

Introdução à Análise de Dados com Aprendizado de Máquina – Anuário ANTAQ 2016-2019



Universidade Católica de Santos - UNISANTOS

Laboratório de Inteligência Artificial

Prof. Me. Marcio Luiz Piva

06/06/2020

Professor dos Cursos de Ciência da Computação e Sistemas de Informação – desde 2005 – 15 anos

Disciplinas: Inteligência Artificial, Programação Paralela e Concorrente, Sistemas Distribuídos e Sistemas Operacionais

Mestre em Engenharia da Computação – Universidade Federal de Uberlândia – Área de Pesquisa: Inteligência Artificial

Pós-Graduações: Engenharia da Computação, MBA Portos e Logística e Administração de Sistemas da Informação

Gerente de Projetos PMP – Project Managment Institute (PMI)

Gerente de Tecnologia da Informação com mais de 25 anos de experiência, sendo 18 anos na área Portuária

PARTE 1 – Breve Introdução

- ✓ *O Laboratório de Inteligência Artificial – LabIA*
- ✓ *Inteligência Artificial – Aspectos Gerais*
- ✓ *Aprendizado de Máquina – Conceitos*
- ✓ *Anuário ANTAQ 2016-2019*

PARTE 2 – Bits e Bytes

- ✓ *Ferramentas Utilizadas*
- ✓ *Machine Learning*
 - *Decision Trees / Random Forest – Scikit Learn*
 - *Linear Regression – TensorFlow*
 - *Deep Learning (MLP) – TensorFlow e Scikit Learn*

PARTE 3 – Finalizando

- ✓ *Questões Gerais*

- **INFRAESTRUTURA:**

- 16 estações de trabalho:
 - Intel Core i5 (4 cores – 8 Th);
 - 8 GB RAM DDR5;
 - SSD 256 GB;
 - Geforce RTX 2060:
 - 8 GB VRAM (GDDR6);
 - 2048 Cuda Cores;



- **DISCIPLINAS (algumas):**

- Inteligência Artificial;
- Programação Paralela e Concorrente;
- Sistemas Distribuídos;
- Sistemas Operacionais;
- Arquitetura de Computadores;



“A Inteligência Artificial (IA) pode ser definida como o ramo da ciência da computação que se ocupa da automação do comportamento inteligente.”

“O problema de definir o campo inteiro da inteligência artificial é semelhante ao de definir a própria inteligência.”

George F. Luger – Inteligência Artificial – 6a. Edição



“O aprendizado de máquina é um ramo da inteligência artificial. Usando a computação, projetamos sistemas que podem aprender com os dados de maneira a serem treinados. Os sistemas podem aprender e melhorar com a experiência e com o tempo, refinar um modelo que pode ser usado para prever resultados de perguntas com base no aprendizado anterior.”

Jason Bell – Machine Learning: Hands on for developers and technical professionals

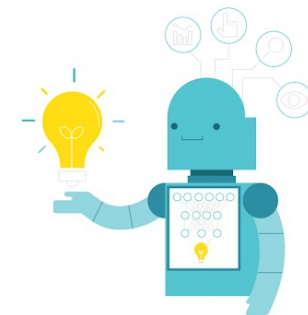
Qual é o foco? Que resultados procuramos e/ou esperamos?

