

## Projeto 1 SBS - Conceção e implementação de modelos de Machine Learning baseados em árvores

Data Set 1 - **Em escolha** - objetivo principal: melhorar o modelo

Data Set 2 - Competição - objetivo principal: melhorar a accuracy

<https://www.youtube.com/watch?v=dEBdn5XFor8>

### Arquitetura Baseada no CRISP-DM

- **Business and Data Understanding** - Entender os dados **a olho** e definir os objetivos de previsão e de análise de dados;
- **First Extraction, Processing and Loading** - Desenvolver a primeira componente de etl no knime para **preparar os dados para serem analisados**;
- **Data Analysis** - Realizar todas as análises importantes para **devolver informações importantes e relevantes**;
- **Data Preparation** - Preparar os dados para modelo e fazer todas as **transformações necessárias melhoria do modelo**;
- **Modelação** - Partição dos dados em 20/80 para teste e treino, e usar o cross validation com tune grid para random forest e decision tree;
- **Avaliação do Modelo** - Avaliar o modelo;
- **Melhoria do modelo** - Tentar melhorar o modelo mudando as co-variáveis e refazendo os parâmetros desde o passo **Data Preparation**;

### Desenvolvimento do Relatório

1. Introdução
2. Metodologias -> Introduzir o Machine Learning
  - a. CRISP-DM
  - b. Modelos Baseados em Árvores
3. Arquitetura e Ferramentas
4. **1º Dataset**
  - a. Contexto do data set
  - b. Análise e compreensão de dados (**Business and Data Understanding;First Extraction, Processing and Loading;Data Analysis**)
  - c. Modelação (**Data Preparation ;Modelação;**)
  - d. Análise de Resultados (**Avaliação do Modelo; Melhoria do modelo**)
5. **2º Dataset**
  - a. Contexto do data set
  - b. Análise e compreensão de dados (**Business and Data Understanding;First Extraction, Processing and Loading;Data Analysis**)
  - c. Modelação (**Data Preparation ;Modelação;**)
  - d. Análise de Resultados (**Avaliação do Modelo; Melhoria do modelo**)
6. Conclusão

**Ferramentas a Utilizar:**

- Knime Analytics Platform

**Data de Entrega:**

- 25 de novembro de 2020

**Data de Apresentação:**

- 26 de novembro de 2020