



DCA

PPgEEC

Machine Learning

Course Outline

A blurred, colorful night cityscape with light streaks from traffic and buildings.

#tenyearschallenge



PLuto

April 4, 2022

10:34:25 AM
PM

1994





2002



2022

1996 - 2020

TOMB RAIDER



PS1

PS2

PS3

PS4

PS5

2010



2022



2018

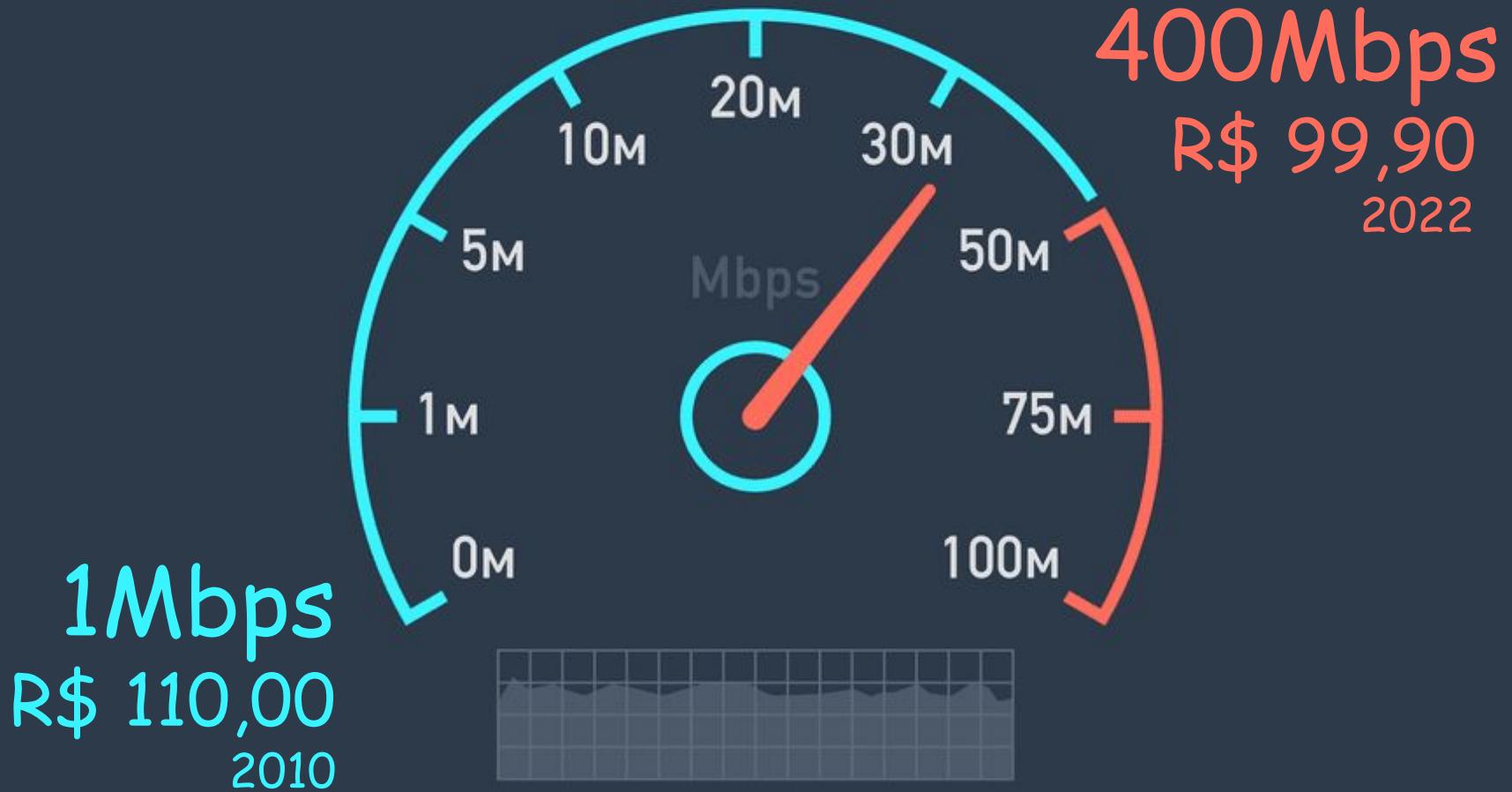


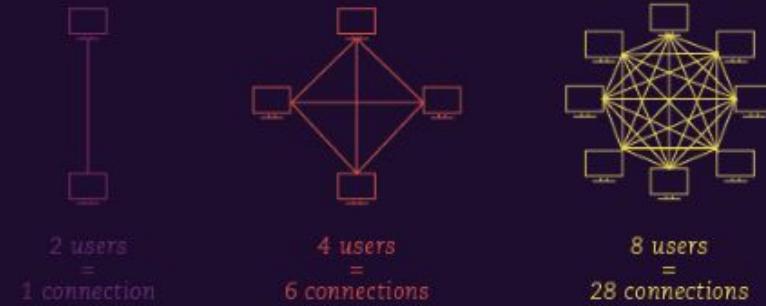
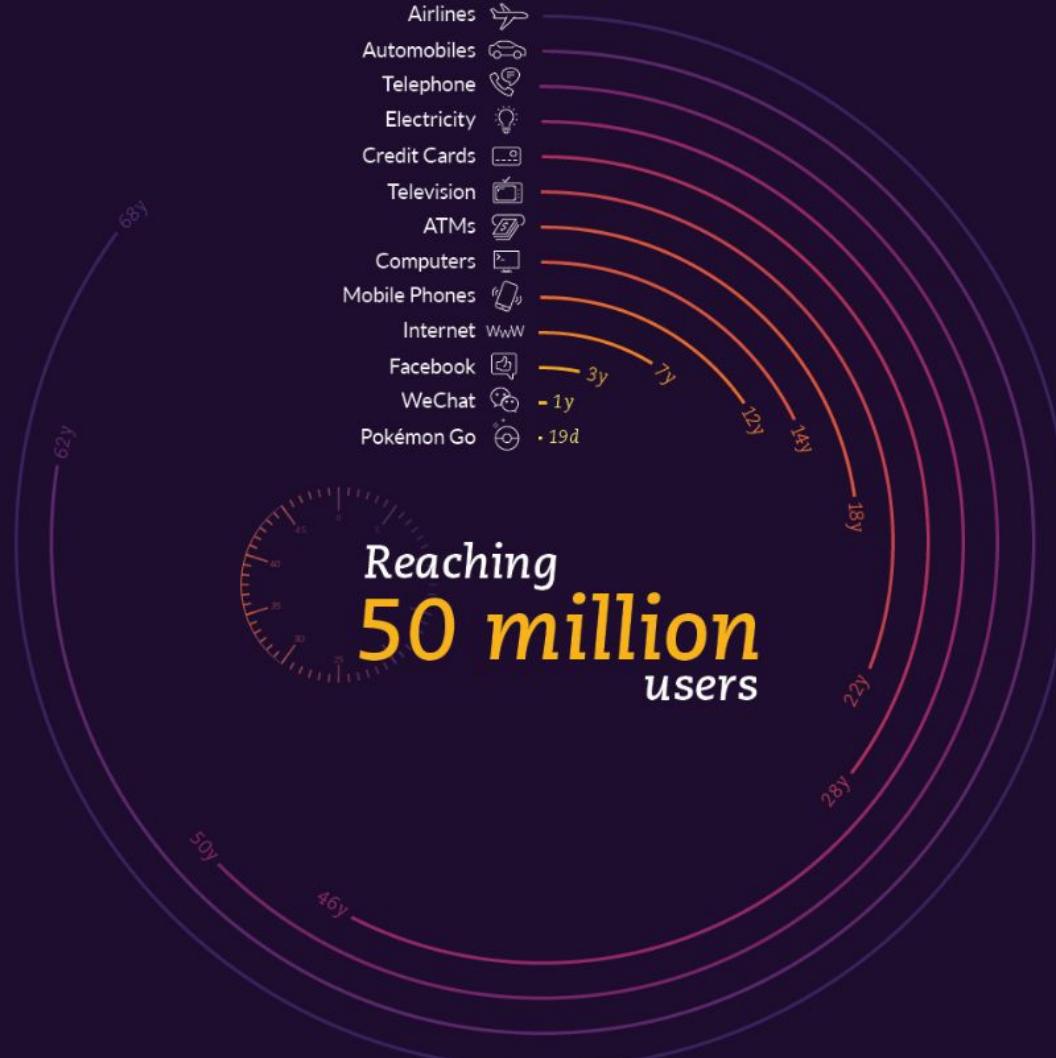
2010