- ★ **Predição**
- ★ **Classificação**
- ★ **Agrupamento (clusterização)**

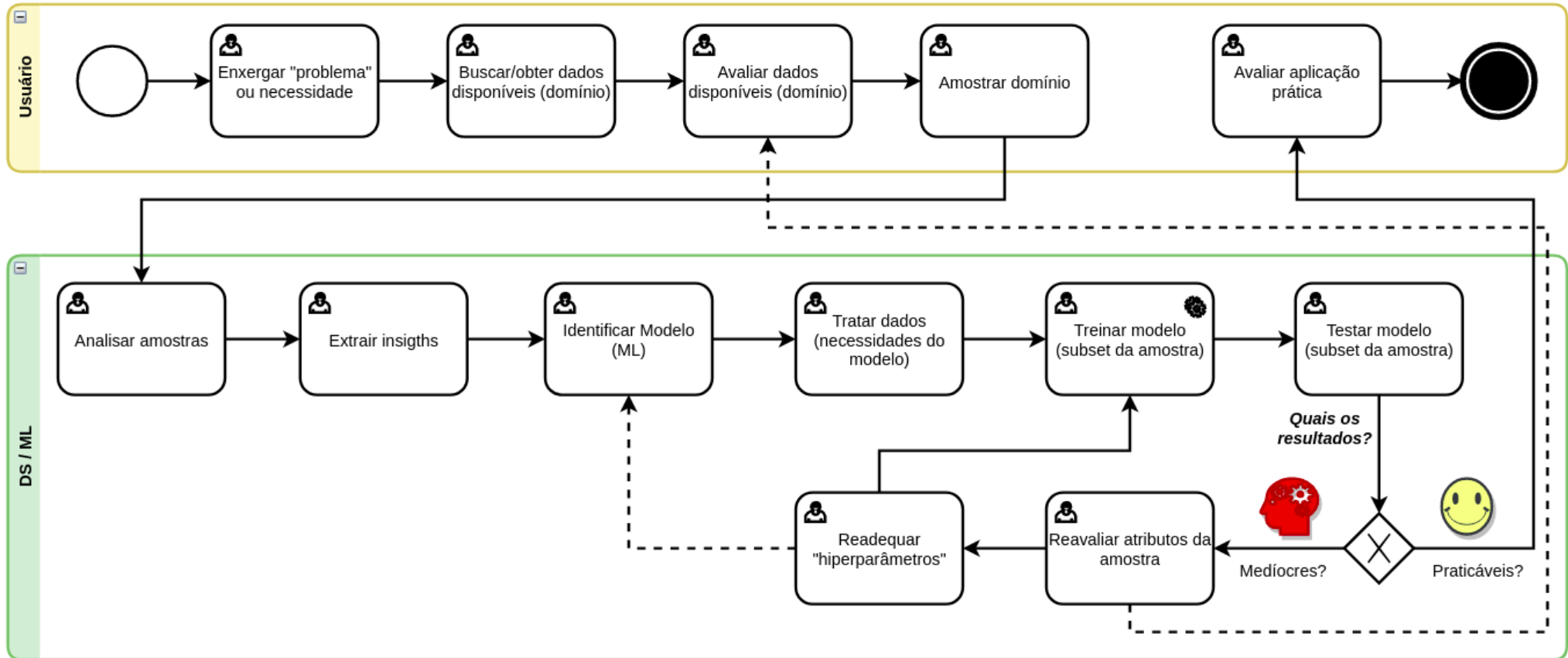


O que temos em mãos? O que sabemos sobre o problema?

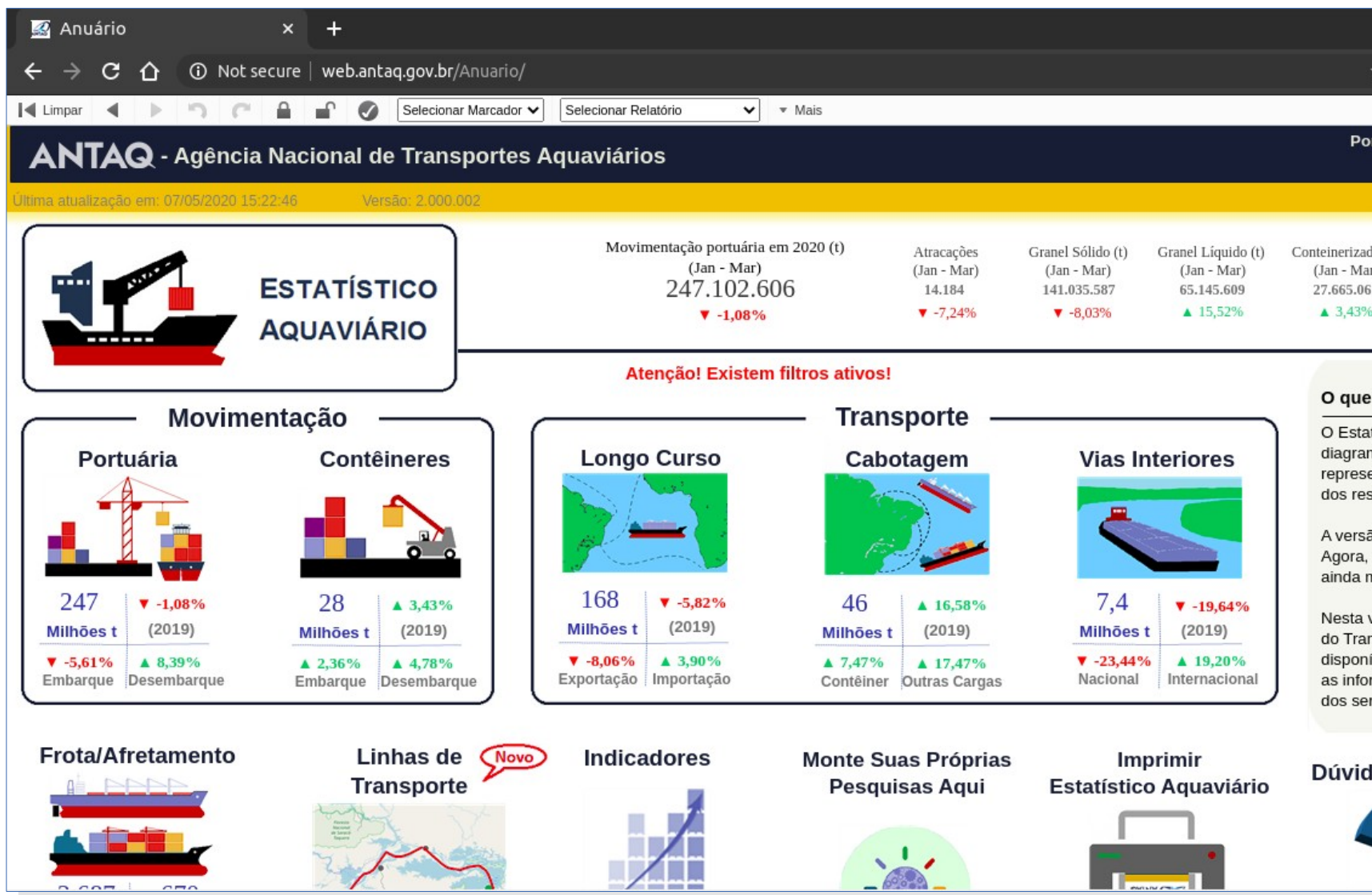
- ★ **Aprendizado Supervisionado**
- ★ **Aprendizado não-Supervisionado.**



