

# MIT BIG DATA

## OBTENÇÃO, ANÁLISE E VISUALIZAÇÃO DE DADOS: VELOCIDADE

PROJETO DE BLOCO



# AGENDA

- Apresentações
- Objetivos e Competências
- Disciplinas
- Referências
- Relevância
- O Projeto de Bloco
- Papo Técnico
  - IA e ML
  - CRISP-DM



# Apresentações



# Professor e Coordenador

- [Cassius Figueiredo](#), M.Sc., M.B.A.  
([cassius.figueiredo@prof.infnet.edu.br](mailto:cassius.figueiredo@prof.infnet.edu.br))
- Formação:
  - Bacharel em Ciência da Computação (UERJ).
  - MBA em Finanças: Investimentos e Risco (FGV).
  - Mestre em Modelagem Matemática da Informação (EMAp – FGV).
- Experiência:
  - Cientista de dados.
  - Engenheiro Sênior na Órama Investimentos.
  - Prof. do Instituto Infnet desde 2012.
  - Coordenador MIT Big Data.
  - Prof. da Fundação Getúlio Vargas.



**Cassius Figueiredo**  
Data Scientist and Instructor





# Alunos

- Breve introdução
  - Nome.
  - O que faz?
  - Conhecimento prévio em Analytics, ML, Python ou R.
  - Onde está no curso?
  - Expectativas.

# Objetivos e Competências





# Objetivos

- Com base no estudo de casos reais, elaborar questões, coletar dados, realizar análises e apresentar resultados que respondam às questões levantadas.

# Competências em desenvolvimento

- Construir uma visão geral sobre aplicações de soluções adotadas no mundo Big Data;
- Desenvolver raciocínio analítico;
- Apresentar resultados.



# Disciplinas



# Big Data Analytics

- Linguagem R;
- Formulação de Problemas;
- Análises Quantitativa e Qualitativa;
- Modelagem;
- Avaliação e visualização de Resultados.



# Cases

- Apresentar soluções utilizadas em áreas diversas;
- Apresentar pipelines para soluções de problemas.

