivo que se deseja alcançar



# CIÊNCIA DE DADOS

19

## □ Etapas do Processo



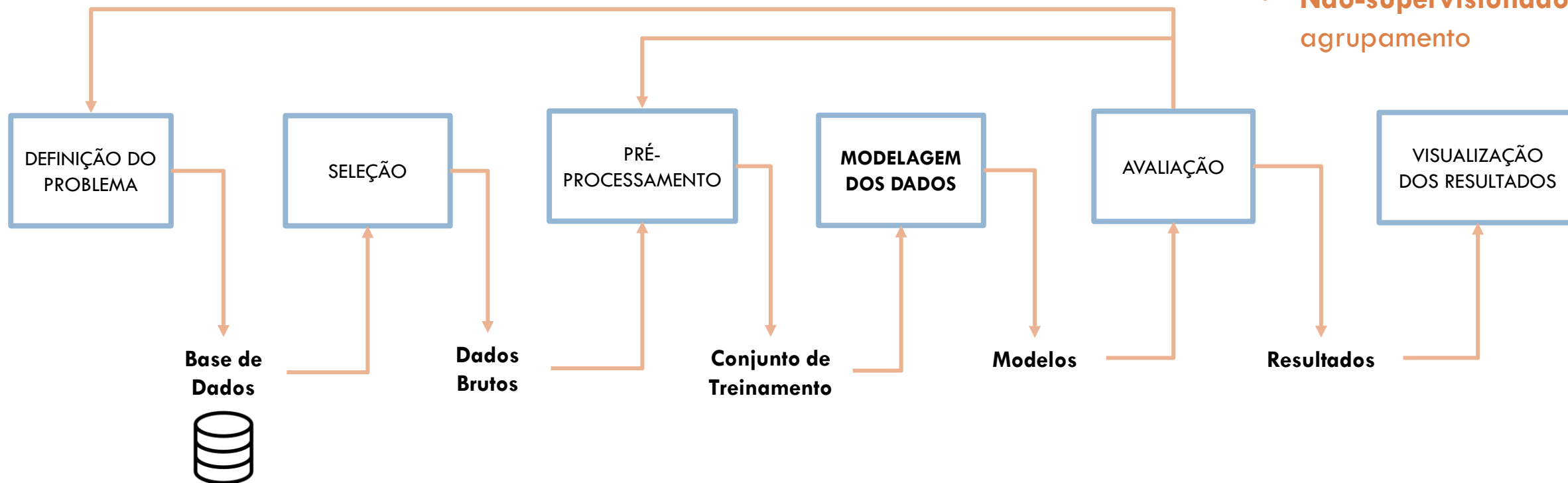
### Pré-processamento:

- Dados ausentes
- Remoção de Outliers
- Redução da dimensão
- Seleção de variáveis
- Normalização

# CIÊNCIA DE DADOS

20

## □ Etapas do Processo



### Modelagem dos dados:

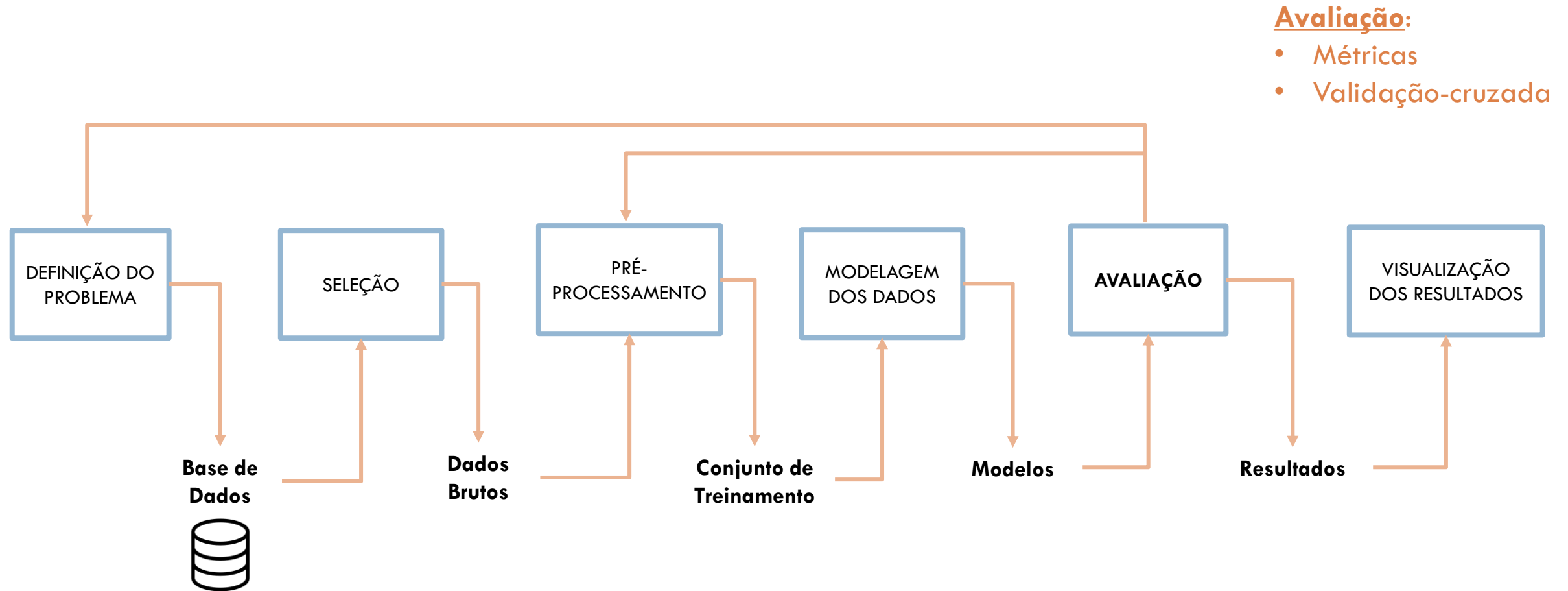
Machine Learning, deep learning

- **Supervisionados:** classificação, regressão.
- **Não-supervisionados:** agrupamento

