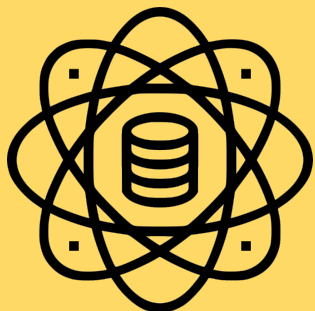


Exercício de análise de dados

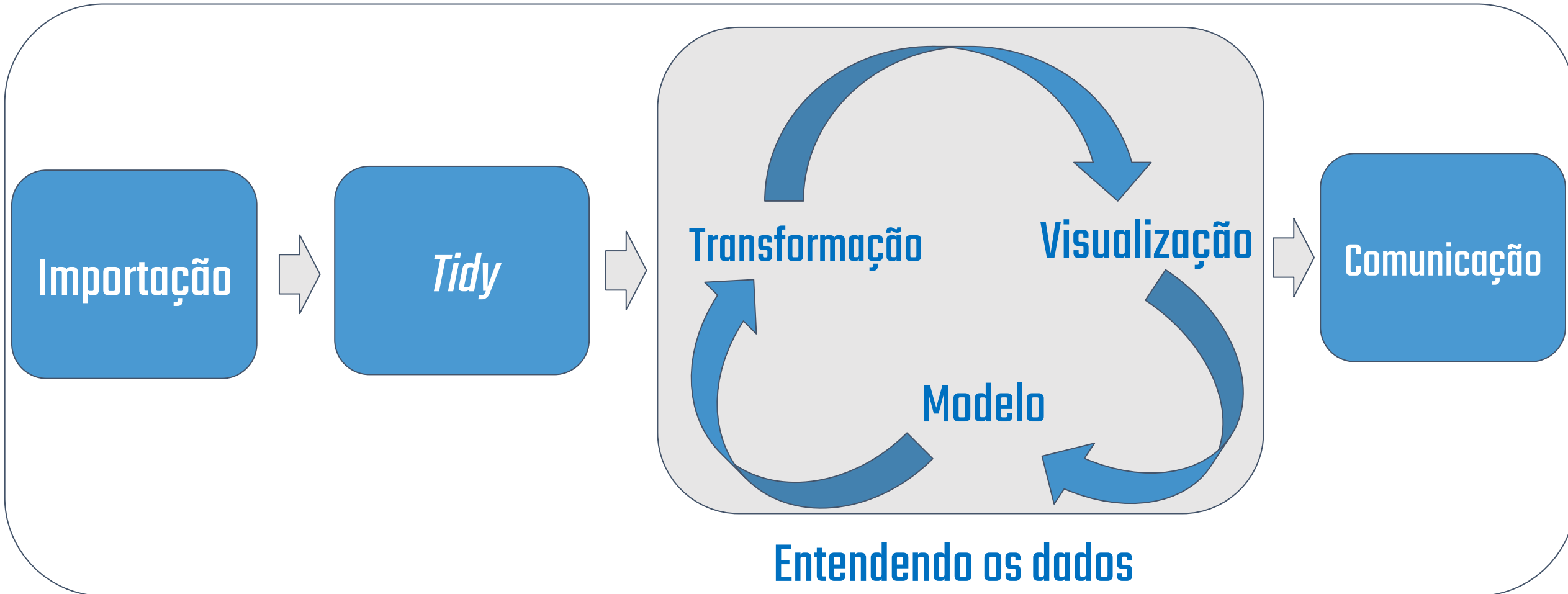


13 de Agosto de 2021
São José dos Campos - SP

Adriano Almeida
Felipe Carvalho
Felipe Menino

Nas aulas anteriores...

Programação



Exercício de análise de dados

Objetivo

Utilizar o fluxo de análise de dados e os conceitos vistos nas aulas anteriores para **organizar, explorar e classificar** o conjunto de dados.