2022



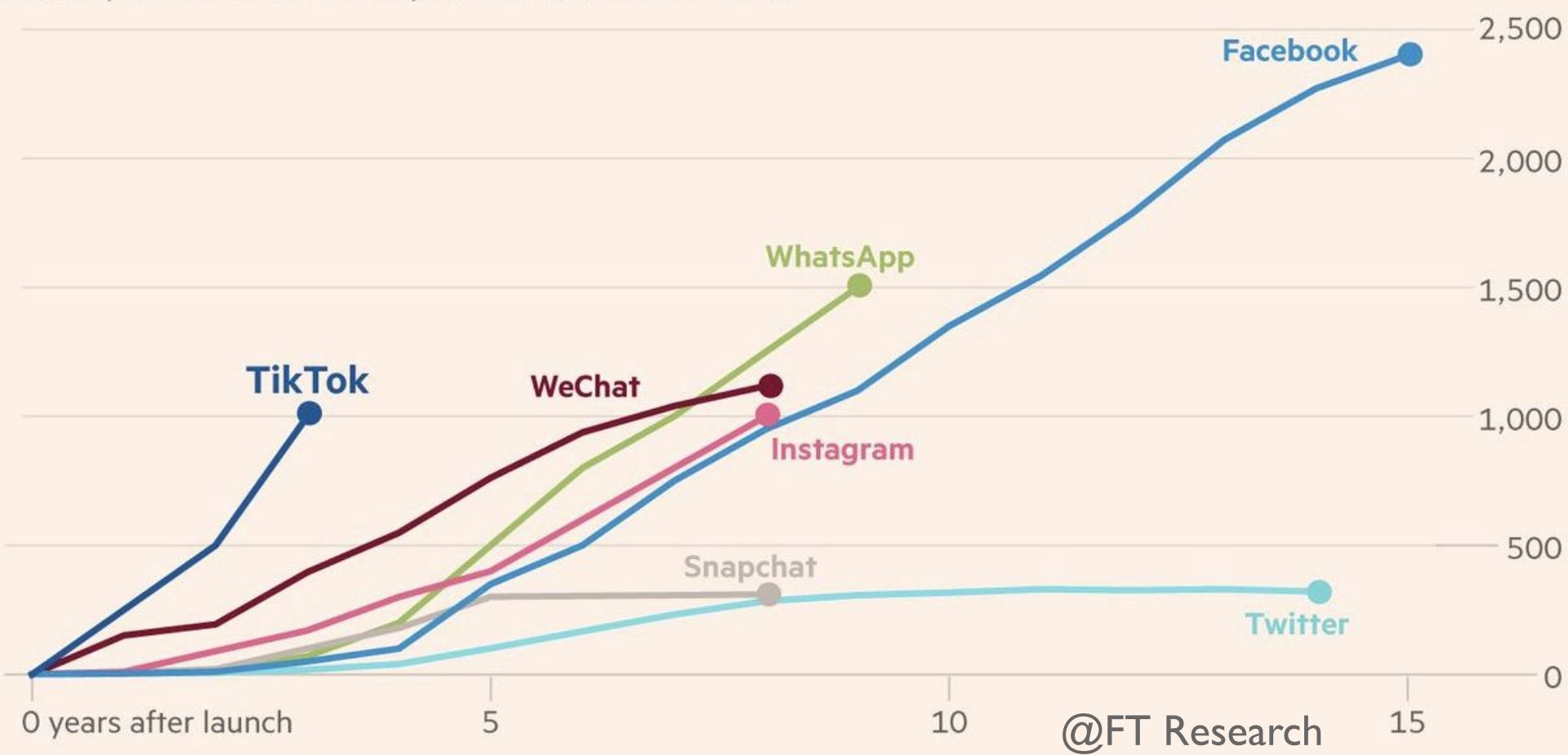




The impact of moving to digital and the power of network effects

TikTok has reached 1bn users faster than any other social media app

Monthly active users since product launch (millions)



2020 | 73 days

2010 3-5 years

1980 7 years

1950 50 years

EVOLUTION Of Medical Knowledge



Discovering temporal scientometric knowledge in COVID-19 scholarly production

Breno Santana Santos^{1,2} · Ivanovitch Silva¹ · Luciana Lima³ ·
Patricia Takako Endo⁴ · Gislian Alves¹ · Marcel da Câmara Ribeiro-Dantas⁵

Received: 7 October 2021 / Accepted: 22 December 2021 / Published online: 16 January 2022

© Akadémiai Kiadó, Budapest, Hungary 2022

Abstract

The mapping and analysis of scientific knowledge makes it possible to identify the dynamics and/or growth of a particular field of research or to support strategic decisions related to different research entities, based on bibliometric and/or scientometric indicators. However, with the exponential growth of scientific production, a systematic and data-oriented approach to the analysis of this large set of productions becomes increasingly essential. Thus, in this work, a data-oriented methodology was proposed, combining Data Analysis, Machine Learning and Complex Network Analysis techniques, and Data Version Control (DVC) tool, for the extraction of implicit knowledge in scientific production bases. In addition, the approach was validated through a case study in a COVID-19 manuscripts dataset, which had 199,895 articles published on arXiv, bioRxiv, medRxiv, PubMed and Scopus databases. The results suggest the feasibility of the proposed methodology, indicating the most active countries and the most explored themes in each period of the pandemic. Therefore, this study has the potential to instrument and expand strategic decisions by the scientific community, aiming at extracting knowledge that supports the fight against the COVID-19 pandemic.

More than 200k
COVID-19
manuscripts
published!!!!



\$0.00
2010



\$46,551.30
2022



NVDA



M



Indicadores



Alerta

Replay



Salvar



Publicar

Abr 273.75 Máx. 275.58 Mín. 262.67 Fch 273.60 +0.74 (+0.27%)

273.60 0.00 273.60

USD

350.00

325.00

300.00

273.60
24d 10h

250.00

225.00

200.00

175.00

150.00

125.00

100.00

75.00

50.00

25.00

0.00

-25.00

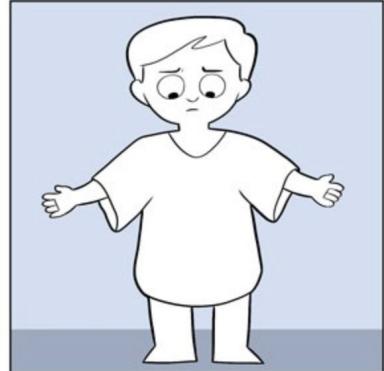
2000 2002 2004 2006 2008 2010 2012 2014 2016 2018 2020 2022



2022

MTAILOR

????



CollegeHumor

Coding in 2022

send_tweet.py

10 |

11

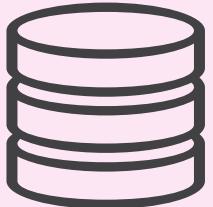
12

13

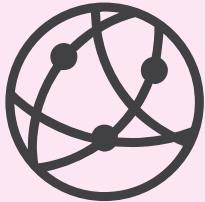
14

15

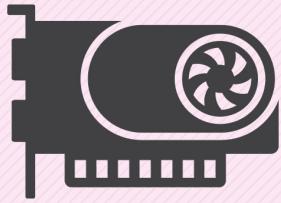
What do these
techs have in
common?



