



AULA 1 - INTRODUÇÃO A CIÊNCIA DE DADOS

#BootcampMIA2022 #SomosMIA



Quem somos?



Bárbara Barbosa

Data Manager na Orderchamp

- Mestra em Sistemas de Informação - com foco em Inteligência Artificial e NLP
- Organizadora do Rails Girls SP, Women Dev Summit e Women in Data Science SP 2019/2020/2021



bahbbc



bahbbc



Quem somos?



Carol Oliveira

Data Scientist na NTT DATA

- Bacharel em Engenharia da Computação na Universidade Veiga de Almeida (UVA)
- Embaixadora do Women in Data Science RJ e co-fundadora do Mulheres em IA



Coliverfelt



oliverfelt



Uma inspiração...

"Nunca estou realmente satisfeita quanto a entender alguma coisa; porque, até onde entendo, a minha compreensão só pode ser uma fração infinitesimal de tudo o que eu quero compreender."

(Ada Lovelace)



Sumário

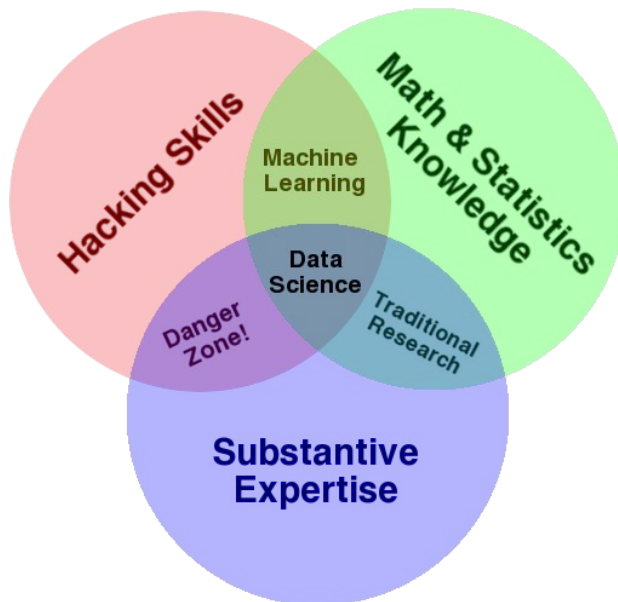
- O que é Ciência de Dados?
- Os papéis de Dados
- Etapas de um projeto
- Problemas comuns da área
- Machine Learning
- Tecnologias
- Minha história em Data Science
- Dinâmica

O que é Ciência de Dados?



O que é Ciência de Dados?

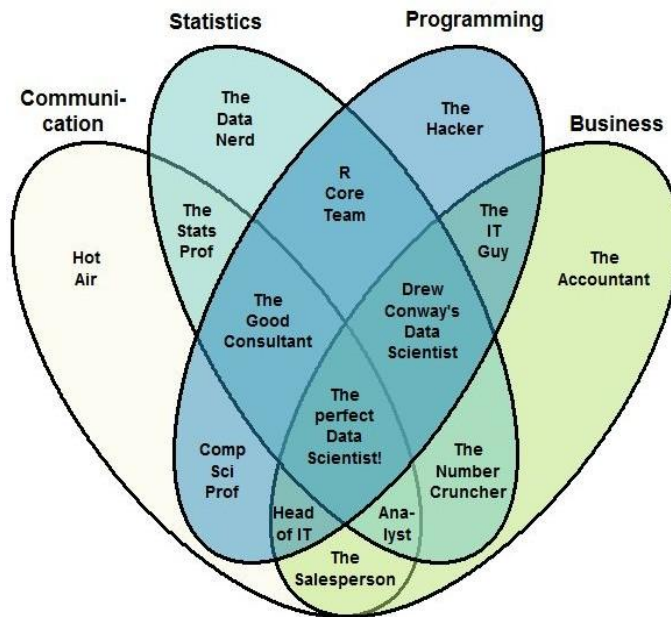
- **Combinação de conhecimentos e práticas para gerar insights a partir de previsões ou classificações**





O que é Data Science?