Exercício de análise de dados

Contextualização

Remote Sensing Letters
Vol. 3, No. 6, November 2012, 491–499



Using geographically weighted variables for image classification

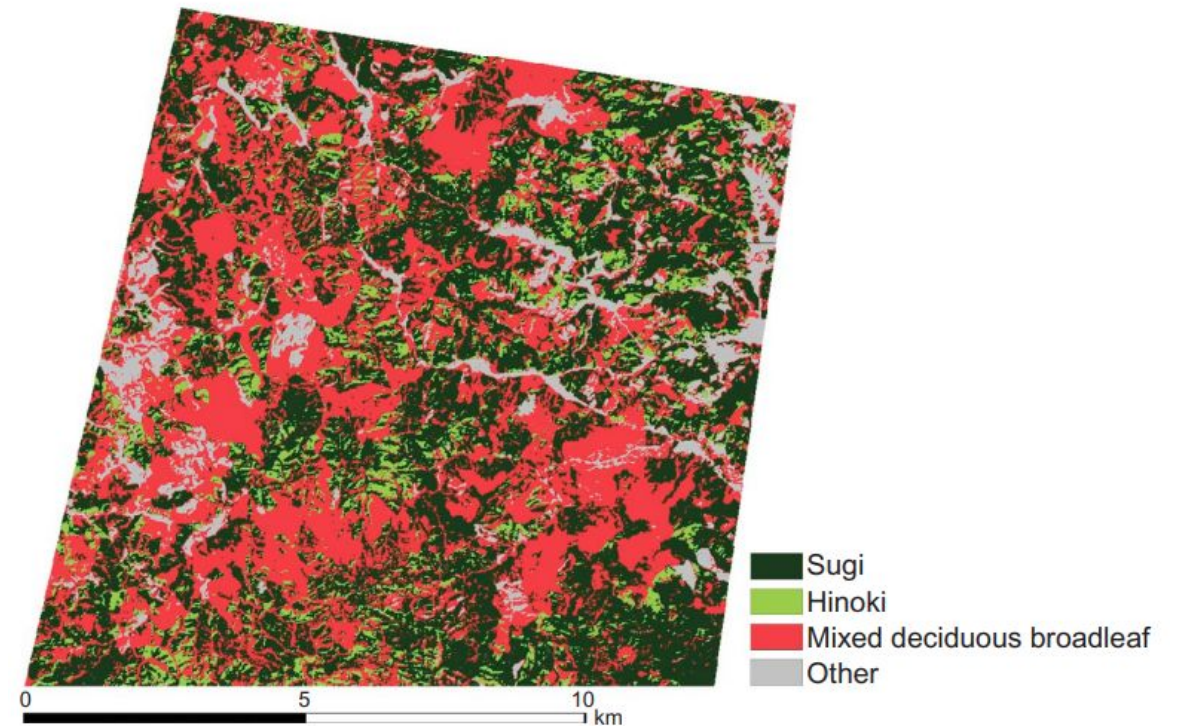
BRIAN JOHNSON*†, RYUTARO TATEISHI† and ZHIXIAO XIE†

†Department of Geosciences, Florida Atlantic University, Boca Raton, FL 33431, USA

‡Center for Environmental Remote Sensing (CEReS), Chiba University, Chiba, Japan

(Received 22 July 2011; in final form 1 October 2011)

In this study, geographically weighted variables calculated for two tree species, *Cryptomeria japonica* (Sugi) and *Chamaecyparis obtusa* (Hinoki), were used in addition to spectral information to classify the two species and one mixed forest class. Spectral values (digital numbers for each band) of ‘Sugi’ and ‘Hinoki’ training samples were used to predict the spectral values for the two species at other locations using the inverse distance weighting (IDW) interpolation method. Next, the similarity between each pixel’s spectral values and their IDW predicted values was calculated for both of the tree species. The similarity measures are considered to be geographically weighted because nearer training samples have more of an impact on their calculation. The use of geographically weighted variables resulted in an increase in overall accuracy from 82.2% to 85.9% and an increase in the kappa coefficient from 0.740 to 0.795 for a support vector machine classification.



Mapa classificado

(JOHNSON, TATEISHI E XIE, 2012)

Exercício de análise de dados

Contextualização

Forest Type Dataset

Conjunto de dados estruturado com a informação de classes de árvores e seu comportamento espectral extraído de imagens de Sensoriamento Remoto.

- Dados extraídos de imagens ASTER/Terra
- 27 atributos
- 523 amostras
 - 209 para treino
 - 314 para teste