# CIÊNCIA DE DADOS

21

## □ Etapas do Processo



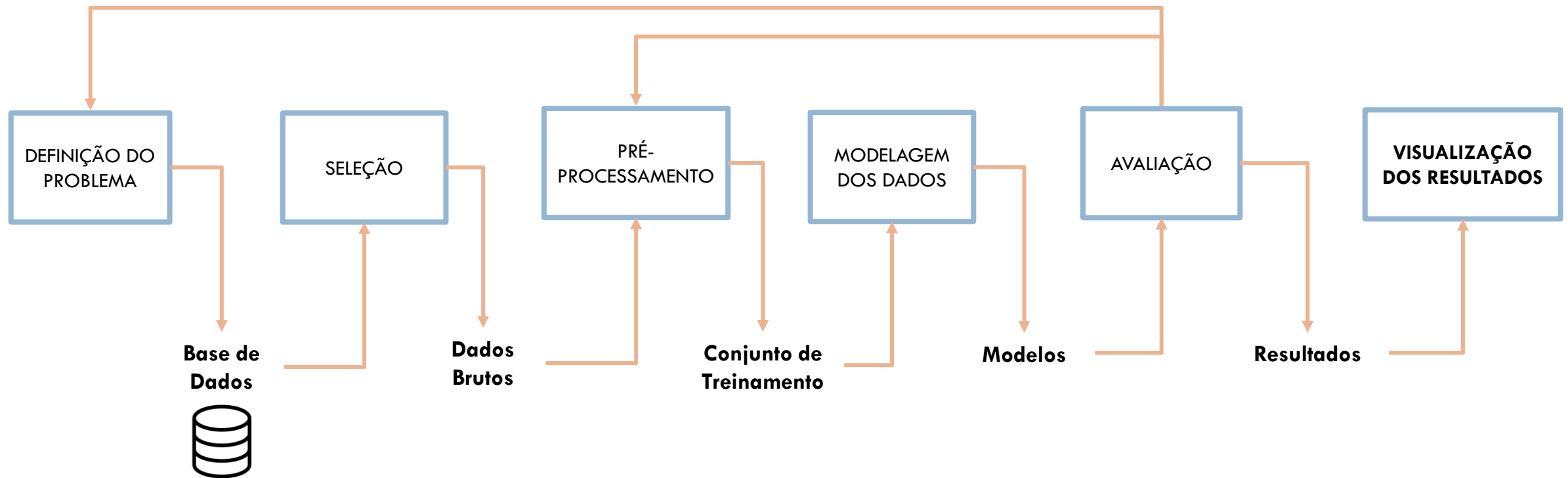
# CIÊNCIA DE DADOS

22

## □ Etapas do Processo

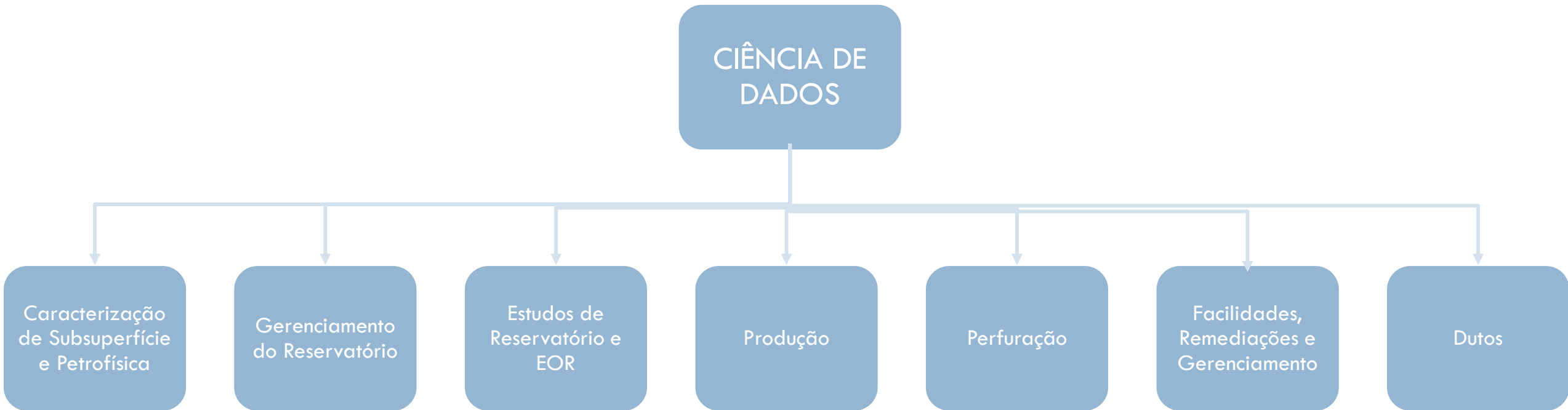
### Visualização de Resultados:

- Gráficos
- Dashboards



# Aplicações na Indústria de Petróleo

23



# Aplicações na Indústria de Petróleo

24

## □ Caracterização de Subsuperfície e Petrofísica:

TRABALHOS	MÉTODO
Caracterização do reservatório fraturado	Redes neurais e lógica Fuzzy
Previsão da porosidade e permeabilidade do campo	Support Vector Machine
Análise de permeabilidade do campo	Random Forest Tree
Modelagem das propriedades PVT do gás	SVM e árvore de decisão
Análise de logs	Inteligência artificial

## □ Gerenciamento de Reservatórios:

TRABALHOS	MÉTODO
Metodologia para caracterização do reservatório de gás natural	bootstrapping
Gerenciamento de Reservatórios Integrados	Técnicas de Machine Learning
Otimização do sistema de completação de poços	Data analytics



# Aplicações na Indústria de Petróleo

25

## □ Estudos de Reservatório e EOR

TRABALHOS	MÉTODO
Previsão do <i>holdup</i> em duas fases de reservatórios	SVM otimizando aplicando Algoritmos Genéticos
Modelos Híbridos para aprendizado próprio para melhor entendimento do reservatório	Primeiros Princípios + Redes Neurais
Modelagem de Seleção de candidatos e desempenho de poços tratados com com gel	Naive Bayes

# Aplicações na Indústria de Petróleo

26

## □ Produção:

- ▣ Muitos dados, modelagem em tempo real através de equipamentos de medição.