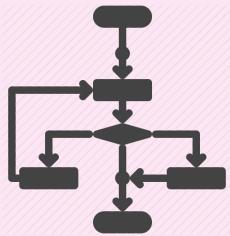
DATA



INTERNET



HARDWARE

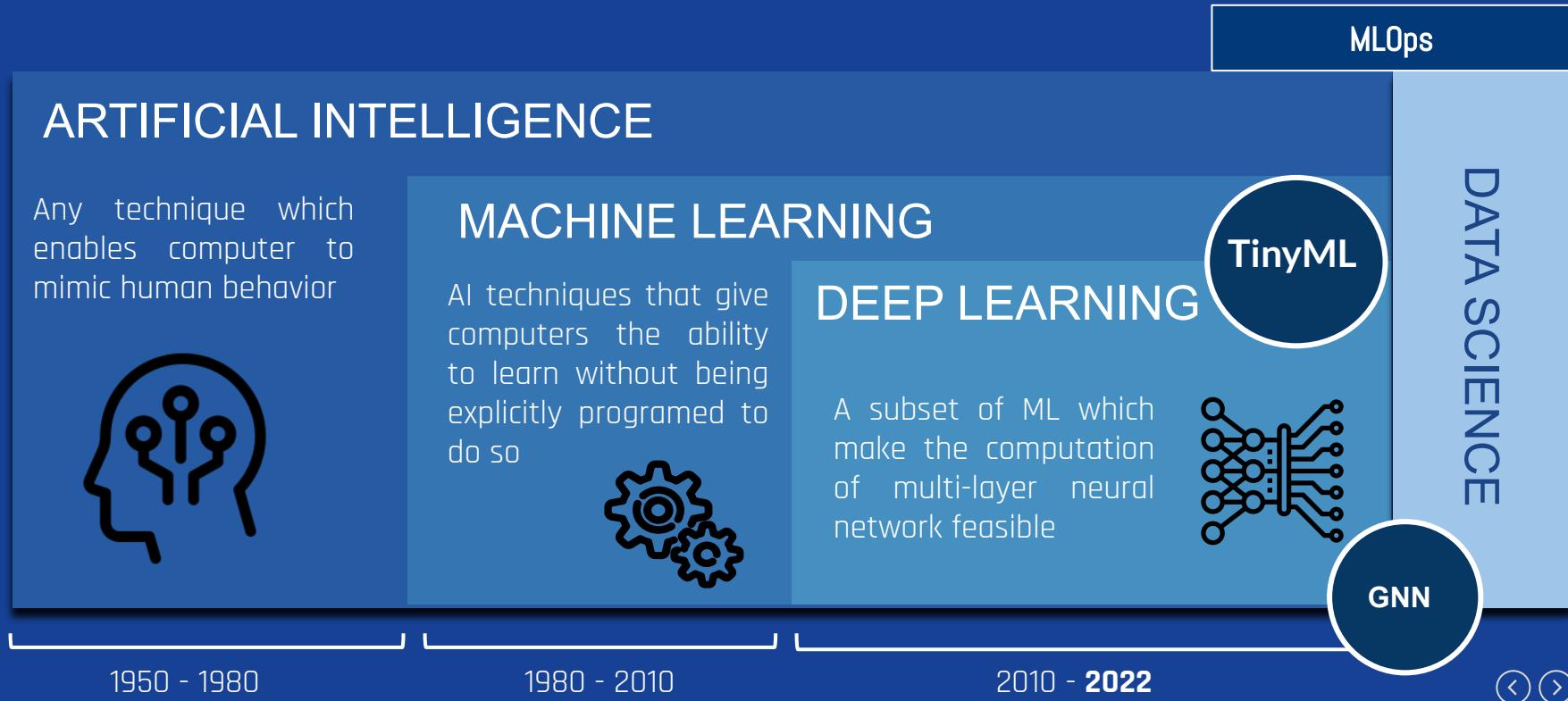


TOOLS &
ALGORITHMS



{AI}

ARTIFICIAL INTELLIGENCE



A Long Pathway

Vector & Matrices

Matrices & Vector Arithmetics
Types, Operations
Factorization

Calculus

Derivatives

@ivanovitchm/imd0033_2019_1

Linear Algebra & Math

Probability & Statistics

Probability
Conditional Probability
Distributions
Bayesian Probability

Statistics
Data Viz, Central Limit Theorem
Hypothesis Tests, Correlation
Resampling Methods

Exploratory Data Analysis

Measurements of Centrality (mean, mode, median, variance, std, z-score)

Data Pipeline

Collect, clean, preparation, model, analysis, interpretation, viz
Deploy, monitoring solution

Data Science

@ivanovitchm/datascience2020.6

Machine Learning

Supervised Learning
KNN, Linear regression, Logistic Regression, Decision Tree, Random Forest, Ensemble, XGBoost, MLP

Unsupervised Learning

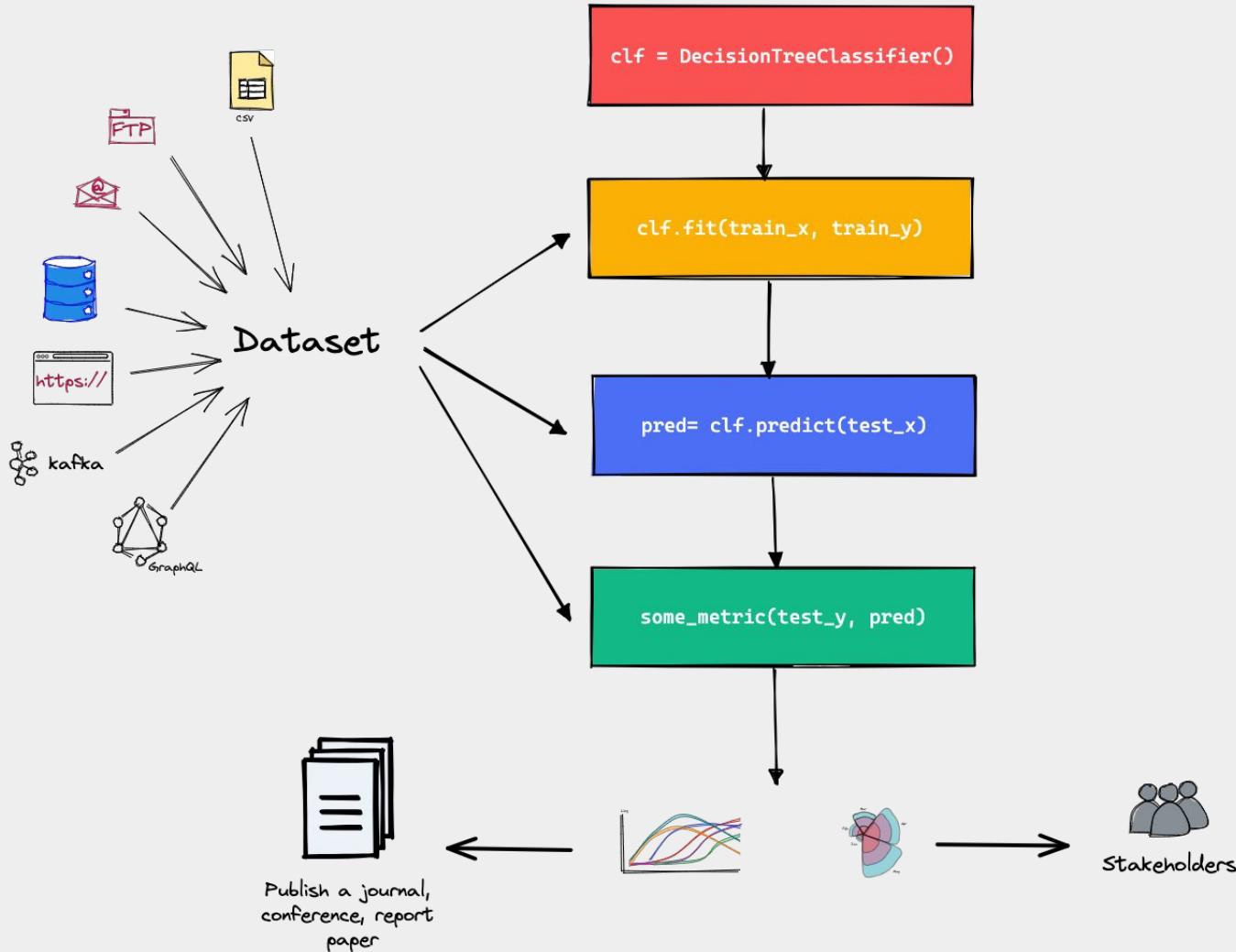
K-Means, PCA

Deep Learning

Fundamentals of Deep Learning
Better Generalization vs Better Learning
Hyperparameter tuning
Batch normalization
Convolutional Neural Networks
Transfer Learning