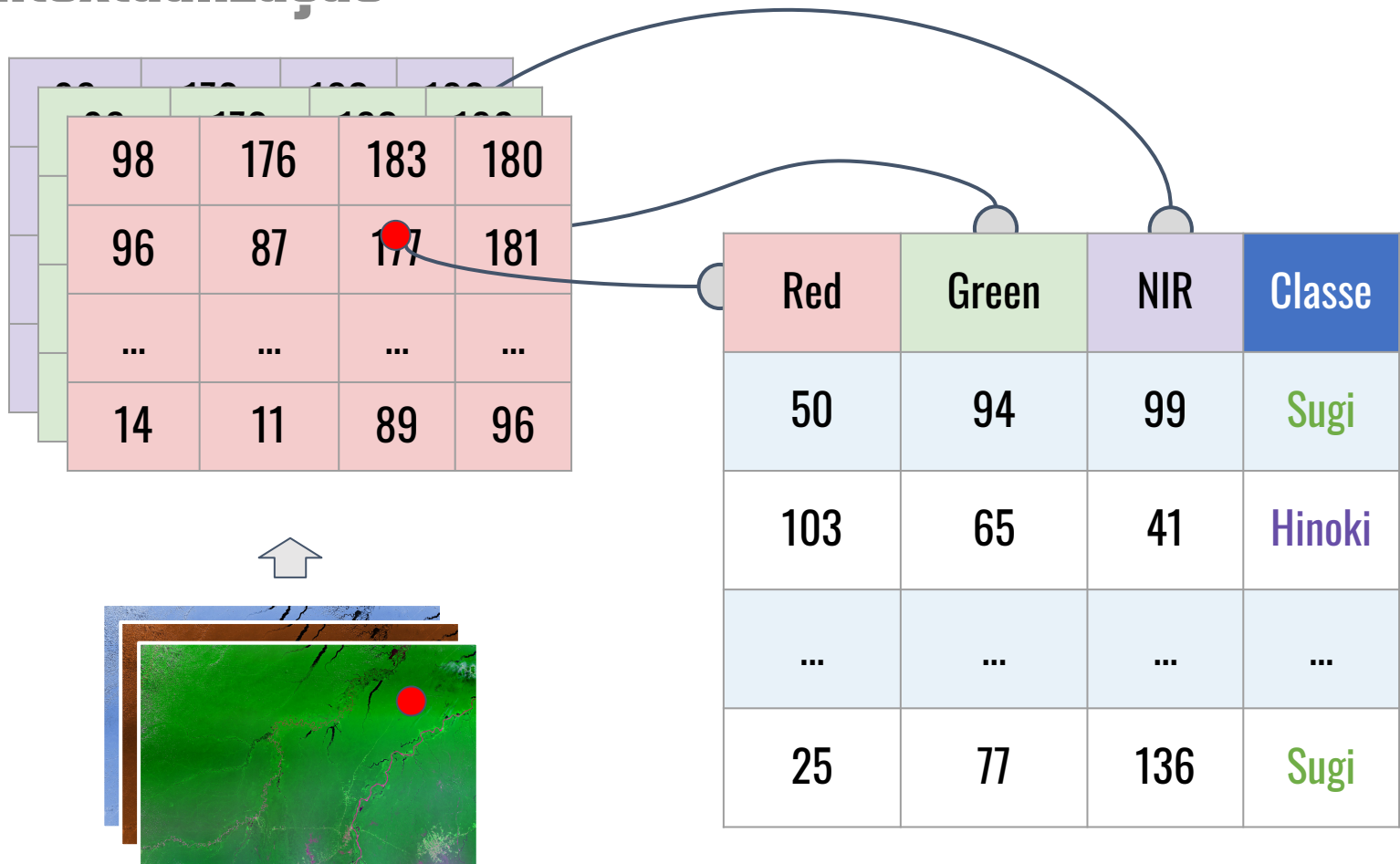
Exercício de análise de dados

Contextualização

Red	Green	NIR	Classe
50	94	99	Sugi
103	65	41	Hinoki
...
25	77	136	Sugi

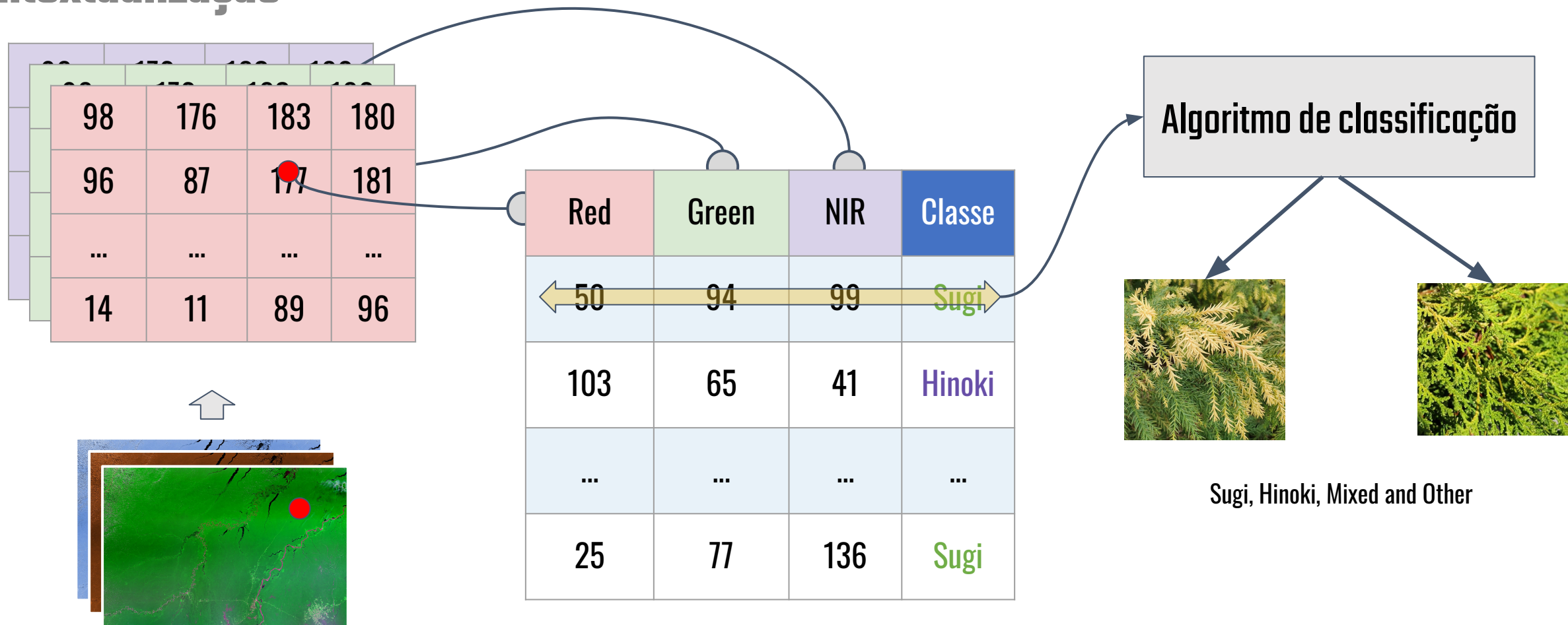
Exercício de análise de dados

Contextualização



Exercício de análise de dados

Contextualização



Exercício de análise de dados

Conjunto de dados

 InClass Prediction Competition

CAP 394 - Exercício de análise de dados

Exercício para aplicar os conceitos de análise de dados

Host

Overview

Data

Code

Leaderboard

Rules

Team

My Submissions

...

Overview

Edit

Description

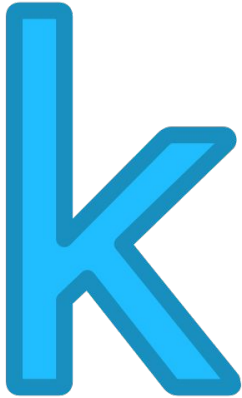
Objective

Tips

CAP 394 - Exercício de análise de dados

Este exercício foi criado para estimular os alunos a fazer análise de dados. Para isso, é proposta como atividade a aplicação dos conceitos de análise de dados apresentados durante as aulas em um conjunto de dados, com o objetivo de fazer sua limpeza e organização, de modo que os dados possam ser

<https://www.kaggle.com/c/cap-394-exercicio-de-analise-de-dados/>



Exercício de análise de dados

Conjunto de dados

ID	Atributo 1	Classe
1	3.1415	A
2	7.77	D
...
N	2.718	E

treino.csv

ID	Atributo 1	Classe
1	2.775	?
2	4.44	?
...	...	?
N	1.309	?