# Referências





# Referências

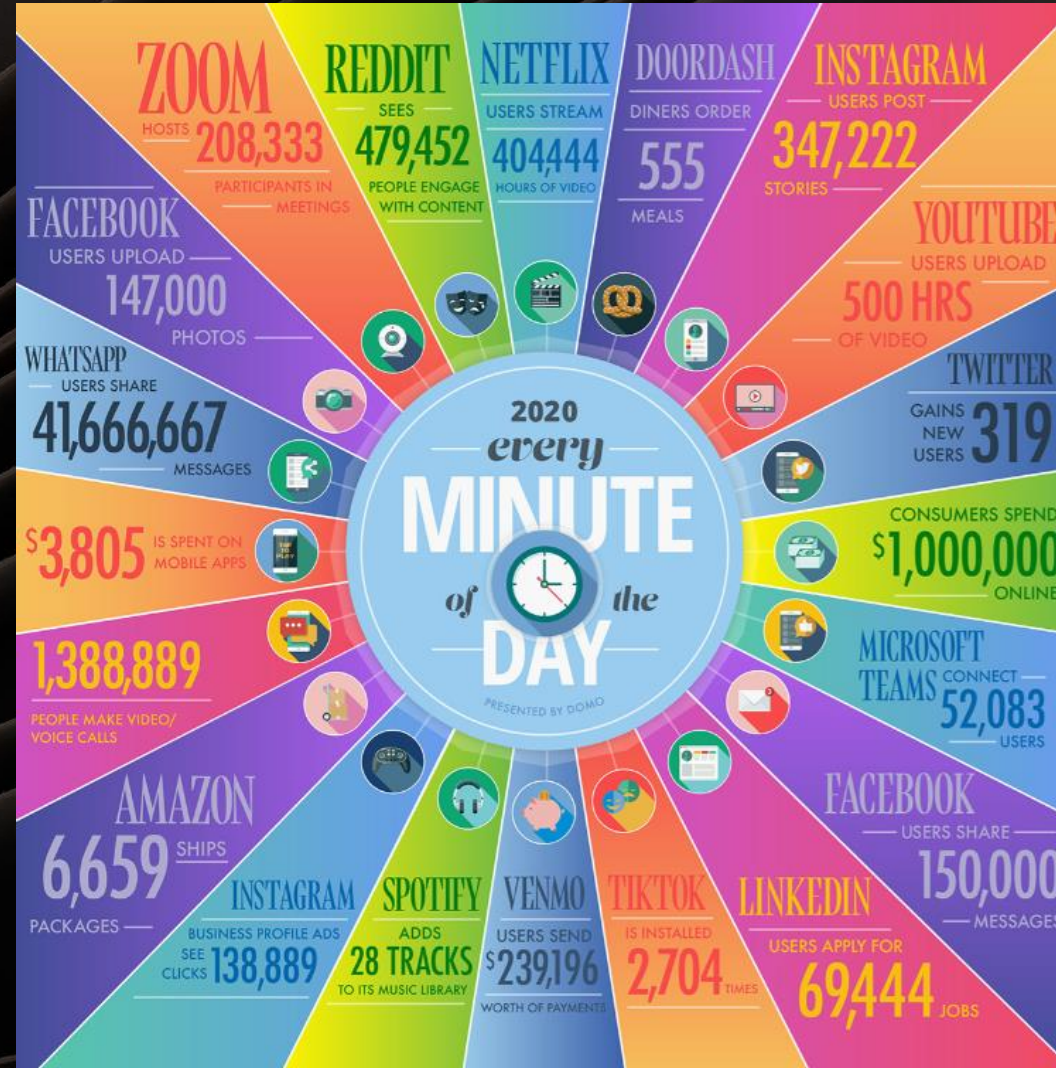
- [Garrett Grolemund, Hadley Wickham; R for Data Science](#)
- [Hadley Wickham; ggplot2: Elegant Graphics for Data Analysis](#)
- [Hadley Wickham; Advanced R](#)
- [Hadley Wickham; R Packages](#)
- [Winston Chang ; R Graphics Cookbook](#)
- [Kieran Healy; Data Visualization, A practical introduction](#)

# Relevância





# Qual a relevância do que veremos aqui?



# O Projeto de Bloco





# Projeto de Bloco

- Entrega que visa mostrar a aplicação dos conhecimentos adquiridos no decorrer do bloco.
- Artefatos:
  - Relatório escrito;
  - Aplicação em R.
- Desenvolvido em grupos, que devem estar definidos até a segunda aula da primeira disciplina do bloco.

# Relatório escrito

- Artefato em PDF que documente todo o processo de execução do projeto de bloco. Deve incluir (mas não está limitado a):
  - Escolha de um problema e base de trabalho;
  - Origem dos dados;
  - Tratamentos e preparações dos dados;
  - Processo de modelagem e avaliação;
  - Apresentação de resultados;
  - Dificuldades;
  - Soluções interessantes.



# Relatório escrito

- Estrutura
  - Colaterais
    - Capa, índice de tópicos e figuras, referências.
  - Introdução
  - Desenvolvimento
    - Arcabouço teórico;
    - Metodologia aplicada.
  - Apresentação e discussão de resultados
  - Conclusão

# Colaterais

- Capa
  - Título
  - Alunos
- Sumário
- Índice de figuras
- **Referências**



# Introdução

- Motivação e objetivos
- No relatório, apresentar um levantamento inicial do trabalho assim como explicar a(s) motivação(ões) e objetivo(s).
- Na apresentação, estas informações devem estar claras e sucintas.

# Desenvolvimento

- Descrição detalhada de como o projeto foi desenvolvido no decorrer do bloco.
- Todos os pontos, mesmos aqueles iniciais, devem ser apresentados de forma a demonstrar uma coerência do que foi realizado.



# Apresentação e discussão de resultados

- Os resultados são produtos gerados durante o projeto.
- Podem ser apresentados diversos pontos de execuções de sua aplicação, com a devida análise.
- A análise está relacionada ao que é esperado do projeto, portanto, explique como foi obtido tal resultado e se ele é satisfatório.
- Justifique eventuais problemas ou melhorias, se necessário.
- Visualizações de dados são formas muito eficientes e esperadas de comunicação em apresentações deste tipo!