TRABALHOS	MÉTODO
Previsão do breakthrough da água em reservatórios fraturados	SVM
Previsão da vazão de produção de óleo	ANN, BBN*, Árvores de Decisão
Estratégias de planejamento de ordenação de testes de produção	Random Forests
Controle e automação da produção de <i>shale gas</i>	ML e ANN

\*BBN: Bayesian Belief Networks

# Aplicações na Indústria de Petróleo

27

## □ Perfuração:

### ▣ Controle de risco, controle das taxas de penetração.

TRABALHOS	MÉTODO
Simular o desempenho de diferentes tipos de fluidos de perfuração em diferentes condições do ambiente	SVM
Mapear alvos potenciais de perfuração durante a fase de exploração usando informações de deposição	SVM
Análise de risco e incertezas no gerenciamento da pressão de perfuração	BBN
Sistema de previsão de circunstâncias que levam a <i>blow-out</i>	BNN

# Aplicações na Indústria de Petróleo

28

- Facilidades, remediação e gerenciamento:
  - ▣ Melhor entendimento do desempenho de certos equipamentos da plataforma, cronogramas da plataforma e operações corretivas.

TRABALHOS	MÉTODO
Análise de risco para possíveis ataques terroristas a gasodutos	BBN
Previsão de preços de Gás a curto prazo	RNA + SVM
Mapeamento de várias estratégias para gerenciamento de plataformas offshore para reduzir riscos e evitar perdas	BBN

# Aplicações na Indústria de Petróleo

29

## □ Dutos:

- ▣ Melhor entendimento das condições dos dutos sob diferentes condições, assim como riscos envolvidos.

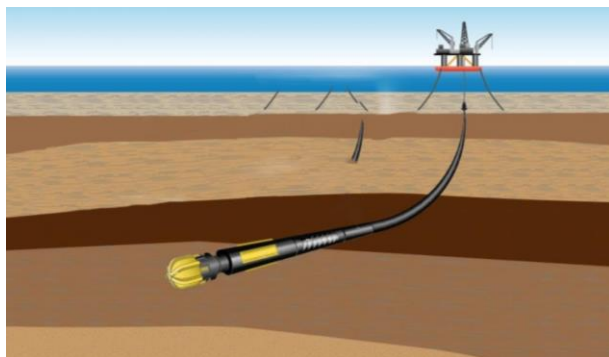
TRABALHOS	MÉTODO
Sistema de previsão de falha de duto para reduzir intervenção humana	SVM
Planejamento de inspeção ideal para problemas nos dutos	Árvores de decisão + BNN + Algoritmos Genéticos
Verificar a relação de causa-efeito para analisar o risco associado a dutos offshore	BBN

# Exemplo de Aplicação

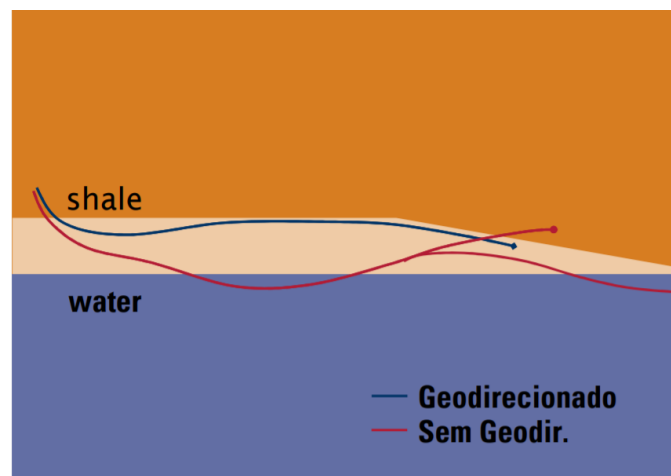
▣ Artigo: *Looking Ahead of the Bit Using Surface Drilling and Petrophysical Data: Machine-Learning-Based Real-Time*

*Geosteering in Volve Field*

Gupta et al. (2020)



<https://iogsolutions.com/index.php/blog/item/3-current-trend-in-geosteering-where-are-we-heading>



[http://www.ukbrocean.coppe.ufrj.br/Eventos\\_data/20080526/Slides.pdf](http://www.ukbrocean.coppe.ufrj.br/Eventos_data/20080526/Slides.pdf)

## Geodirecionamento

### ▣ Geodirecionamento do Poço:

- Controle da direção de um poço baseado nos resultados de medidas de perfis, para manter dentro de uma zona de interesse.

### ▣ Quais informações precisamos?

- Propriedades mecânicas das rochas envolta da broca.

# Exemplo de Aplicação:

Artigo: *Looking Ahead of the Bit Using Surface Drilling and Petrophysical Data: Machine-Learning-Based Real-Time Geosteering in Volve Field*

31

- Dados são coletados em tempo real durante a perfuração. A partir dessas medidas, é possível definir em qual tipo de formação a broca está? Caso ela passe para outra formação, é possível identificar em tempo real?

## Medidas obtidas durante a perfuração (MWD)

- Taxa de penetração
- Peso da broca,
- Torque
- Pressão e Temperatura de fundo, etc

**MACHINE  
LEARNING**

## Qual tipo de formação que se está perfurando?

- Arenito
- Xisto
- Calcário

# Exemplo de Aplicação

Artigo: *Looking Ahead of the Bit Using Surface Drilling and Petrophysical Data: Machine-Learning-Based Real-Time Geosteering in Volve Field*

32

## Etapa 1: Pré-processamento

- Análises estatísticas;
- Remoção de outliers;
- Análise de correlações de variáveis;
- PCA.

