Introduction to machine learning

Mardônio França / Vinicius Sampaio

boitatá lab

Machine Learning

Machine learning (ML) is the study of computer algorithms that improve automatically through experience. (wikipedia).

Machine learning is about extracting knowledge from data. It is a research field at the intersection of statistics, artificial intelligence, and computer science and is also known as predictive analytics or statistical learning

about boitatá lab

boitatá lab is a group of research about data science.

about mardônio França

data scientist, poet

about vinicius sampaio

data scientist

publications

- Temporal Analysis and Visualisation of Music - Bracis 2020

- A Brief Survey of Deep Learning based methods, against OpenNLP NameFinder for Named Entity Recognition on Portuguese Literary Texts - STIL 2019

- Atomic Force Microscopy (AFM) feature extraction and fault diagnosis using Clustering Algorithm based on Minimum Spanning Tree (MST) - xxxiv-encontro-de-fisicos-do-norte-e-nordeste

Temporal Analysis and Visualisation of Music

Luan Misael Gomes de Moura [1,5], **Carlos Henrique Quartucci Forster**[2], **Emanuel P. Fontelles**[3,5], **Vinicius A. Sampaio**[4,5], **Mardônio J. C. França**[5]

[1]Itaú Unibanco

[2]Instituto Tecnológico de Aeronáutica

[3]Universidade Federal do Ceará

[4]Universidade Estadual do Ceará

[5]Boitatá Lab

luan.moura@itau-unibanco.com.br, forster@ita.br,

emanuelfontelles@fisica.ufc.br, vinicius.sampaio@aluno.uece.br,

mardfranca@gmail.com

Abstract. *This paper proposes a temporal analysis for music metadata using a generative probabilistic model for collections the discrete datasets such as text corpora. This method is also a topic model that is used for discovering abstract topics from a collection of documents. The method is then applied to audio metadata and song lyrics extracted with Echo Nest® engine, Spotify® Lyrics Genius® API. Song data time series are generated by grouping data items by release date, genre and dominant topics (from LDA analysis). Using a technique from Network Theory we visualise how these topics, in this case, genres, are related to each other through time.*

2020

A Brief Survey of Deep Learning based methods against OpenNLP NameFinder for Named Entity Recognition on Portuguese Literary Texts

1st Vinicius Amaro Sampaio
Ceará State University
Fortaleza, Brazil
vinicius.sampaio@aluno.uece.br

2nd Mardônio J. C. França
Casa Magalhães
Fortaleza, Brazil
mardfranca@gmail.com

3rd Paulo Bruno Lopes da Silva
University of São Paulo
Fortaleza, Brazil
paulobruno.ls.fr@usp.br

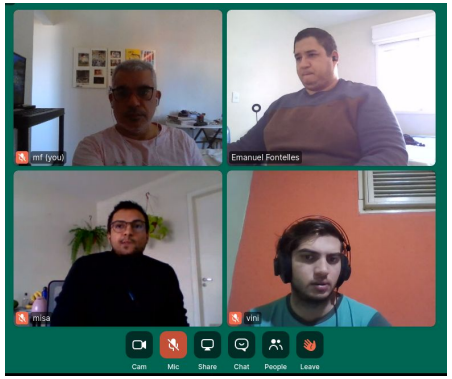
4th Gustavo Augusto Lima de Campos
Ceará State University
Fortaleza, Brazil
gustavo@larces.uece.br

5th Lara Domingos Hissa
Federal University of Ceará
Fortaleza, Brazil
larahissa@fisica.ufc.br

Atomic Force Microscopy (AFM) feature extraction and fault diagnosis using Clustering Algorithm based on Minimum Spanning Tree (MST)

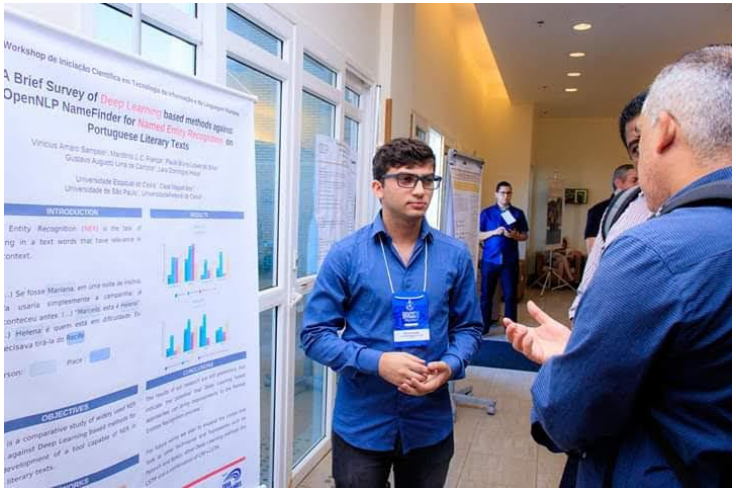
Lara D. Hissa, Vinicius A. Sampaio, Ramon R.R.X. Chaves, Daniel B. Araújo, Ludwing F. M. Camacho, Mardônio F. J. C. França
Federal University of Ceará, State University of Ceará, Casa Magalhães

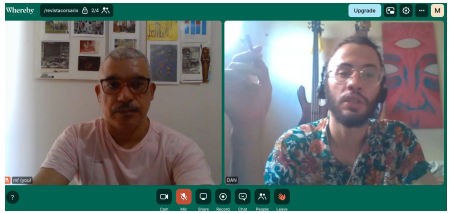
2019


mf (you)
Emanuel Fontelles
misa
vini
Cam
Mic
Share
Chat
People
Leave

2020

2019


Workshop de Iniciação Científica
A Brief Survey of Deep Learning based methods against OpenNLP NameFinder for Named Entity Recognition on Portuguese Literary Texts
INTRODUCTION
OBJECTIVES


Wherebe
Upgrade

Introduction to machine learning

Introduction to machine learning

Mardônio França / Vinicius Sampaio

boitatá lab

Topics

1. Introduction
2. Supervised Learning
3. Unsupervised Learning and Preprocessing
4. Representing Data and Engineering Features
5. Model Evaluation and Improvement
6. Algorithm Chains and Pipelines
7. Working with Text Data
8. Wrapping Up

Introduction

Introduction

1 - Why machine learning ?

2 - Python

3 - Essential Libraries

- 3 - 1 - Jupyter notebook
- 3 - 2 - numpy
- 3 - 3 - pandas \ pyspark
- 3 - 4 - matplotlib
- 3 - 5 - scikit-learn

Introduction

4 - A first application : classifying Iris Species

- 4- 1 - Meet the Data
- 4 - 2 - Measuring Success: Training and Testing Data
- 4 - 3 - First Things First : Look at Your Data
- 4 - 4 - Building Your First Model: k-Nearest Neighbors
- 4 - 5 - Making Predictions
- 4 -6 - Evaluating the Model

numpy

```
import numpy as np

x = np.array([[1,2,3],[4,5,6]])
```

scipy

```
from scipy import sparse

eye = np.eye(4)

sparse_matrix = sparse.csr_matrix(eye)
```

matplotlib

```
%matplotlib inline

import matplotlib.pyplot as plt

x = np.linspace(-10, 10, 100)

y = np.sin(x)

plt.plot(x, y, marker="x")
```

pandas

```
import pandas as pd

data = {'Name': ["John", "Anna", "Peter", "Linda"],

 'Location' : ["New York", "Paris", "Berlin", "London"],

 'Age' : [24, 13, 53, 33]

 }

data_pandas = pd.DataFrame(data)
```

Thank you !