# Conclusão

- Todo projeto deve ter uma conclusão coerente!
- A conclusão parte da análise feita dos resultados, logo, apresente de forma resumida os resultados do projeto e como eles agregam valor aos objetivos e motivações apresentadas na introdução.
- Podem incluir sugestões de melhorias para o projeto.



# Apresentação de projeto

- Cada grupo deverá apresentar o seu projeto em um tempo pré-definido. Podem incluir demonstrações reais de seu programa funcionando.
- Todos os componentes do grupo devem participar da apresentação!
- Não esqueça de testar tudo antes para não se atrapalhar com o tempo.
- Ensaiai a apresentação é uma ótima forma de ganhar confiança!
- Uma outra alternativa é gravar uma execução de sua aplicação e apresentar o vídeo durante a apresentação. Desta forma, você pode administrar melhor seu tempo de apresentação.

# Observações gerais

- Será definida a ordem e o horário de cada apresentação.
- A participação de toda a turma é importante e poderá ser utilizada como critério de avaliação.
- **Teremos um cronograma definido de entregas intermediárias e aulas de acompanhamento e suporte, principalmente na segunda disciplina do bloco (Cases)**



# Aplicação

- Desenvolvida em R;
- Deverá ser bem documentada, no decorrer do código e no relatório escrito;
- Deve conter todas as etapas de solução propostas;
- Não precisa ser limitado a técnicas e pacotes vistos em sala.

# Cronograma (em construção)

- **23/01/2021 – Aulas 01 e 02 PB**
- **Entregas intermediárias**
  - No decorrer das disciplinas teremos aulas dedicadas ao PB onde definiremos os marcos intermediários.
- **19/06/2021 – Aulas 03 e 04 PB**
  - Apresentações (cronograma a ser definido)



# Avaliação

- Definiremos os critérios de avaliação no decorrer das disciplinas, sendo a nota ao menos uma composição dos seguintes critérios:
  - Qualidade do relatório
  - Qualidade da apresentação
  - Participação na apresentação
  - Aderência ao tempo de apresentação

# TEMA PRINCIPAL

Analytics como ferramenta de apoio  
para a evolução social brasileira





# Tema

- Propor soluções para problemas brasileiros, com dados brasileiros.
- Propor um problema e apresentar a proposta de solução.

# Bases

- Alguns sites com bases de dados brasileiras
  - <http://dados.gov.br/>
  - <http://www.ipeadata.gov.br/Default.aspx>
  - <https://dadosabertos.bcb.gov.br/>
  - <http://portal.inep.gov.br/web/guest/dados>
  - <http://www.data.rio/>
  - DataSUS



# DESAFIO



# Desafio!

- Se algum grupo estiver disposto seria uma ótima experiência publicar o trabalho.
- Darei todo o apoio necessário.
- Alvos:
  - Publicações
  - Congressos