## Etapa 2: Agrupamento

- Identificar eletrofácies a partir das variáveis de RG (raio gama), DT (acústico), RHOB (densidade).
- Métodos: K-means, SOMs, clusterização hierárquica

## Etapa 3: Significância Petrofísica

- Análise dos 3 diferentes clusters encontrados em termos de mineralogia usando dados de testemunho.
  - Arenito, xisto e calcário.

## Etapa 4: Previsão da litologia

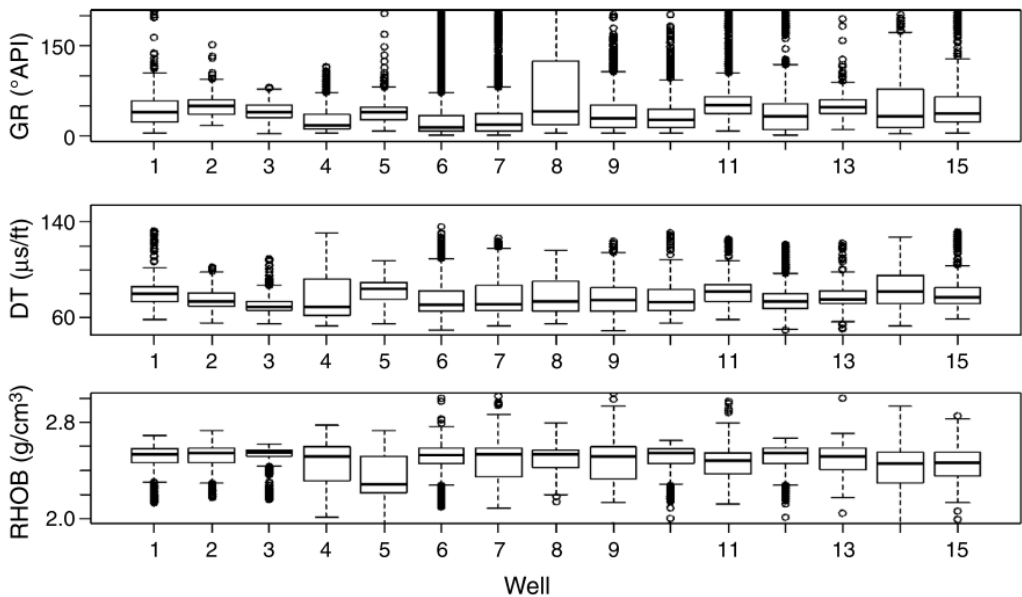
- Prever as 3 classes do agrupamento, usando ML.
  - Entradas: MWD
  - Saídas: Cluster 1,2 ou 3.



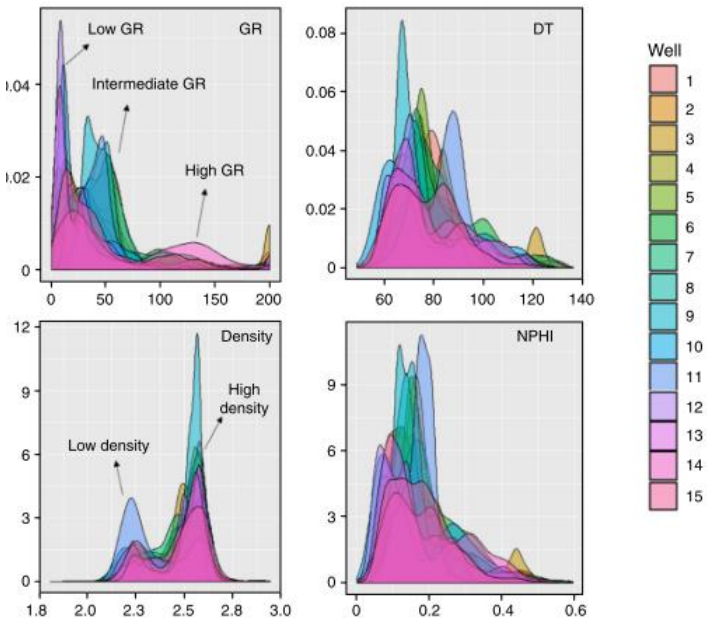
# Exemplo de Aplicação

Looking Ahead of the Bit Using Surface Drilling and Petrophysical Data: Machine-Learning-Based Real-Time Geosteering in Volve Field

## Etapa 1: Pré-processamento



Boxplot das variáveis



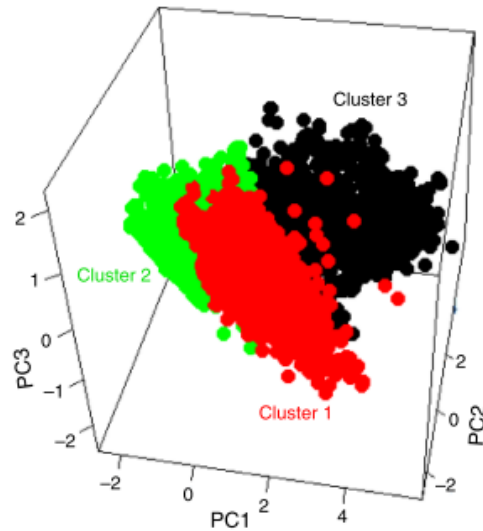
Distribuições das variáveis

Variables	Correlation Coefficient
DT/NPHI	0.93
Pressure/ECD	0.84
GR/NPHI	0.78
Temperature/pressure	0.76
Temperature/ECD	0.49
Temperature/torque	0.46
Temperature/(rev/min)	0.43
Torque/pressure	0.37
Torque/ECD	0.27
Torque/mud-flow rate	0.21