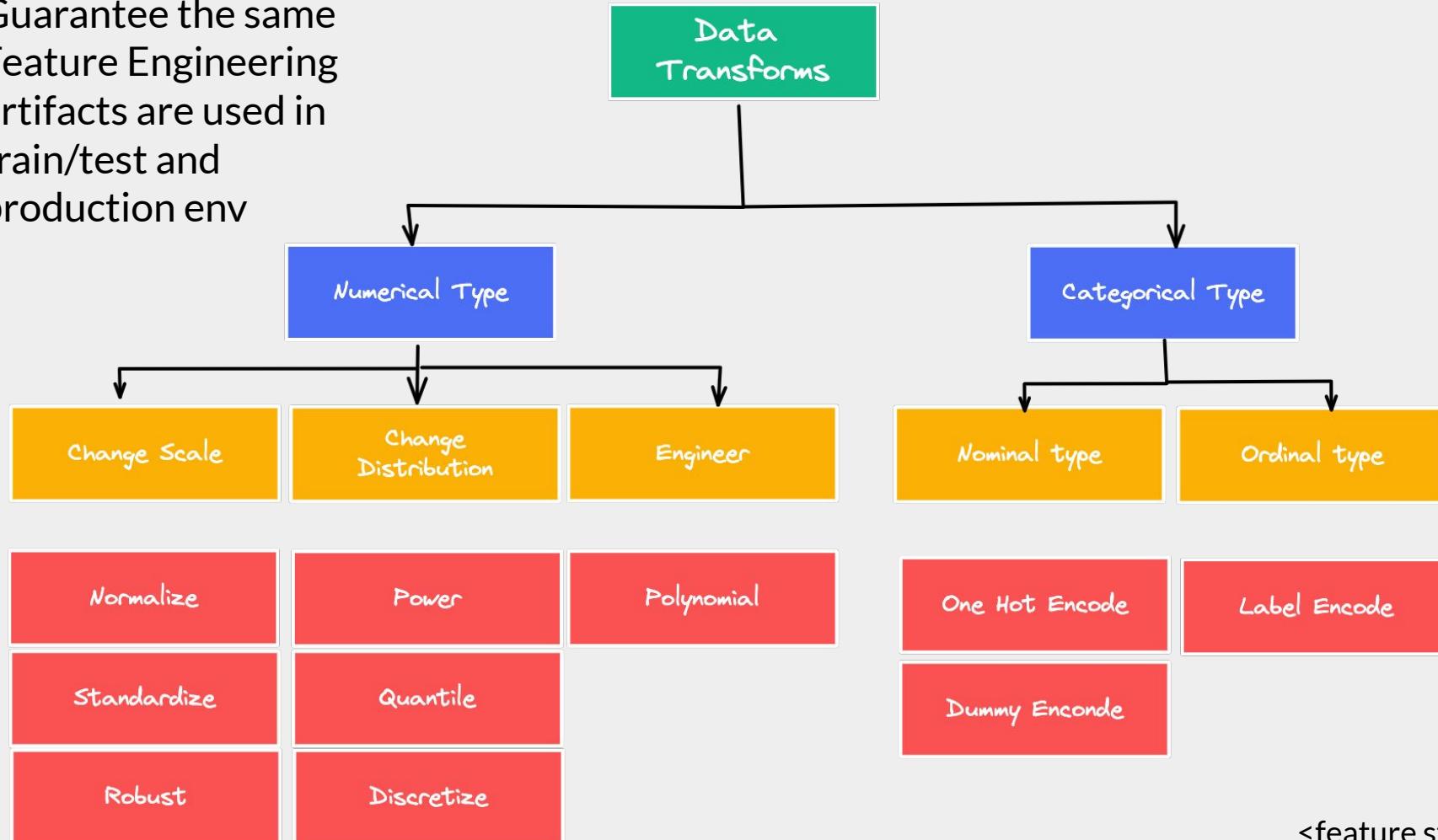


A typical ML workflow





Guarantee the same
Feature Engineering
artifacts are used in
train/test and
production env



<feature store>

Article

Predictive Models for Imbalanced Data: A School Dropout Perspective

Thiago M. Barros ^{1,*†}, Plácido A. Souza Neto ^{1,†} and Ivanovitch Silva ^{2,†}
and Luiz Affonso Guedes ^{2,†}

¹ Federal Institute of Rio Grande do Norte (IFRN), 1559 Tirol Natal, Brazil; placido.neto@

² Federal University of Rio Grande do Norte (UFRN), 59078-970 Natal, Brazil; ivan@imd.
affonso@dca.ufrn.br (L.A.G.)

* Correspondence: thiago.medeiros@ifrn.edu.br

† These authors contributed equally to this work.

Received: 24 July 2019; Accepted: 10 November 2019; Published: 15 November 2019

Abstract: Predicting school dropout rates is an important issue for the smooth functioning of the educational system. This problem is solved by classifying students into two classes using activities related statistical datasets. One of the classes must identify the students who have the tendency to persist. The other class must identify the students who have the tendency to dropout. This problem often encounters a phenomenon that masks out the obtained results. This study delves into this phenomenon and provides a reliable educational data mining technique that accurately predicts the dropout rates. In particular, the three data classifying techniques, namely, decision tree, neural networks and Balanced Bagging, are used. The performances of these classifiers are tested with and without the use of a downsample, SMOTE and ADASYN data balancing. It is found that among other parameters geometric mean and UAR provides reliable results while predicting the dropout rates using Balanced Bagging classifying techniques.

Keywords: dropout rates; accuracy paradox; imbalanced learning; downsample; g-mean predict; mlp; decision tree; Balanced Bagging; UAR; SMOTE; ADASYN



Evasão escolar de crianças e adolescente aumenta 171% na pandemia, diz estudo

Levantamento da organização Todos Pela Educação mostra que 244 mil crianças de 6 a 14 anos estavam fora da escola no segundo trimestre de 2021.

Por g1 São Paulo

Publicado há 3 meses



Concept/Data Drift

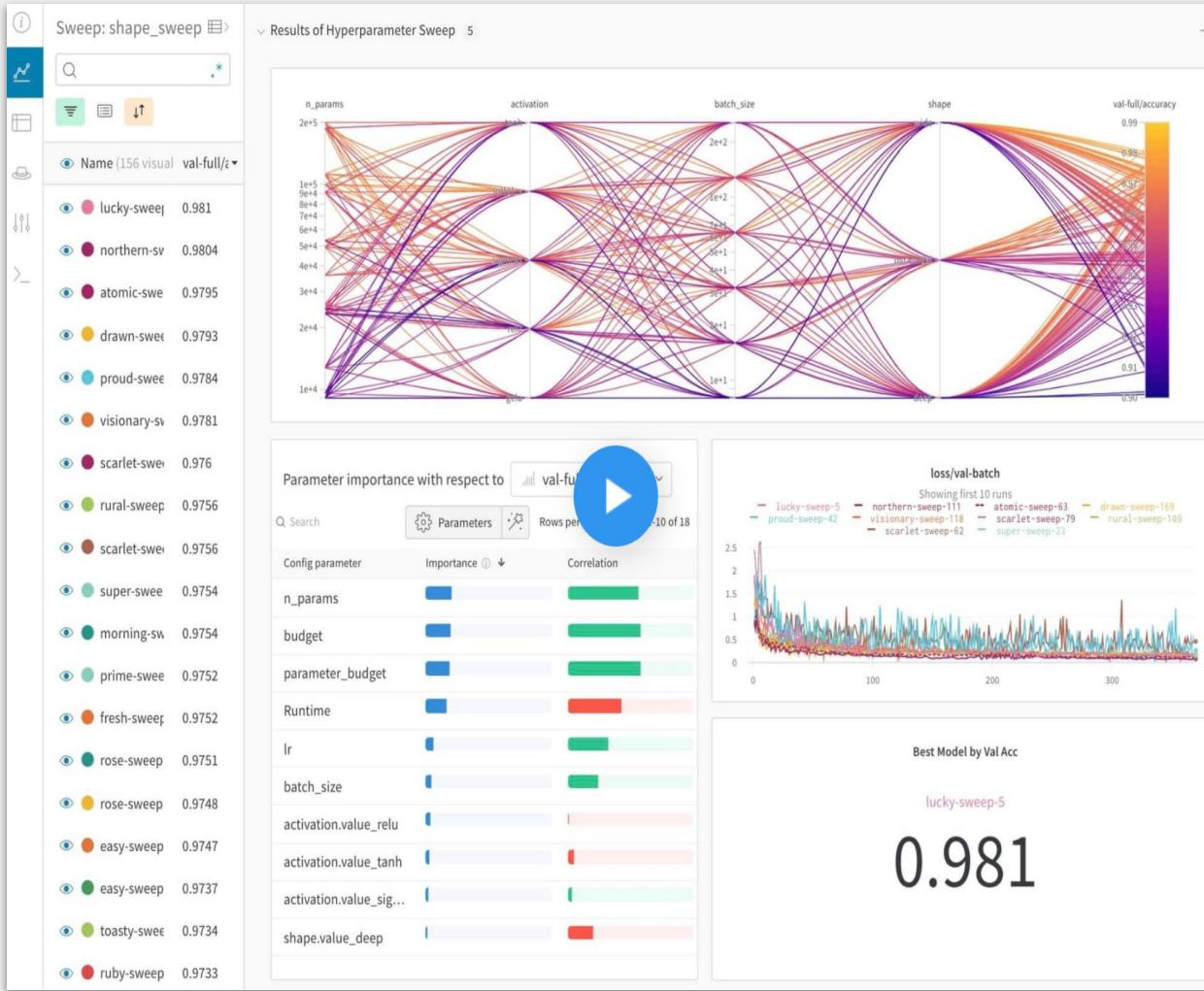
Mais de 650 mil crianças saíram da escola durante a pandemia