# Papo técnico



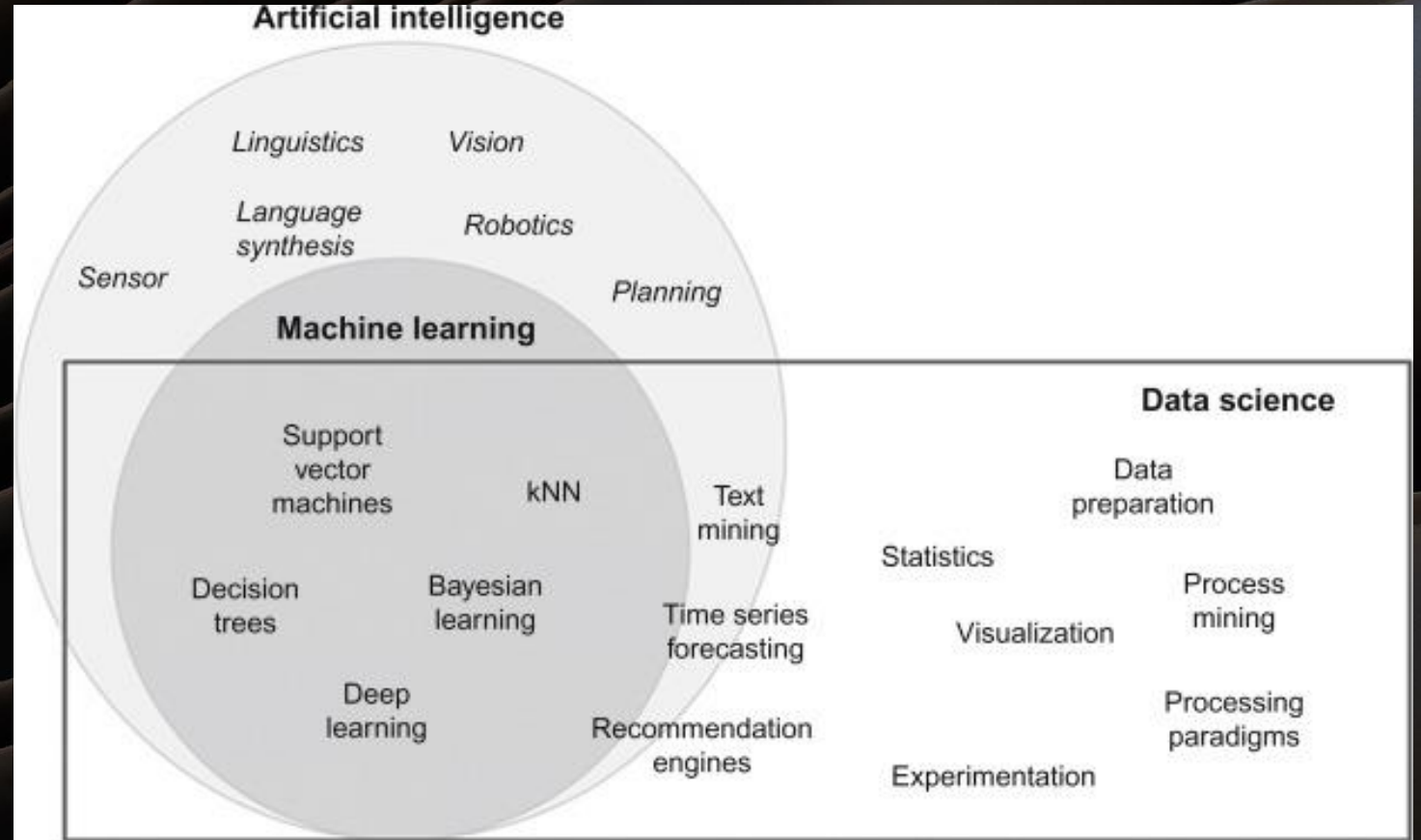
# Inteligência Artificial e Machine Learning