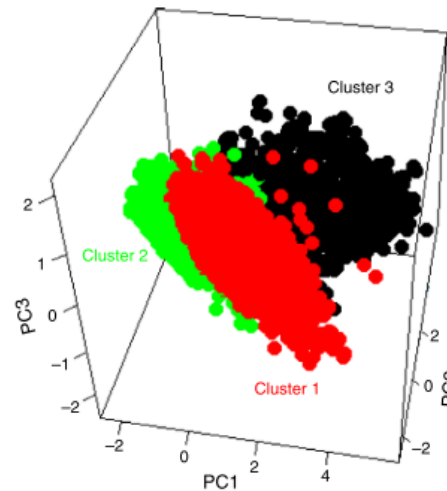
Correlação de Variáveis

# Exemplo de Aplicação

- ▣ *Looking Ahead of the Bit Using Surface Drilling and Petrophysical Data: Machine-Learning-Based Real-Time Geosteering in Volve Field*

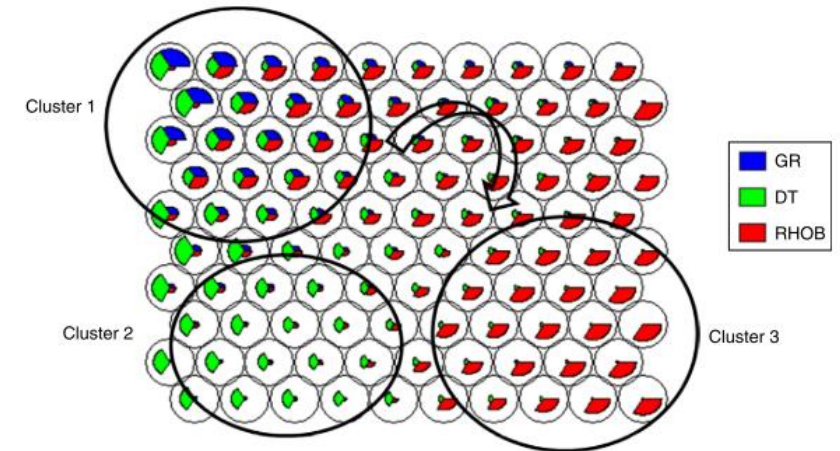


K-means



Clusterização  
Hierárquica

## Etapa 2: Agrupamento



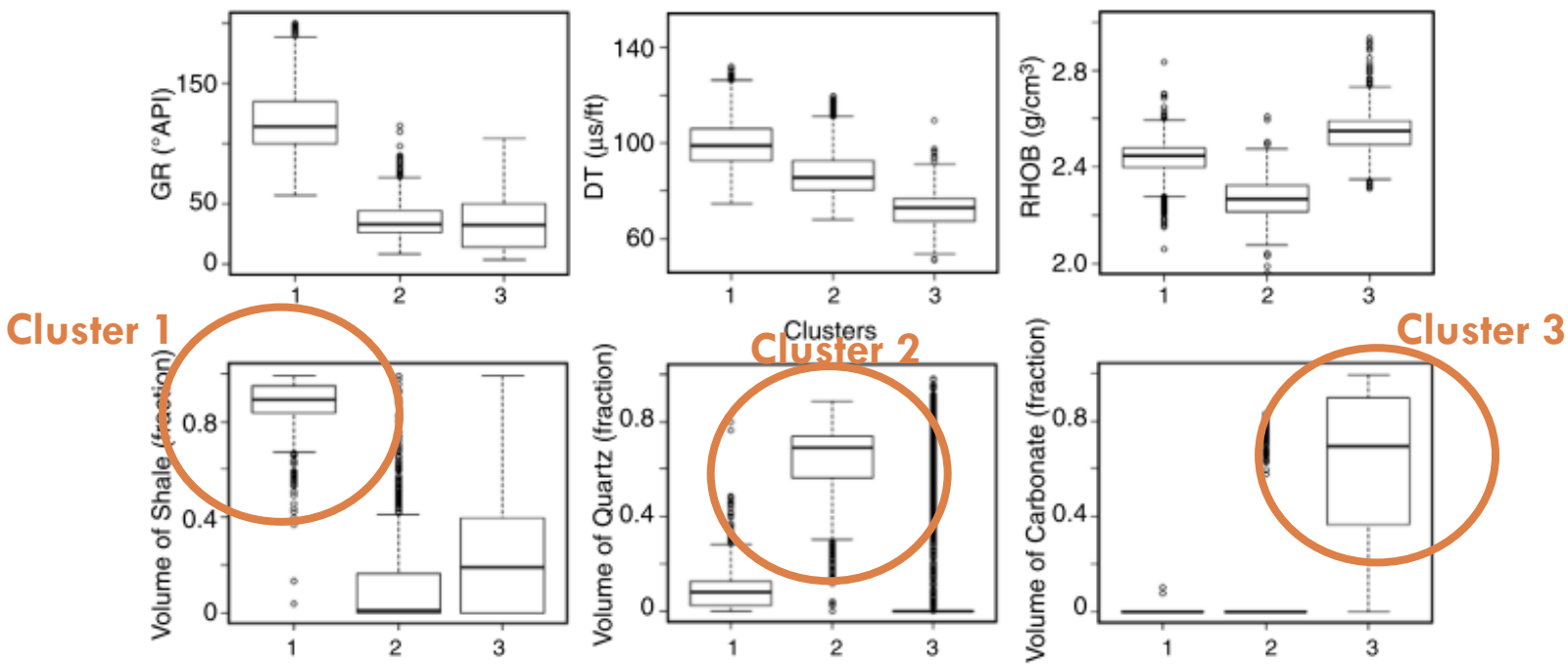
SOM (self-organizing  
maps, no português  
Mapas de Kohonen)

# Exemplo de Aplicação

Looking Ahead of the Bit Using Surface Drilling and Petrophysical Data: Machine-Learning-Based Real-Time Geosteering in Volve Field