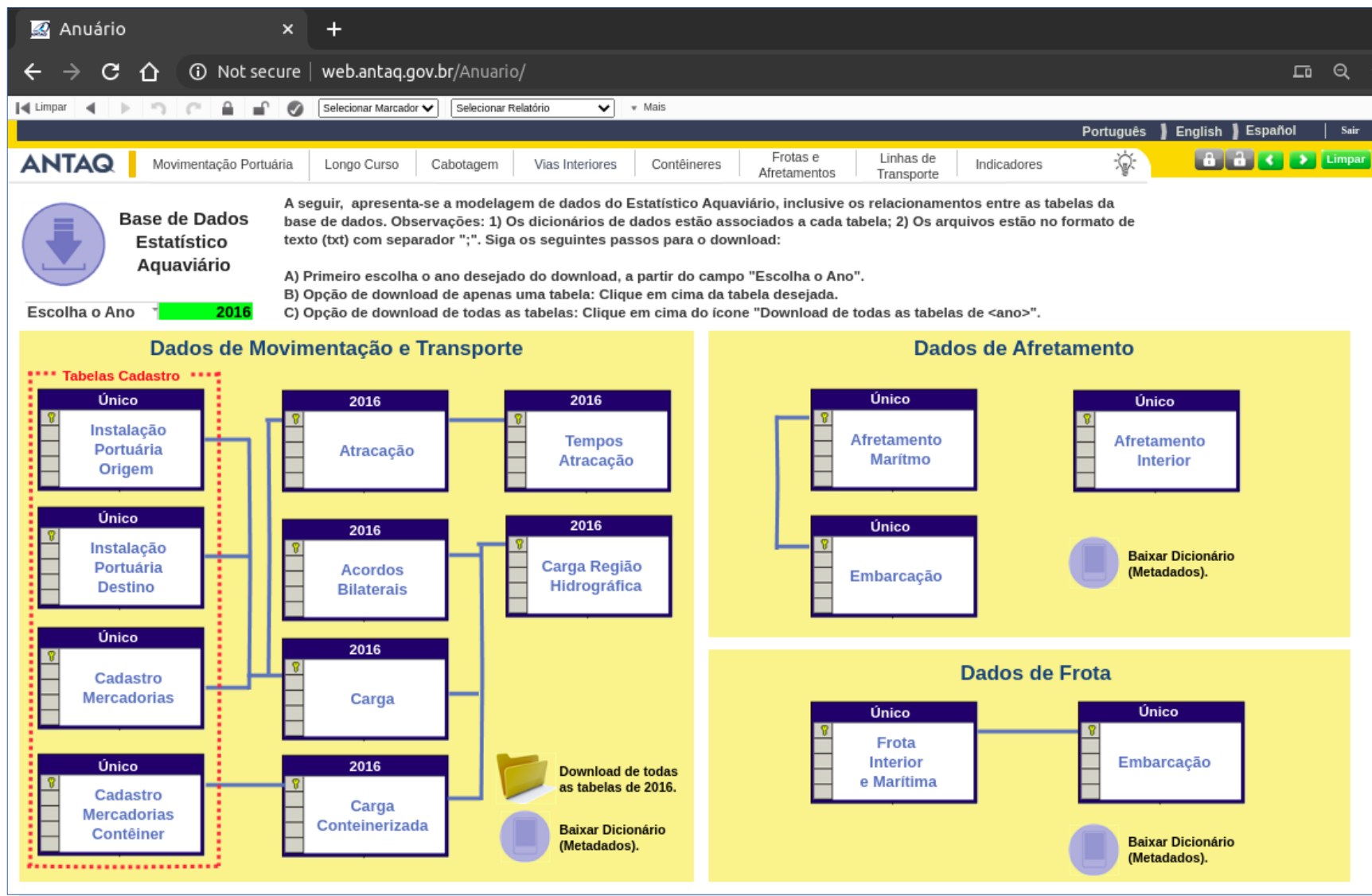
FLUXO GERAL SIMPLIFICADO PARA ANÁLISE DE DADOS E APLICAÇÃO DE APRENDIZADO DE MÁQUINA