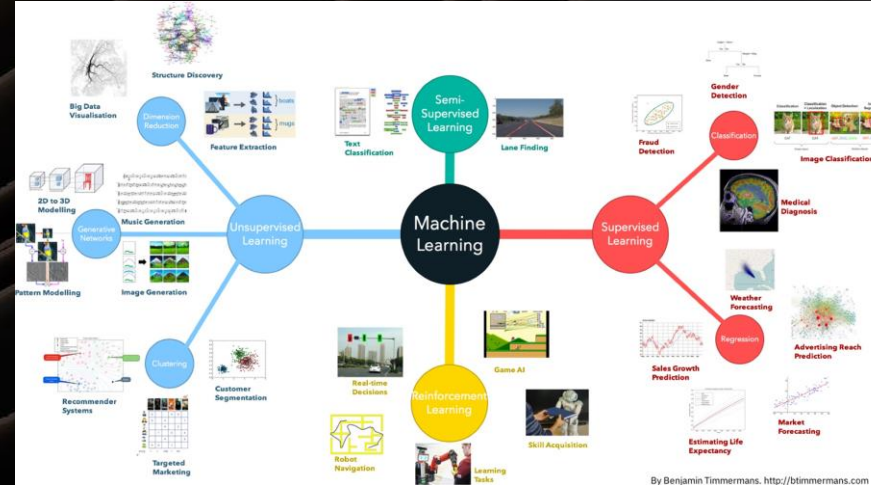
# Inteligência Artificial

- Evolução no tempo
- Áreas



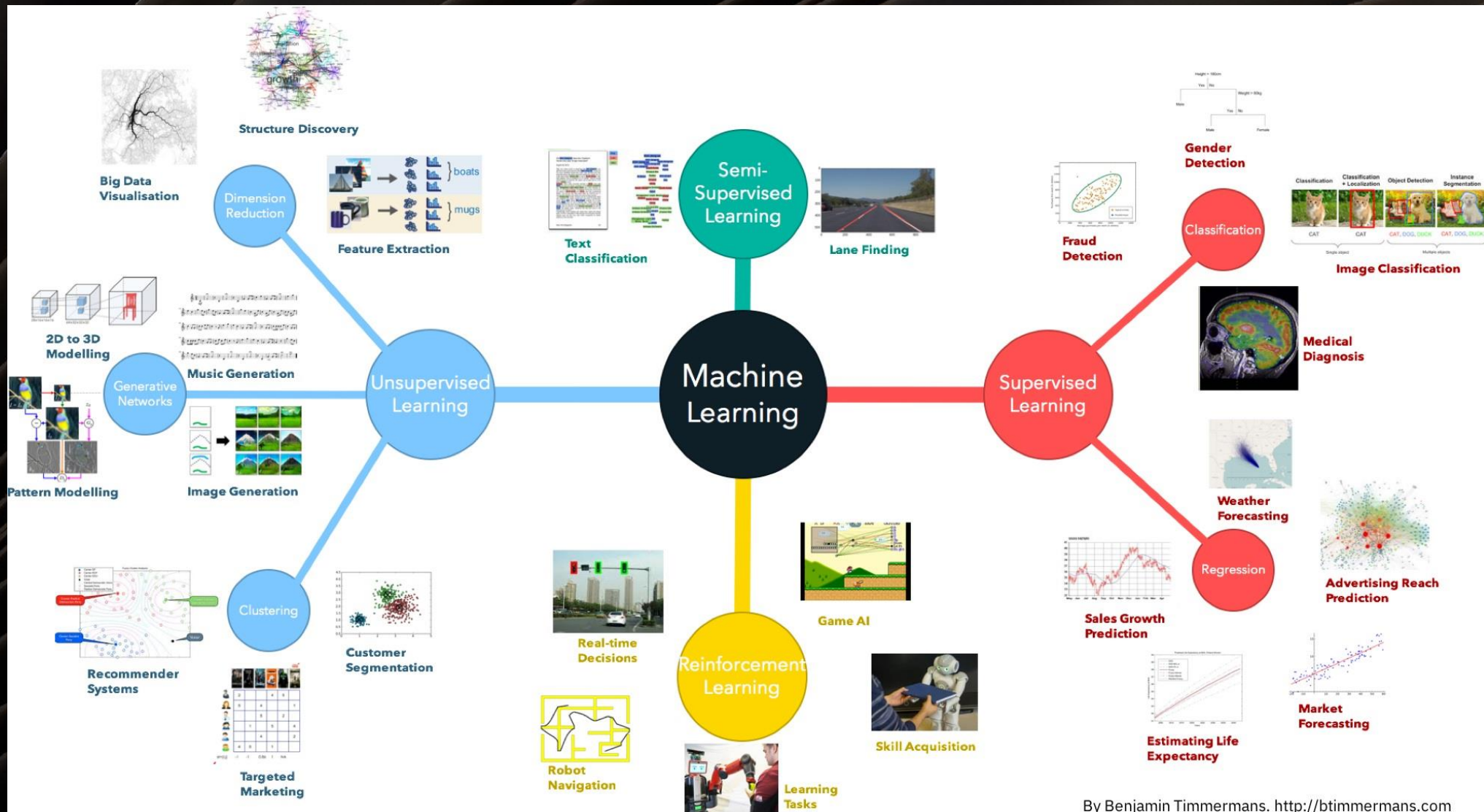