The Data Scientist Venn Diagram

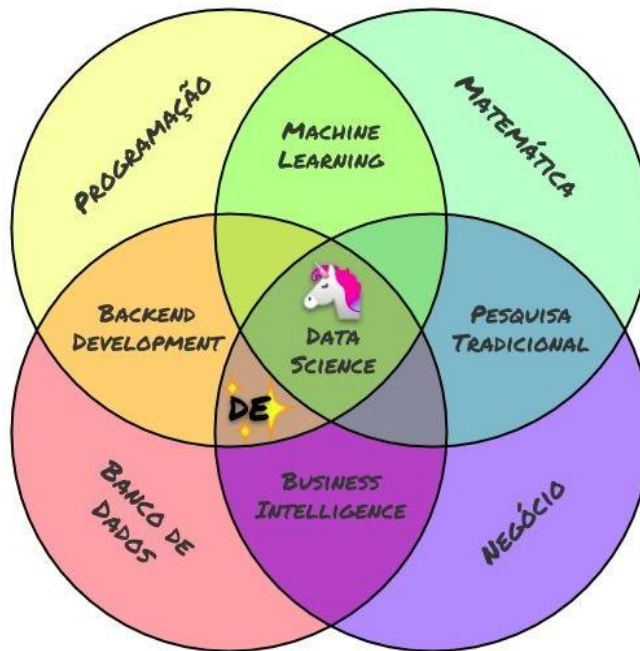


Os papéis de Dados



Os papéis de dados

- **Data Scientist**
- Data Engineering
- Data Analyst
- Machine Learning Engineering
- Chief Data Officer



Unicórnio de Ciência de Dados:
fez uma análise e descobriu



que não existe.

Etapas de um projeto



Etapas de um projeto

- **Identificação do problema**





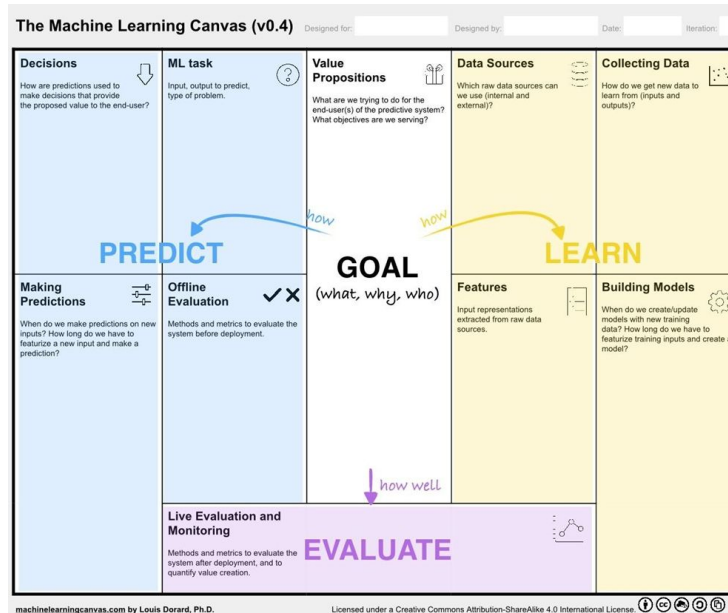
Etapas de um projeto

- **Identificação do problema**
- Diferenciar bons e maus pagadores
- Identificar clientes que podem abandonar a operação (churn)
- Identificar objetos



Etapas de um projeto

■ Identificação do problema





Etapas de um projeto

- **Coleta, preparação e análise**





Etapas de um projeto

■ Coleta, preparação e análise

■ Mergulhe nos dados...

- SQL
- Json
- XML
- Imagem
- Etc.





Etapas de um projeto

■ Coleta, preparação e análise

■ Busque ajuda...

- Data Engineer
- BI
- Produto
- DEVs





Etapas de um projeto

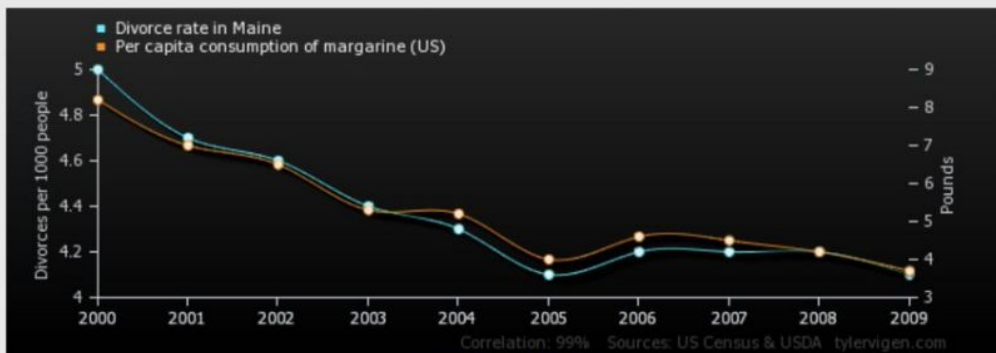
- **Coleta, preparação e análise**
- Análise exploratória...