## Etapa 3: Significância Petrofísica

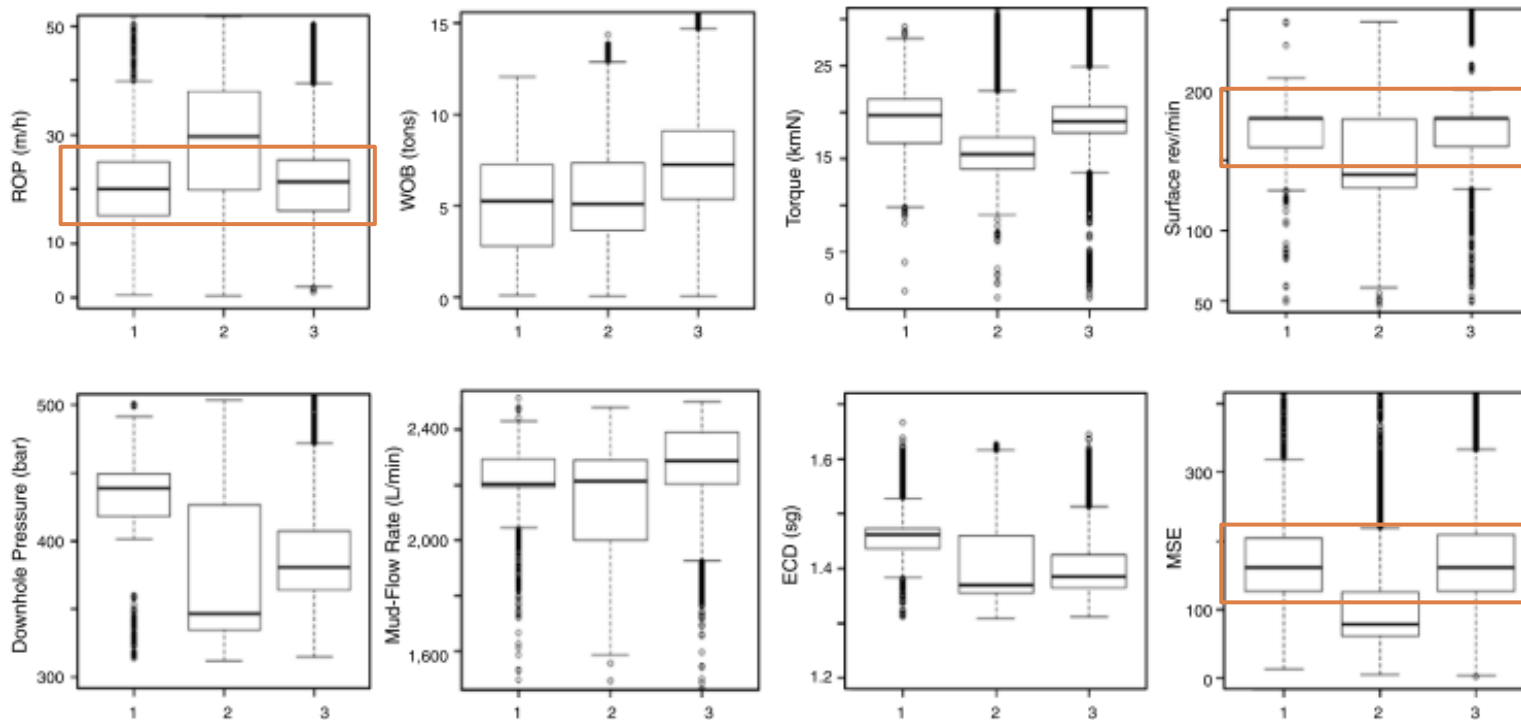
- Cluster 1: fácies rica em xisto.
- Cluster 2: fácies rica em arenito.
- Cluster 3: fácies rica em calcário/dolomita.



	Cluster 1	Cluster 2	Cluster 3
GR (°API)	144	38	23
DT (μs/ft)	102	93	69
RHOB (g/cm³)	2.4	2.2	2.6

# Exemplo de Aplicação

- Looking Ahead of the Bit Using Surface Drilling and Petrophysical Data: Machine-Learning-Based Real-Time Geosteering in Volve Field



Variáveis de Perfuração em cada cluster

## Etapa 4: Previsão da litologia

- Variáveis muito parecidas entre clusters. Necessário usar outras técnicas para automaticamente identificar em qual cluster está, a partir das variáveis de entrada.

# Exemplo de Aplicação

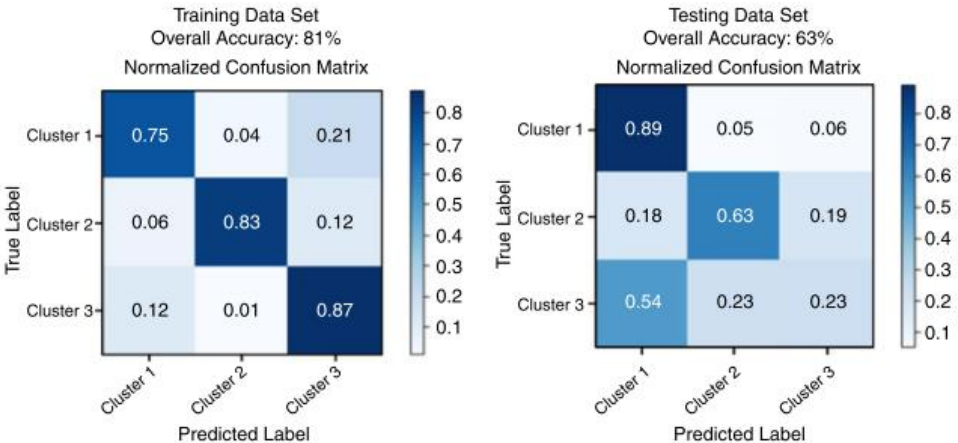
▣ *Looking Ahead of the Bit Using Surface Drilling and Petrophysical Data: Machine-Learning-Based Real-Time Geosteering in Volve Field*

## Etapa 4: Previsão da litologia

Technique	Training/Testing	Accuracy	AUC	Kappa	Log Loss
DTs	Training	0.90	0.93	0.79	0.32
	Testing	0.80	0.73	0.51	0.70
Gradient boosting	Training	0.94	0.98	0.89	0.40
	Testing	0.77	0.83	0.49	1.93
Random forest	Training	1.00	1.00	1.00	0.019
	Testing	0.75	0.86	0.46	0.60

Table 4—Accuracy metrics for the different tree-based regression techniques. AUC = area under the ROC curve.