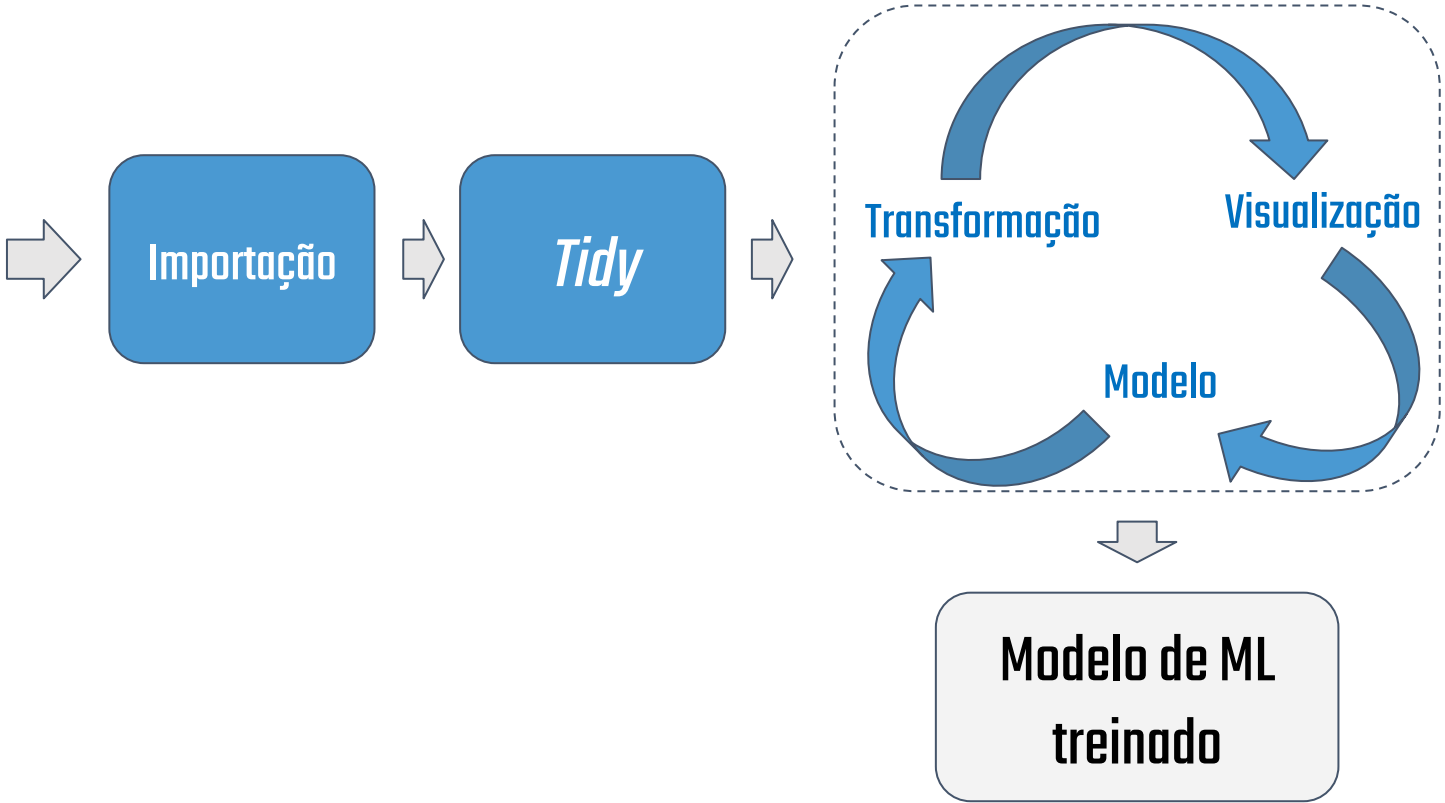
teste.csv

Exercício de análise de dados

Resultados

ID	Atributo 1	Classe
1	3.1415	A
2	7.77	D
...
N	2.718	E

treino.csv



Exercício de análise de dados

Resultados

ID	Atributo 1	Classe
1	2.775	?
2	4.44	?
...	...	?
N	1.309	?

teste.csv



Modelo de ML
treinado



ID	Atributo 1	Classe
1	2.775	D
2	4.44	E
...
N	1.309	A

Resultados da predição

Exercício de análise de dados

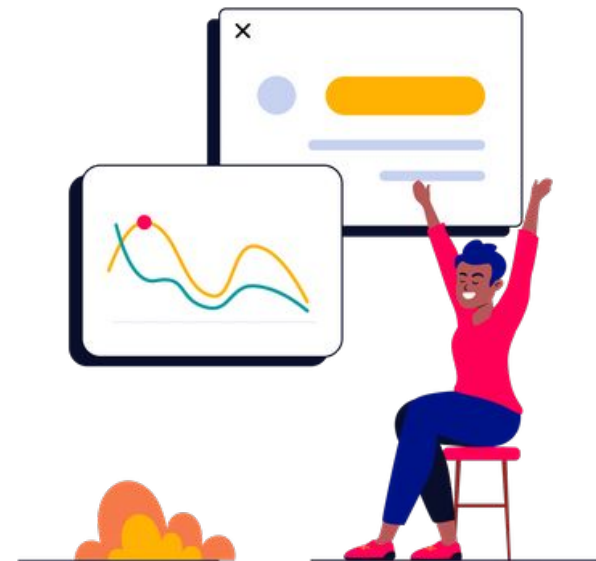
Recomendações

- Explorar os conceitos vistos em aula;
- Trocar informações com os colegas;
- Entender os dados é a chave do problema.