Etapas de um projeto

Divorce rate in Maine correlates with Per capita consumption of margarine (US)



[Upload this chart to imgur](#)

	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009
Divorce rate in Maine Divorces per 1000 people (US Census)	5	4.7	4.6	4.4	4.3	4.1	4.2	4.2	4.2	4.1
Per capita consumption of margarine (US) Pounds (USDA)	8.2	7	6.5	5.3	5.2	4	4.6	4.5	4.2	3.7
Correlation: 0.992558										



Etapas de um projeto

- **Construção do modelo**





Etapas de um projeto

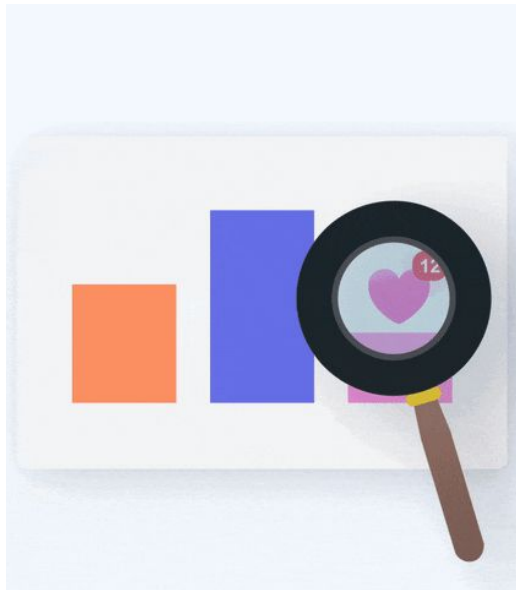
- **Construção do modelo**
- Feature engineering
- Modelagem dos dados





Etapas de um projeto

- **Avaliação do modelo**





Etapas de um projeto

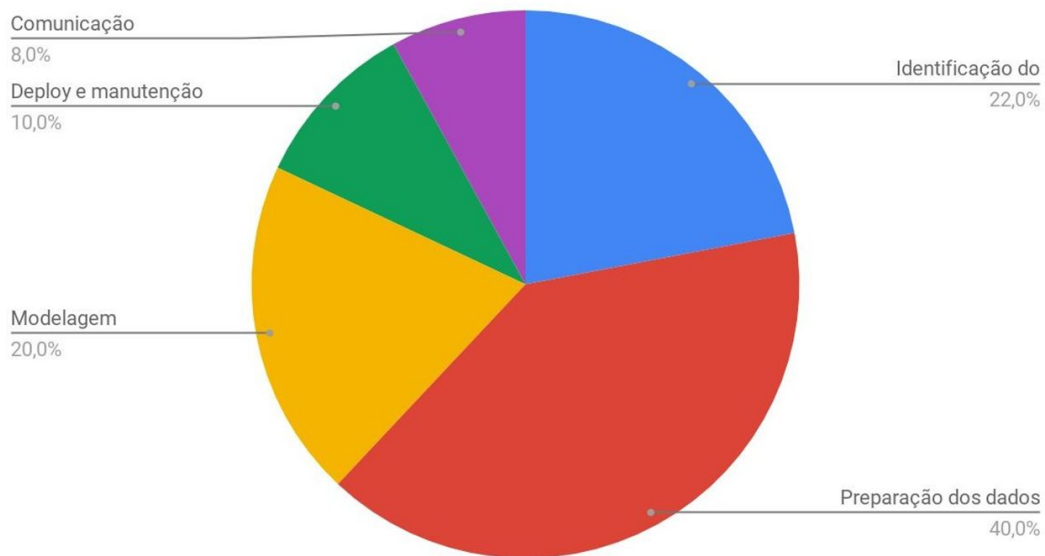
- **Deploy do modelo**





Etapas de um projeto

Tempos de um projeto



Problemas comuns Da área



Problemas comuns da área

■ Organização

■ Falta de alinhamento:

- Data Science
- Negócio
- TI

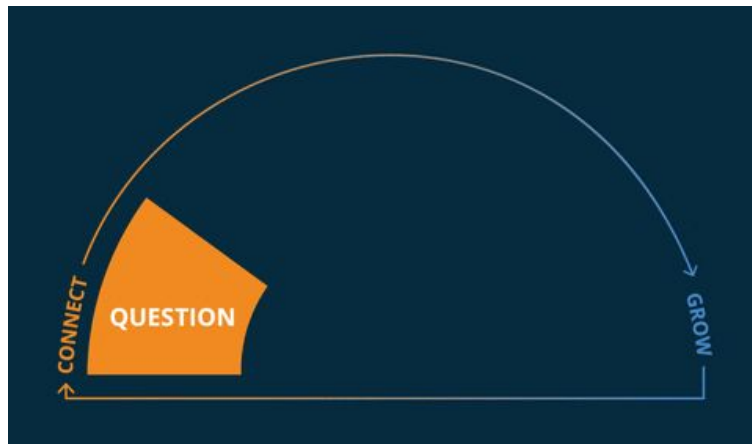




Problemas comuns da área

■ Processo

- **Falta de coordenação:**
- Desenvolvimento
- Deploy
- Manutenção dos modelos





Problemas comuns da área

■ Tecnologia

- Infraestrutura
- Sistemas incompatíveis
- Problemas de acesso a dados



Machine Learning



Por que queremos que as máquinas aprendam?