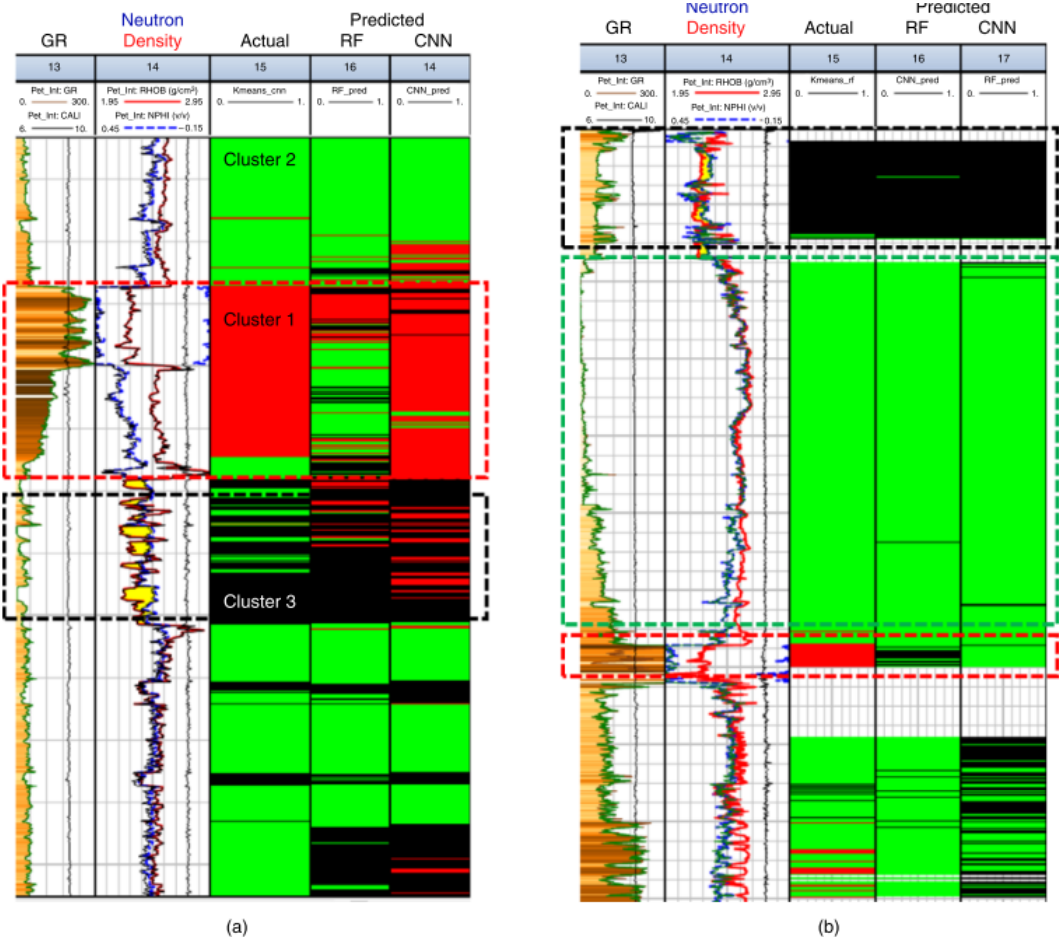
### Métricas de avaliação



### Matriz de confusão

# Exemplo de Aplicação

Looking Ahead of the Bit Using Surface Drilling and Petrophysical Data: Machine-Learning-Based Real-Time Geosteering in Volve Field