Pela primeira vez desde 2005, país registrou queda de matrículas na educação infantil

PERSPECTIVA 2022

Evasão escolar bate recordes durante a pandemia

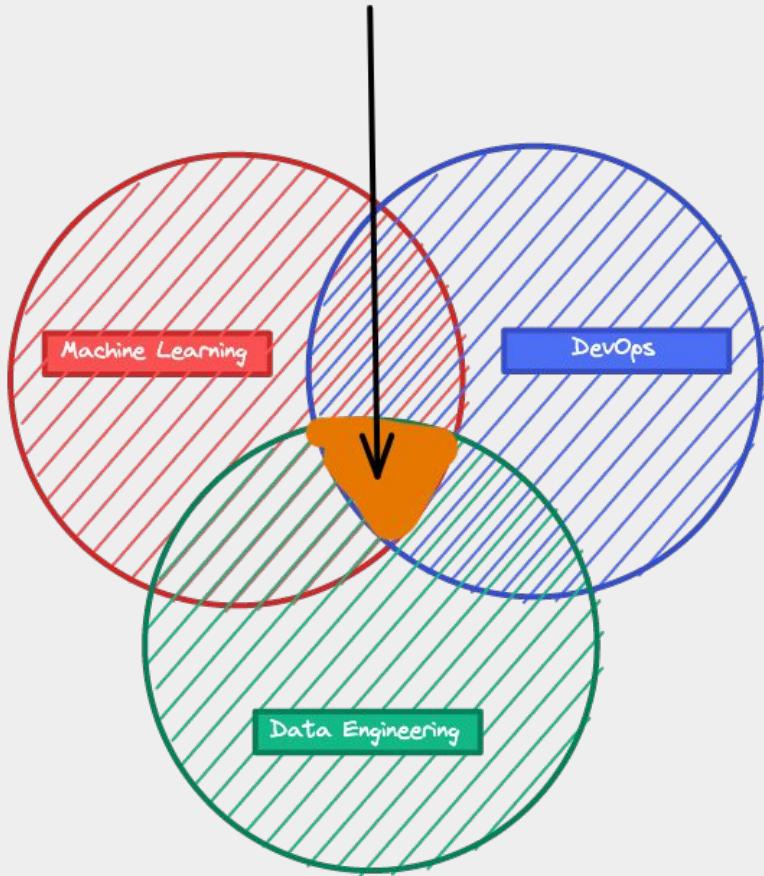
Colégios abertos, mas sem alunos. Com cerca de 240 mil estudantes fora das salas de aula a desistência é o maior desafio de 2022. Especialistas afirmam que esse é o pior cenário em 20 anos



What settings were used in the last experiment?

experiment tracking
dataset versioning
model management

MLOps



Computer

ARTIFICIAL INTELLIGENCE AND SOFTWARE ENGINEERING: ARE WE READY?



IEEE

IEEE
COMPUTER
SOCIETY

vol. 55 no. 3

www.computer.org/computer

S. Warnett and U. Zdun, "Architectural Design Decisions for the Machine Learning Workflow" in **Computer**, vol. 55, no. 03, pp. 40-51, 2022. doi: 10.1109/MC.2021.3134800

H. Washizaki, et al., "Software-Engineering Design Patterns for Machine Learning Applications" in **Computer**, vol. 55, no. 03, pp. 30-39, 2022. doi: 10.1109/MC.2021.3137227

A. Mashkoor, T. Menzies, A. Egyed and R. Ramler, "Artificial Intelligence and Software Engineering: Are We Ready?" in **Computer**, vol. 55, no. 03, pp. 24-28, 2022.
doi: 10.1109/MC.2022.3144805

R. Sangwan, Y. Badr, S. Srinivasan and P. Mukherjee, "On the Testability of Artificial Intelligence and Machine Learning Systems" in **Computer**, vol. 55, no. 03, pp. 101-105, 2022.
doi: 10.1109/MC.2021.3132710

DATA VIOLENCE

How Bad
Engineering
Choices Can
Damage Society



Por que a demissão de pesquisadora negra do Google se transformou em escândalo global

O silenciamento e a saída de Timnit Gebru geram novas dúvidas sobre o compromisso das grandes empresas de tecnologia com seus propósitos éticos



Jeff Dean (@)
@JeffDean

On behalf of the entire Google Research & @GoogleAI communities, I'm excited to share an overview of some of our research in 2020.

Thanks to everyone who helped make this work possible!



Google Research: Looking Back at 2020, and Forward to 2...
Posted by Jeff Dean, Senior Fellow and SVP of Google Research and Health, on behalf of the entire Google ...
ai.googleblog.com

5:40 PM · Jan 12, 2021 · Twitter Web App

333 Retweets 82 Quote Tweets 1.5K Likes

Model Cards for Model Reporting

Margaret Mitchell, Simone Wu, Andrew Zaldivar, Parker Barnes, Lucy Vasserman, Ben Hutchinson, Elena Spitzer, Inioluwa Deborah Raji, Timnit Gebru
 {mmitchellai,simonewu,andyzaldivar,parkerbarnes,lucyvasserman,benhutch,espitzer,tgebru}@google.com
 deborah.raji@mail.utoronto.ca

ABSTRACT

Trained machine learning models are increasingly used to perform high-impact tasks in areas such as law enforcement, medicine, education, and employment. In order to clarify the intended use cases of machine learning models and minimize their usage in contexts for which they are not well suited, we recommend that released models be accompanied by documentation detailing their performance characteristics. In this paper, we propose a framework that we call model cards, to encourage such transparent model reporting. Model cards are short documents accompanying trained machine learning models that provide benchmarked evaluation in a variety of conditions, such as across different cultural, demographic, or phenotypic groups (e.g., race, geographic location, sex, Fitzpatrick skin type [15]) and intersectional groups (e.g., age and race, or sex and Fitzpatrick skin type) that are relevant to the intended application domains. Model cards also disclose the context in which models are intended to be used, details of the performance evaluation procedures, and other relevant information. While we focus primarily on human-centered machine learning models in the application fields of computer vision and natural language processing, this framework can be used to document any trained machine learning model. To solidify the concept, we provide cards for two supervised models: One trained to detect smiling faces in images, and one trained to detect toxic comments in text. We propose model cards as a step towards the responsible democratization of machine learning and related artificial intelligence technology, increasing transparency into how well artificial intelligence technology works. We hope this work encourages those releasing trained machine learning models to accompany model releases with similar detailed evaluation numbers and other relevant documentation.