# Machine Learning

- Tipos de problemas
- Aplicações
- Pipeline básico

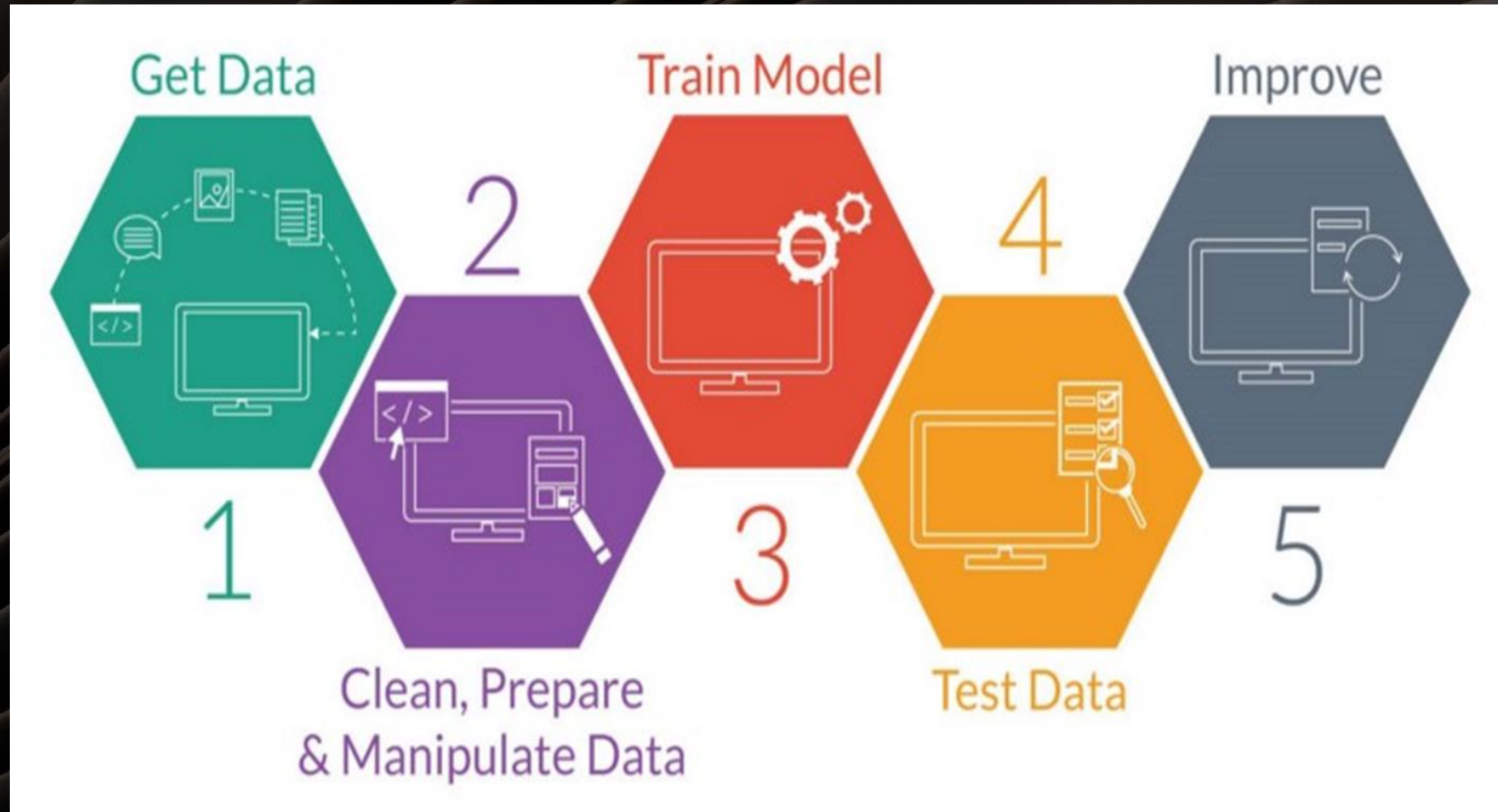




# Problemas e aplicações