Os componentes do aprendizado de máquina

- Dados





Os componentes do aprendizado de máquina

■ Atributos





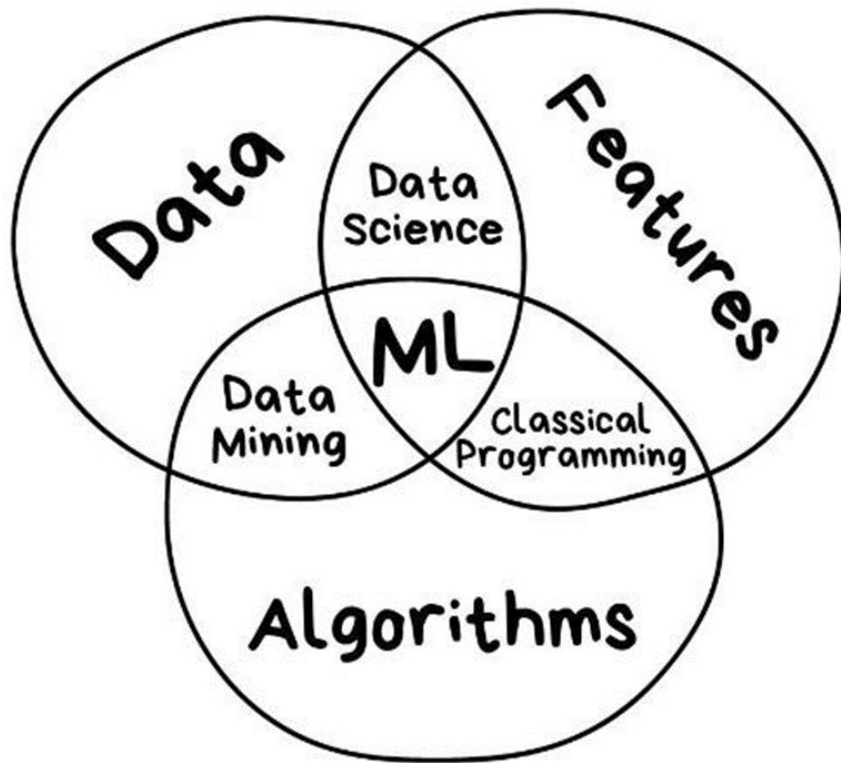
Os componentes do aprendizado de máquina

■ Algoritmos





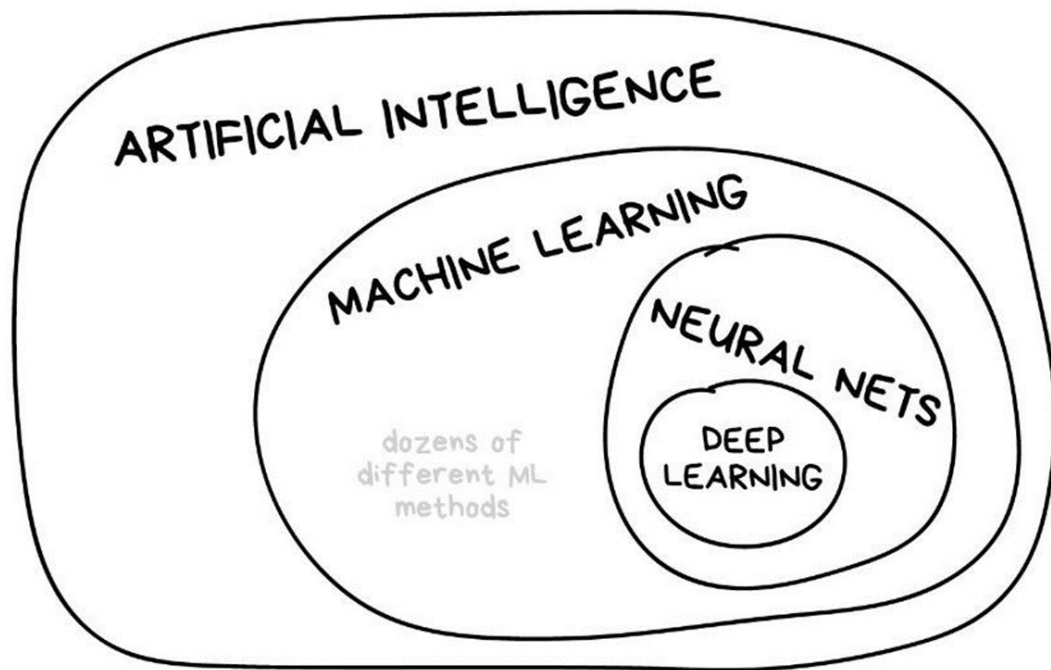
Os componentes do aprendizado de máquina