CCS CONCEPTS

- General and reference → Evaluation; • Social and professional topics → User characteristics; • Software and its engineering → Use cases; Documentation; Software evolution; • Human-centered computing; • Walkthrough evaluations

KEYWORDS

datasheets, model cards, documentation, disaggregated evaluation, fairness evaluation, ML model evaluation, ethical considerations

ACM Reference Format:

Margaret Mitchell, Simone Wu, Andrew Zaldivar, Parker Barnes, Lucy Vasserman, Ben Hutchinson, Elena Spitzer, Inioluwa Deborah Raji, Timnit Gebru. 2019. Model Cards for Model Reporting. In *FAT* '19: Conference on Fairness, Accountability, and Transparency*, January 29–31, 2019, Atlanta, GA, USA. ACM, New York, NY, USA, 10 pages. <https://doi.org/10.1145/3287560.3287596>

1 INTRODUCTION

Currently, there are no standardized documentation procedures to communicate the performance characteristics of trained machine learning (ML) and artificial intelligence (AI) models. This lack of documentation is especially problematic when models are used in applications that have serious impacts on people's lives, such as in health care [14, 42, 44], employment [1, 13, 29], education [23, 45] and law enforcement [2, 7, 20, 34].

Researchers have discovered systematic biases in commercial machine learning models used for face detection and tracking [4, 9, 49], attribute detection [5], criminal justice [10], toxic comment detection [11], and other applications. However, these systematic errors were only exposed after models were put into use, and negatively affected users reported their experiences. For example, after MIT Media Lab graduate student Joy Buolamwini found that commercial face recognition systems failed to detect her face [4], she collaborated with other researchers to demonstrate the disproportionate errors of computer vision systems on historically marginalized groups in the United States, such as darker-skinned women [5, 41]. In spite of the potential negative effects of such reported biases, documentation accompanying trained machine learning models (if supplied) provide very little information regarding model performance characteristics, intended use cases, potential pitfalls, or other information to help users evaluate the suitability of these systems to their context. This highlights the need to have detailed documentation accompanying machine learning models, including





+100k papers



COVID-19: A scholarly production dataset report for research analysis

Breno Santana Santos^{a,b,*}, Ivanovitch Silva^a, Marcel da Câmara Ribeiro-Dantas^c, Gislainy Alves^a, Patricia Takako Endo^d and Luciana Lima^a

^a Universidade Federal do Rio Grande do Norte (UFRN), Rio Grande do Norte, Brazil

^b Núcleo de Pesquisa e Prática em Inteligência Competitiva (NUPIC), Universidade Federal de Sergipe (UFS), Itabaiana/SE, Brazil

^c Institut Curie (UMR168), Sorbonne Université (EDITE), Paris, France

^dUniversidade de Pernambuco (UPE), Pernambuco, Brazil

ARTICLE INFO

Keywords:
COVID-19
SARS-CoV-2
Pandemic
Data Science
Bibliometrics
Scientometrics

ABSTRACT

COVID-19 has been recognized as a global threat, and several studies are being conducted in order to contribute to the fight and prevention of this pandemic. This work presents a scholarly production dataset focused on COVID-19, providing an overview of scientific research activities, making it possible to identify countries, scientists and research groups most active in this task force to combat the coronavirus disease. The dataset is composed of 40,212 records of articles' metadata collected from Scopus, PubMed, arXiv and bioRxiv databases from January 2019 to July 2020. Those data were extracted by using the techniques of Python Web Scraping and pre-processed with Pandas Data Wrangling. In addition, the pipeline to preprocess and generate the dataset are versioned with the Data Version Control tool (DVC) and are thus easily reproducible and auditable.



Brazil is one of the most densely populated countries in the world. The outbreak has affected more than 600,000 people and put the country on the front line of the global pandemic. As the outbreak continues to spread, the health and socioeconomic reforms of the president and his government have been criticised for being overly harsh. This analysis attempts to understand the reasons behind the policies and why they are being so harshly criticised, and how the institutional changes and the administration have been ineffective in dampening the disease. In particular, the reasons for the policies are discussed. It is argued that the policies are overly harsh not only because of the lack of economic growth but also because of the lack of social and health security, making it difficult to pay the healthcare bill. The authors conclude that the policies are counterproductive and the policies need to be reformed. This study provides a framework for analysing the policies of the government and the subsequent failure in their implementation.



Created by GPT-2

Who would be the author of that text?



100%
Unique

0%
Plagiarism

What if?

Instituto Nacional da Propriedade Industrial

Buscar no Site



MARCAS

PATENTES

DESENHOS
INDUSTRIALIS

INDICAÇÕES
GEOGRÁFICAS

PROGRAMAS DE
COMPUTADOR

TOPOGRAFIAS
DE CIRCUITOS
INTEGRADOS

CONTRATOS DE
TECNOLOGIA E
DE FRANQUIA

ACADEM
DO INP

Alguma dúvida?



Chatbot

Alguma dúvida?
Chatbot do INPI
DISPONÍVEL 24/7

DE
MAIS

VALOR

À SUA
CRIAÇÃO!

DE
MAIS

State of the Art

*non traditional





arXiv.org

[Login](#)

Search...

All fields



Search

[Help | Advanced Search](#)

arXiv is a free distribution service and an open-access archive for 1,833,415 scholarly articles in the fields of physics, mathematics, computer science, quantitative biology, quantitative finance, statistics, electrical engineering and systems science, and economics. Materials on this site are not peer-reviewed by arXiv.

Subject search and browse:

Physics

[Search](#)[Form Interface](#)[Catchup](#)**News**

Read about recent news and updates on [arXiv's blog](#). (View the former "what's new" pages here). Read [robots beware](#) before attempting any automated download.

COVID-19 Quick Links

See COVID-19 SARS-CoV-2 preprints from

- [arXiv](#)
- [medRxiv](#) and [bioRxiv](#)

Important: e-prints posted on arXiv are not peer-reviewed by arXiv; they should not be relied upon without context to guide clinical practice or health-related behavior and should not be reported in news media as established information without consulting multiple experts in the field.

Physics

- **Astrophysics (astro-ph new, recent, search)**
includes: [Astrophysics of Galaxies](#); [Cosmology and Nongalactic Astrophysics](#); [Earth and Planetary Astrophysics](#); [High Energy Astrophysical Phenomena](#); [Instrumentation and Methods for Astrophysics](#); [Solar and Stellar Astrophysics](#)
- **Condensed Matter (cond-mat new, recent, search)**
includes: [Disordered Systems and Neural Networks](#); [Materials Science](#); [Mesoscale and Nanoscale Physics](#); [Other Condensed Matter](#); [Quantum Gases](#); [Soft Condensed Matter](#); [Statistical Mechanics](#); [Strongly Correlated Electrons](#); [Superconductivity](#)
- **General Relativity and Quantum Cosmology (gr-qc new, recent, search)**
- **High Energy Physics – Experiment (hep-ex new, recent, search)**
- **High Energy Physics – Lattice (hep-lat new, recent, search)**
- **High Energy Physics – Phenomenology (hep-ph new, recent, search)**
- **High Energy Physics – Theory (hep-th new, recent, search)**
- **Mathematical Physics (math-ph new, recent, search)**
- **Nonlinear Sciences (nlin new, recent, search)**
includes: [Adaptation and Self-Organizing Systems](#); [Cellular Automata and Lattice Gases](#); [Chaotic Dynamics](#); [Exactly Solvable and Integrable Systems](#); [Pattern Formation and Solitons](#)



Papers With Code

<https://paperswithcode.com/>

Computer Vision



Semantic
Segmentation

80 benchmarks

1462 papers with code



Image
Classification

184 benchmarks

1275 papers with code



Object
Detection

299 benchmarks

1076 papers with code



Image
Generation

134 benchmarks

509 papers with code



Denoising

95 benchmarks

467 papers with code

[See all 965 tasks](#)

Medical



Medical Image
Segmentation

171 benchmarks

131 papers with code



Drug
Discovery

14 benchmarks

105 papers with code



Lesion
Segmentation

5 benchmarks

75 papers with code



Brain Tumor
Segmentation

7 benchmarks

44 papers with code



COVID-19
Diagnosis

40 papers with code

[See all 199 tasks](#)

@paperwithcode

Natural Language Processing



Machine
Translation

56 benchmarks

977 papers with code



Language
Modelling

19 benchmarks

962 papers with code



Question
Answering

66 benchmarks

863 papers with code



Sentiment
Analysis

50 benchmarks

584 papers with code



Text
Generation

49 benchmarks

416 papers with code

[See all 363 tasks](#)

Graphs



Link Prediction

52 benchmarks

304 papers with code



Node
Classification

61 benchmarks

249 papers with code



Graph
Embedding

1 benchmark

168 papers with code



Graph
Classification

46 benchmarks

139 papers with code



Community
Detection

12 benchmarks

107 papers with code

[See all 63 tasks](#)

@paperwithcode





TWO MINUTE
PAPERS | WHAT A TIME TO BE ALIVE!