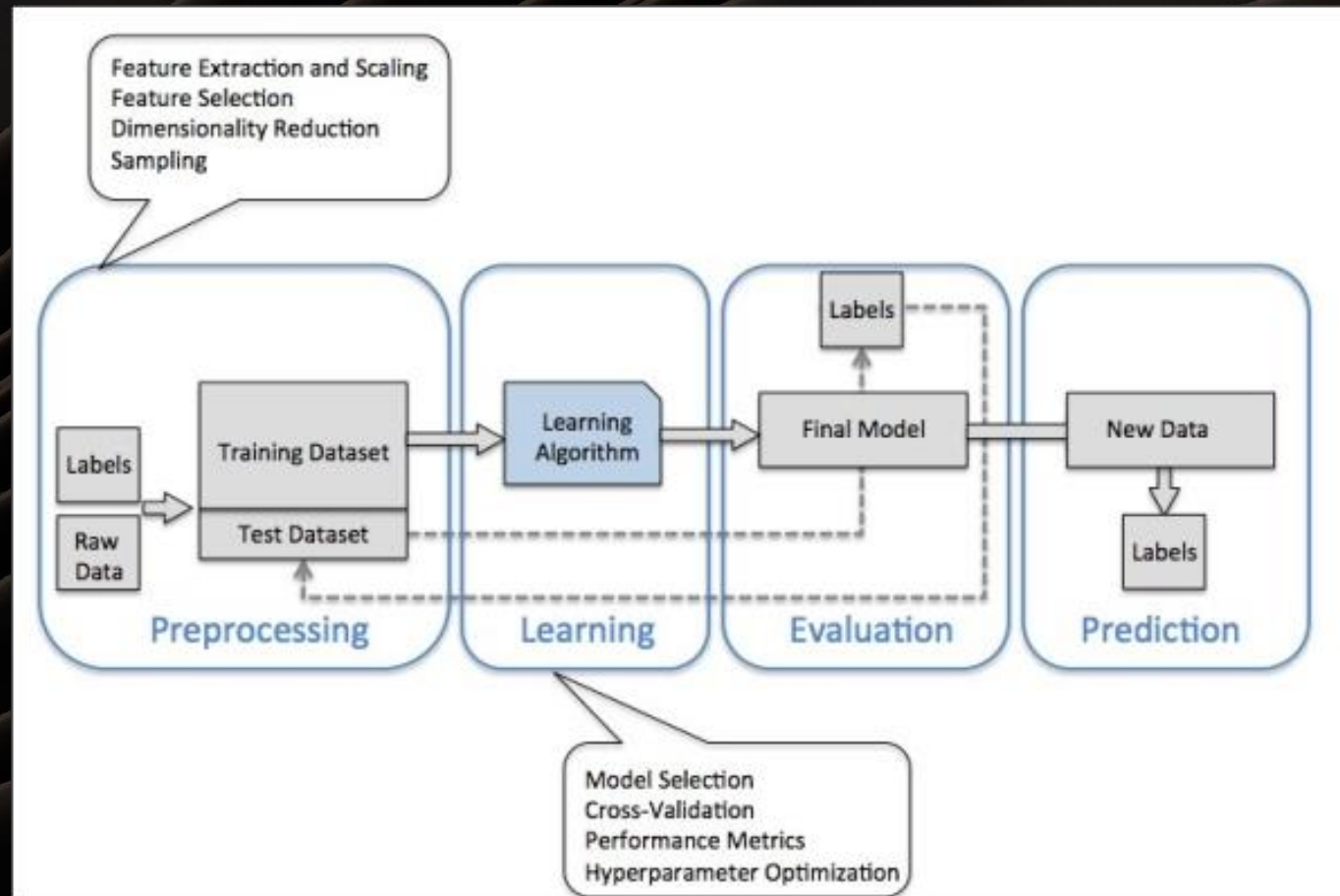


# Pipeline





# Pipeline



# CRISP-DM





# CRISP-DM

- Cross-Industry Standard Process for Data Mining