Os componentes do aprendizado de máquina

■ Diagrama IA vs ML



Tecnologias



Tecnologias

■ Big Data



Minha história em Data Science



Dúvidas?

- **Bárbara**
 - **LinkedIn:** [linkedin.com/in/bahbbbc](https://www.linkedin.com/in/bahbbbc)
 - **Twitter:** @bahbbbc
- **Carol**
 - **LinkedIn:** [linkedin.com/in/oliverfelt](https://www.linkedin.com/in/oliverfelt)

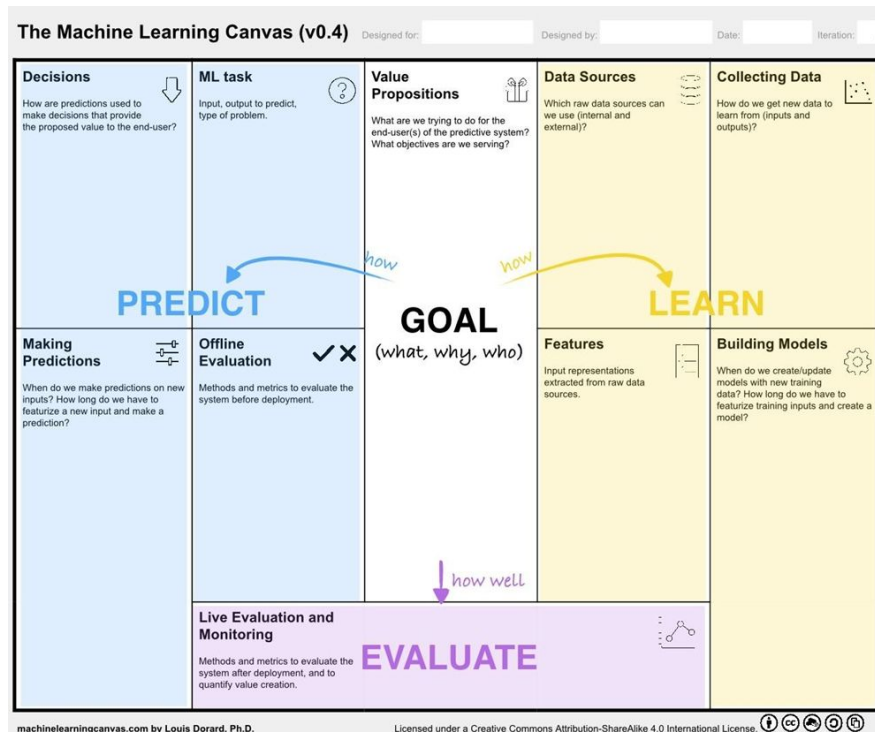


Dinâmica



Dinâmica

Machine Learning Canvas



Para praticar...

- ▷ O repositório do UCI contém alguns datasets para realizar clustering
- ▷ Também pode-se retirar a classe de datasets existentes para problemas supervisionados e aplicar técnicas de aprendizado não supervisionado!





Nossos contatos



mulheres.em.ia@gmail.com



mulheres-em-ia



@mulheres.em.ia



@mulheres.em.ia



@MulheresemInteligenciaArtificial



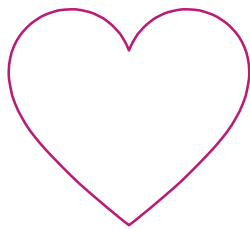
Canal: Mulheres em IA

Linktree

<https://linktr.ee/mulheres.em.ia>

Grupo Telegram para Mulheres

https://t.me/mulheres_em_ia



Muito obrigada!

Dúvidas? Podem nos procurar! 🙄

Por hoje, é isso!

Na próxima quarta falaremos
sobre **Introdução a Engenharia de
Dados** \o/



Vamos preencher o formulário de feedback???



<http://bit.ly/bootcamp-ds-sp-feedback-19>