Website



Two Minute Papers ✓

1.2M subscribers

JOIN

SUBSCRIBED



HOME

VIDEOS

PLAYLISTS

COMMUNITY

CHANNELS

ABOUT



DeepFake Detector AIs Are Good Too!

Two Minute Papers ✓ 1.1M views • 2 years ago

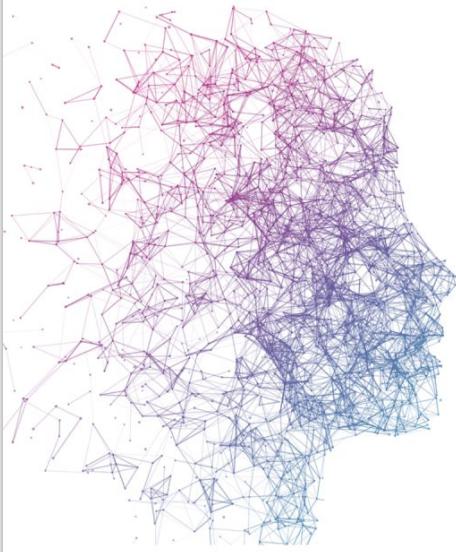
♥ Pick up cool perks on our Patreon page: <https://www.patreon.com/TwoMinutePapers> 🎉 The paper "FaceForensics++: Learning to Detect Manipulated Facial Images" is available here: <http://www....>





UFRN.ai

**EXPANDINDO OS HORIZONTES
DA INTELIGÊNCIA ARTIFICIAL**



INTELIGÊNCIA ARTIFICIAL

DISTRITO . REPORT 2021

REALIZAÇÃO:

DISTRITO

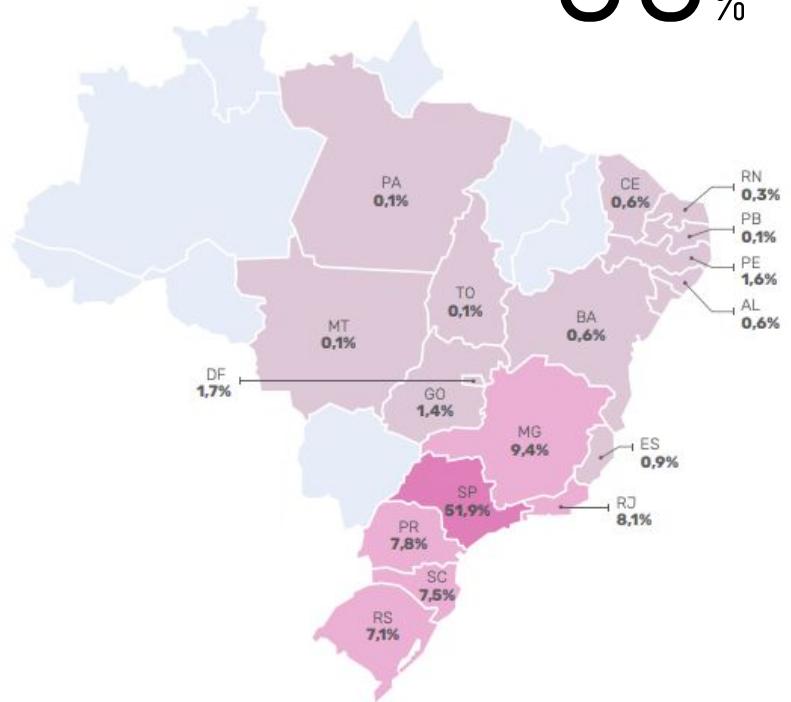
APOIO ESTRATÉGICO:



<https://distrito.me/dataminer/reports/>

AS STARTUPS DE INTELIGÊNCIA
ARTIFICIAL ESTÃO CONCENTRADAS NO
EIXO SUL-SUDESTE

93%



Contextualização

AMÉRICA DO NORTE

ESTADOS UNIDOS

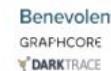


CANADÁ



EUROPA

REINO UNIDO



+50



ÁSIA

CHINA



JAPÃO



CINGAPURA



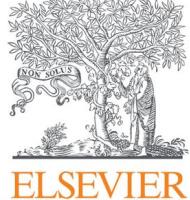
ISRAEL



Decacórnios



Fonte: Hurun Unicorn Index; Crunchbase; Tracxn



Pesquisador@s

Colaboração

Tópicos

Empresas

Dados não-estruturados



↑
Informação



Alphabet

facebook.



twitter

HIKVISION®

amazon

GoPro®
Be a HERO.