Etapa 4: Previsão da litologia

# Sites Indicados

39

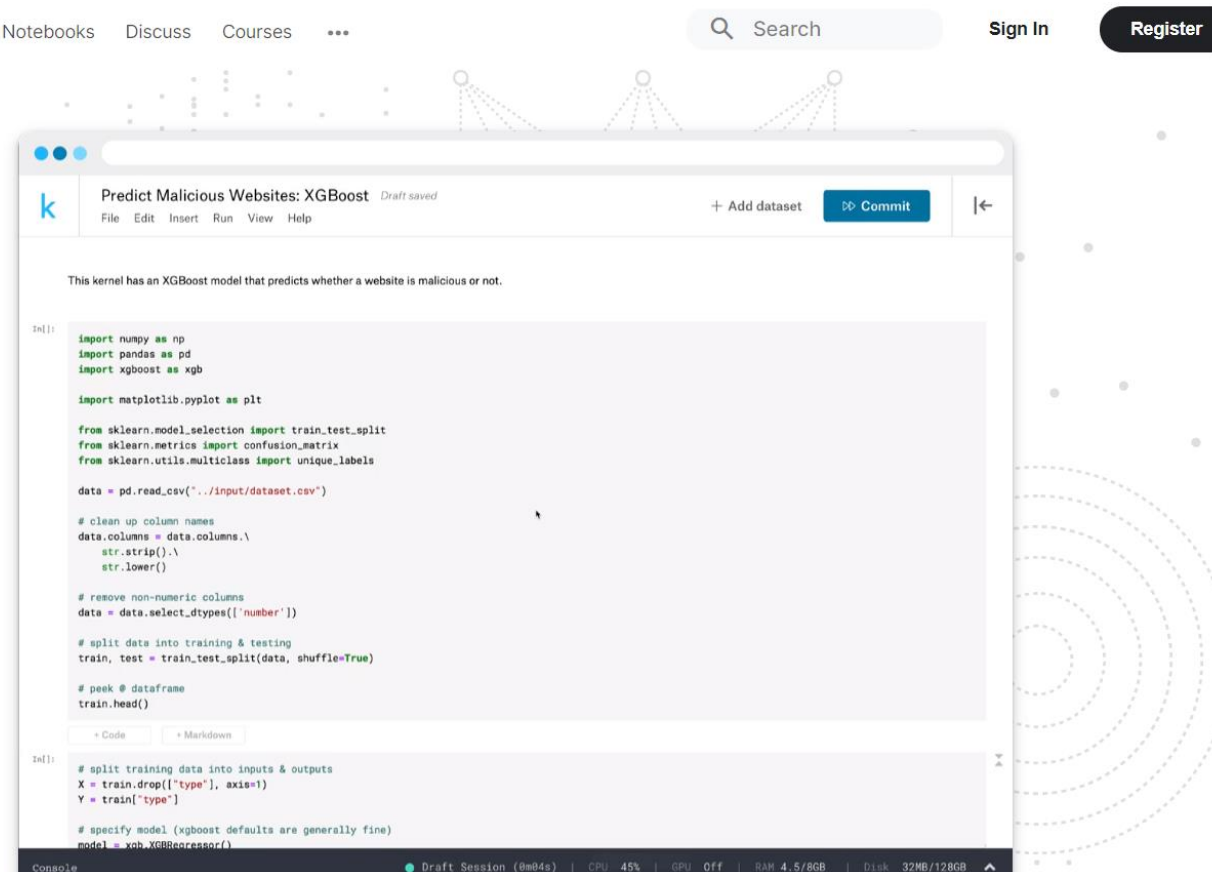
## ▣ Kaggle

### Start with more than a blinking cursor

Kaggle offers a no-setup, customizable, Jupyter Notebooks environment. Access free GPUs and a huge repository of community published data & code.

 REGISTER WITH GOOGLE

Register with Email



<https://www.kaggle.com/>



# Sites Indicados

40

## ▣ KDnuggets



Free Online Statistics Course. Build practical skills in using data to solve problems

Topics: [Coronavirus](#) | [AI](#) | [Data Science](#) | [Deep Learning](#) | [Machine Learning](#) | [Python](#) | [R](#) | [Statistics](#)



Data Science Salon, Online, Sep 22-25. Save 20% w. code KDnuggets



SAS Viya



<https://www.kdnuggets.com/>

Subscribe to KDnuggets News





# Sites Indicados

41

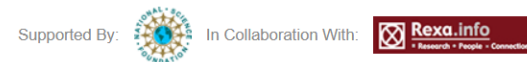
## ■ UCI

<https://archive.ics.uci.edu/ml/index.php>



### Welcome to the UC Irvine Machine Learning Repository!

We currently maintain 557 data sets as a service to the machine learning community. You may [view all data sets](#) through our searchable interface. For a general overview of the Repository, please visit our [About page](#). For information about sets in publications, please read our [citation policy](#). If you wish to donate a data set, please consult our [donation policy](#). For any other questions, feel free to [contact the Repository librarians](#).



#### Latest News:






**09-24-2018:** Welcome to the new Repository admins Dheeru Dua and Efi Karra Taniskidou!  
**04-04-2013:** Welcome to the new Repository admins Kevin Bache and Moshe Lichman!  
**03-01-2010:** [Note](#) from donor regarding Netflix data  
**10-16-2009:** Two new data sets have been added.  
**09-14-2009:** Several data sets have been added.  
**03-24-2008:** New data sets have been added!  
**06-25-2007:** Two new data sets have been added: UJI Pen Characters, MAGIC Gamma Telescope

#### Featured Data Set: [Wine](#)



**Task:** Classification  
**Data Type:** Multivariate  
**# Attributes:** 13  
**# Instances:** 178

#### Newest Data Sets:

**07-22-2020:**  [Facebook Large Page-Page Network](#)  
**07-17-2020:**  [Amphibians](#)  
**07-12-2020:**  [Early stage diabetes risk prediction dataset](#)  
**06-28-2020:**  [Taiwanese Bankruptcy Prediction](#)  
**06-20-2020:**  [South German Credit \(UPDATE\)](#)  
**06-17-2020:**  [BitcoinHeistRansomwareAddressDataset](#)  
**06-16-2020:**  [Crop mapping using fused optical-radar data set](#)

#### Most Popular Data Sets (hits since 2007):

**3489961:**  [Iris](#)  
**1902990:**  [Adult](#)  
**1469341:**  [Wine](#)  
**1310621:**  [Breast Cancer Wisconsin \(Diagnostic\)](#)  
**1291556:**  [Heart Disease](#)  
**1286391:**  [Wine Quality](#)  
**1258645:**  [Bank Marketing](#)

# Canais Youtube

42

- ▣ LORDE-COPPE / UFRJ:

- <https://www.youtube.com/channel/UCjc2gQvYyYG17T7syO66JgA>

- ▣ Programação Dinâmica:

- <https://www.youtube.com/c/Programa%C3%A7%C3%A3oDin%C3%A2mica>

- ▣ StatQuest with Josh Starmer:

- <https://www.youtube.com/channel/UCtYLUtHgS3k1Fg4y5tAhLbw>

- ▣ Peixe Babel:

- <https://www.youtube.com/c/PeixeBabel/videos>

# Referências Bibliográficas

- BALAJI, K. et al. Status of Data-Driven Methods and their Applications in Oil and Gas Industry. 2018.
- Gupta, I. et al. (2020). Looking Ahead of the Bit Using Surface Drilling and Petrophysical Data: Machine-Learning-Based Real-Time Geosteering in Volve Field. Society of Petroleum Engineers. doi:10.2118/199882-PA
- Hodaway, K R. **Harness Oil and Gas Big Data with Analytics**. 2014.
- Rasheed A, San O, Kvamsdal T. Digital Twin: Values, Challenges and Enablers. 2019. Disponível em: <https://arxiv.org/abs/1910.01719>.
- Sankaran S, Matringe S, Sidahmed M, et al. Data Analytics in Reservoir Engineering. SPE Technical Report. 2019.