<http://web.antaq.gov.br/Anuario/>



<http://web.antaq.gov.br/Anuario/> - [base de dados]



Anuário

Not secure | web.antaq.gov.br/Anuario/

Seletor de idiomas: Português | English | Español | Sair

ANTAQ | Movimentação Portuária | Longo Curso | Cabotagem | Vias Interiores | Contêineres | Frotas e Afretamentos | Linhas de Transporte | Indicadores

Base de Dados Estatístico Aquaviário

Escolha o Ano: **2016**

A seguir, apresenta-se a modelagem de dados do Estatístico Aquaviário, inclusive os relacionamentos entre as tabelas da base de dados. Observações: 1) Os dicionários de dados estão associados a cada tabela; 2) Os arquivos estão no formato de texto (txt) com separador ";". Siga os seguintes passos para o download:

- A) Primeiro escolha o ano desejado do download, a partir do campo "Escolha o Ano".
- B) Opção de download de apenas uma tabela: Clique em cima da tabela desejada.
- C) Opção de download de todas as tabelas: Clique em cima do ícone "Download de todas as tabelas de <ano>".

Dados de Movimentação e Transporte

Tabelas Cadastro

- Único: Instalação Portuária Origem
- Único: Instalação Portuária Destino
- Único: Cadastro Mercadorias
- Único: Cadastro Mercadorias Contêiner

2016

- Atracação
- Acordos Bilaterais
- Carga
- Carga Containerizada

2016

- Tempos Atracação
- Carga Região Hidrográfica

Download de todas as tabelas de 2016.

Baixar Dicionário (Metadados).