Exercício de análise de dados

Dinâmica

- Horário da aula 13:30 às 15:30;
- A entrega não é obrigatória, mas recomendada;
 - Quanto mais tempo você passa com os dados, mais eles te ensinam
- Não é uma competição!



Exercício de análise de dados

Materiais auxiliares

Prefácio

Grupo DataAt

Nossos livros-textos

Licença

I Introdução

1 Introdução

1.1 Machine learning

II Aprendizado Supervisionado

2 Regressão

2.1 Regressão Linear

2.2 Máquina de vetores de suporte

3 Classificação

3.1 k-Nearest Neighbors

3.2 Árvore de decisão

III Aprendizado Não Supervisionado

4 Agrupamento

4.1 O que é um agrupamento?

4.2 Técnicas de agrupamento

4.3 Kmeans

4.4 Agrupamento Hierárquico - M...

5 Exemplos

IV Apêndice

Introdução ao Machine Learning

Adriano Almeida

Felipe Carvalho

Felipe Menino

Prefácio

Seja bem vinda(o) ao livro-texto do minicurso de **Introdução ao Machine Learning**. Criamos este material para compartilhar o pouco que sabemos e dividir nossas experiências. Neste material, você irá encontrar conteúdos sobre o conceito, técnicas e algumas dicas úteis sobre Machine Learning. Procuramos abordar os conceitos de forma didática, porque sabemos o quão difícil é se inteirar de uma nova área, principalmente para as pessoas que não estão familiarizadas com o assunto. Este livro-texto não tem um público-alvo, escrevemos com o objetivo de atingir o máximo de pessoas em quaisquer áreas. O único pré-requisito para ler este livro-texto é ter curiosidade, porque não são as respostas que movem o mundo, e sim, as perguntas!



WORCAP 2020 Workshop em Computação Aplicada

MC2 - INTRODUÇÃO AO MACHINE LEARNING

Transmissão pelo YouTube: <https://www.youtube.com/c/PGCAPINPE>

8-11 e 14-17 de setembro

Exemplo de aplicação do Método Kmeans

- Instrutores:** Adriano, Felipe Carvalho e Felipe Menino
- Realização:** Dia 15/09
- Descrição:** Objetiva-se apresentar aos alunos exemplos de aplicação de algoritmos de agrupamento.
- Sumário:**
 - Descrição do conjunto de dados
 - Agrupamento dos dados genéticos
 - Agrupamentos dos dados reais
 - Referências
- Links Úteis:**
 - Livro **Introdução ao Machine Learning**
 - Exemplo de **Classificação** em Python
 - Exemplo de **Regressão** com **Máquina de Vetores de Suporte** em Python
 - Exemplo de **Agrupamento Hierárquico** em R

Descrição do Notebook

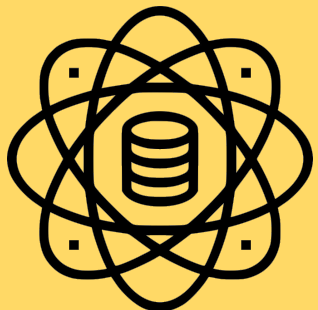
Neste notebook vamos apresentar um exemplo de aplicação do **Kmeans** usando a linguagem **R**. Para isso, vamos usar dois conjunto de dados: dados genéticos com base em uma distribuição normal e dados sobre os **Países do mundo**.

O conjunto de dados de **Países do mundo** contém informações sobre o número de habitantes, a área do país, renda per capita, entre outros. No entanto, para fins didáticos, estamos usando o dado tratado e limpo, deste repositório: **Machine Learning aplicado a dados espaciais**. A pergunta que queremos responder neste agrupamento é: Será que conseguimos obter grupos de países que possuem características semelhantes? Tudo indica que sim, né? Mas, vamos ver que essa tarefa não é nada trivial!



<https://dataat.github.io/introducao-ao-machine-learning/>

Exercício de análise de dados



13 de Agosto de 2021
São José dos Campos - SP

Adriano Almeida
Felipe Carvalho
Felipe Menino