Dados de Afretamento

- Único: Afretamento Marítimo
- Único: Afretamento Interior
- Único: Embarcação

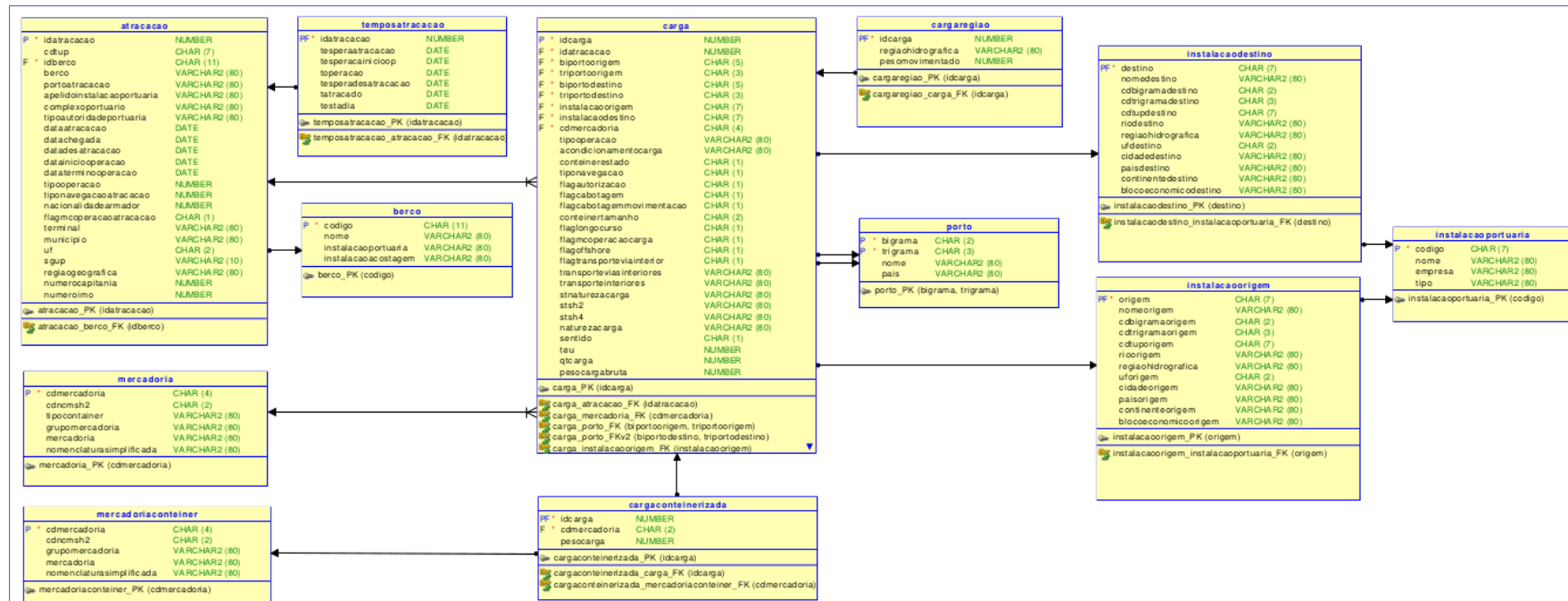
Baixar Dicionário (Metadados).

Dados de Frota

- Único: Frota Interior e Marítima
- Único: Embarcação

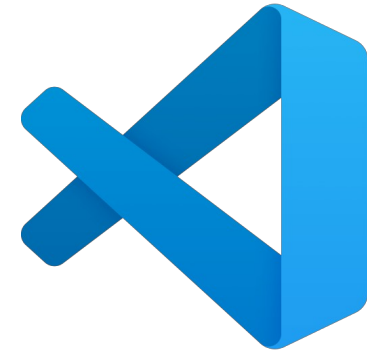
Baixar Dicionário (Metadados).

Modelo de Dados (completo)



Foco das análises

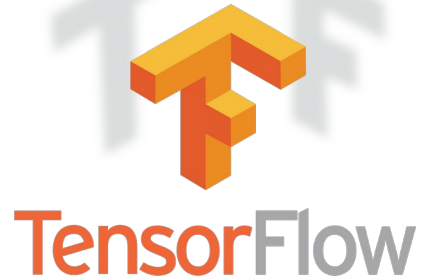
- Atracacao – 296.051 registros
- TemposAtracacao – 294.115 registros
- Cargas – 8.068.230 registros



NumPy

matplotlib

pandas



Bits e Bytes

https://github.com/marciopiva/anuario_antaq

"Although linear models have significant limitations as practical techniques for pattern recognition, particularly for problems involving input spaces of high dimensionality, they have nice analytical properties and form the foundation for more sophisticated models..." - Crhistopher M. Bishop - Pattern Recognition and Machine Learning

"Machine learning (ML) is a branch of artificial intelligence that systematically applies algorithms to synthesize the underlying relationships among data and information." - Mariette Awad and Raul Kanna - Efficient Learning Machines - Theoris, Concepts, and Applications for Engineers and System Designers

"The future belongs to the companies and people that turn data into products." - Mike Loukides - O'Reilly & Associates - What is data science?

"The extracted knowledge is not only available and usable as a black box function, but rather it can be easily understood, interpreted, and controlled by humans in the form of a readable decision tree." - Wolfgang Ertel - Introdution to Artificial Intelligence

"In standard trees, each node is split using the best split among all variables. In a random forest, each node is split using the best among a subset of predictors randomly chosen at that node." - Andy Liaw an Matthew Wiener - Classification and Regression by randomForest

"...a neural network derives its computing power through, first, its massively parallel distributed structure and, second, its ability to learn and, therefore, generalize." - Milan Hajek - Neural Networks

Muito obrigado!

marciopiva@unisantos.br