OpenAI

Uber

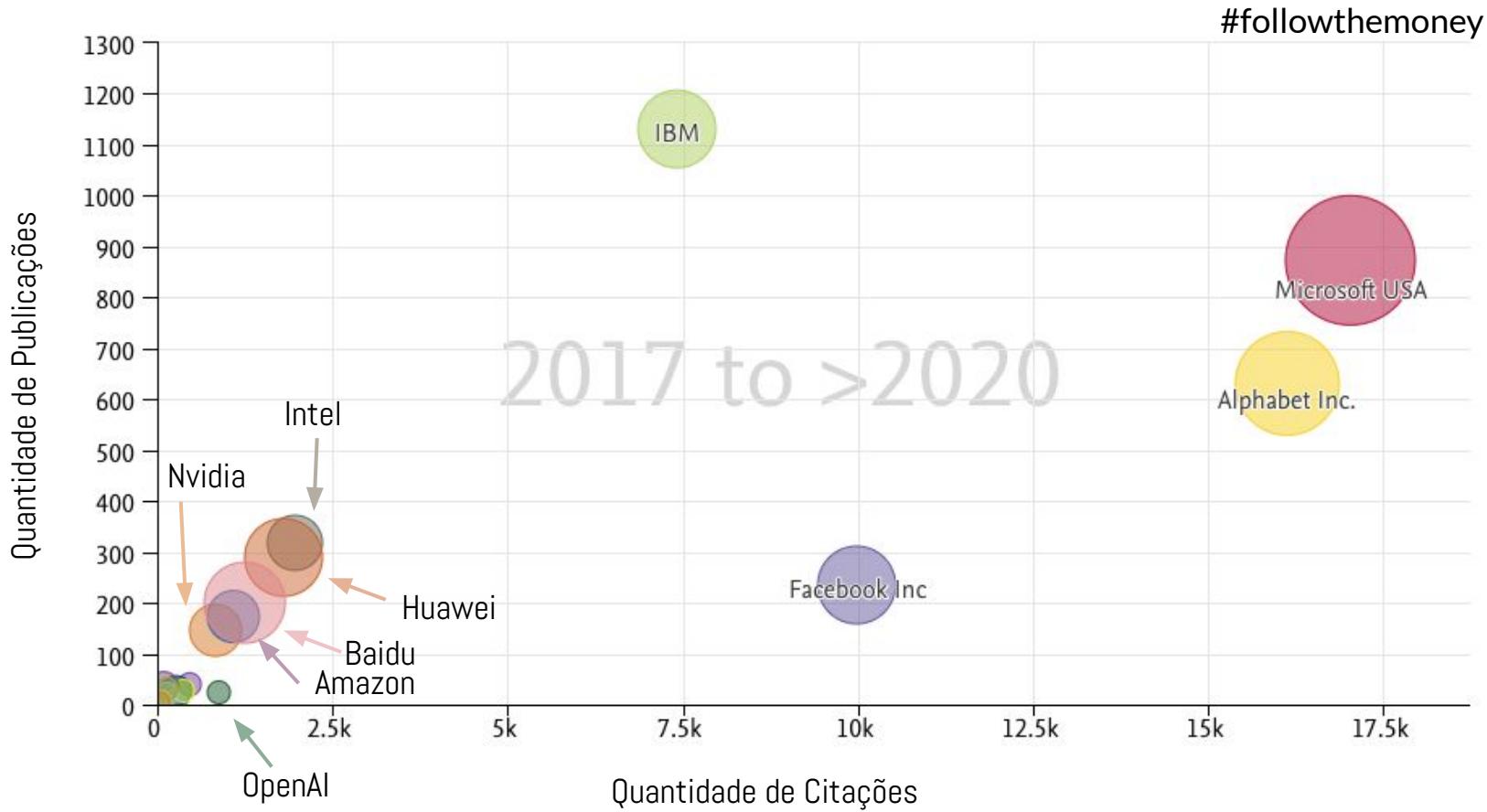


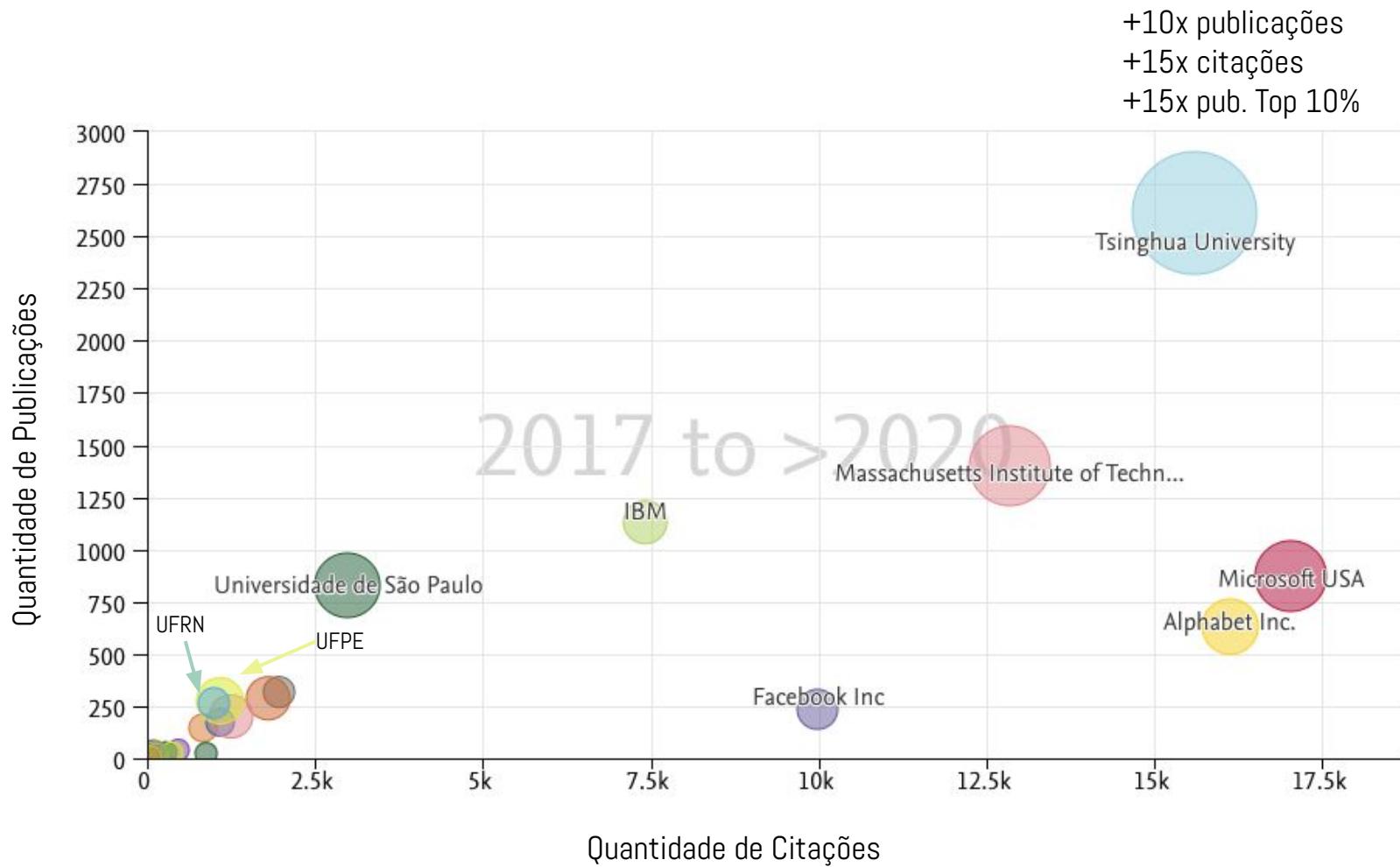
lyft

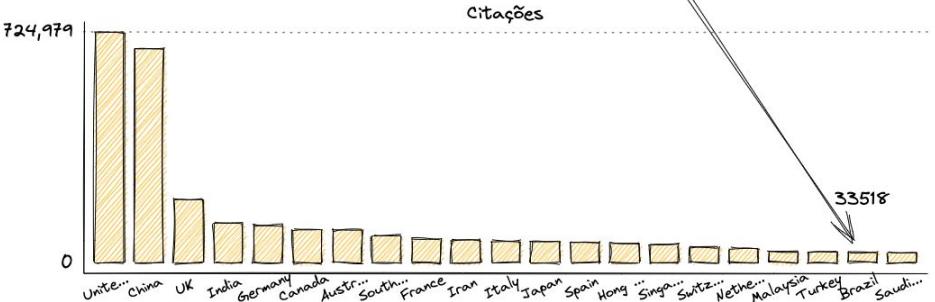
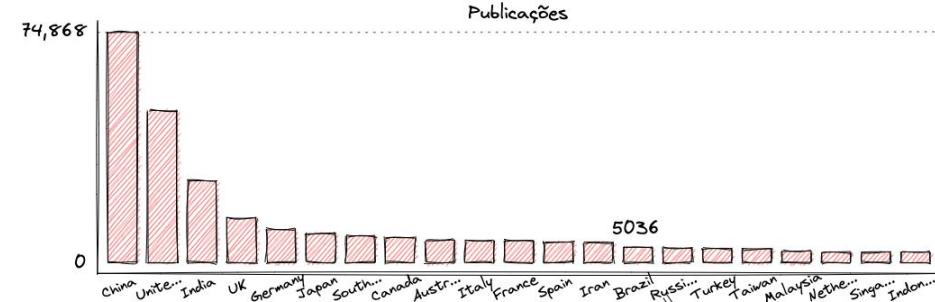
ORACLE

IBM

SMIC



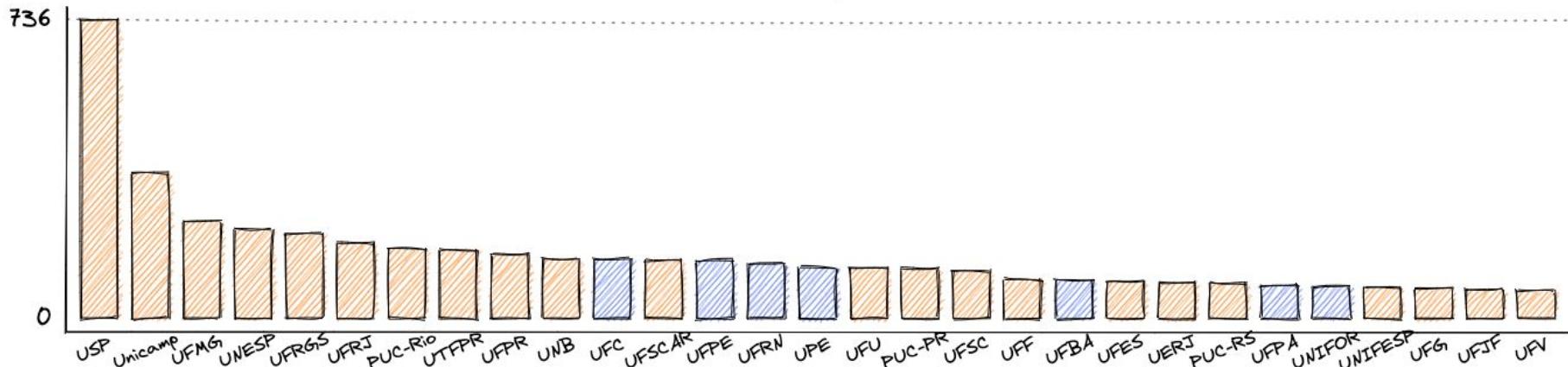




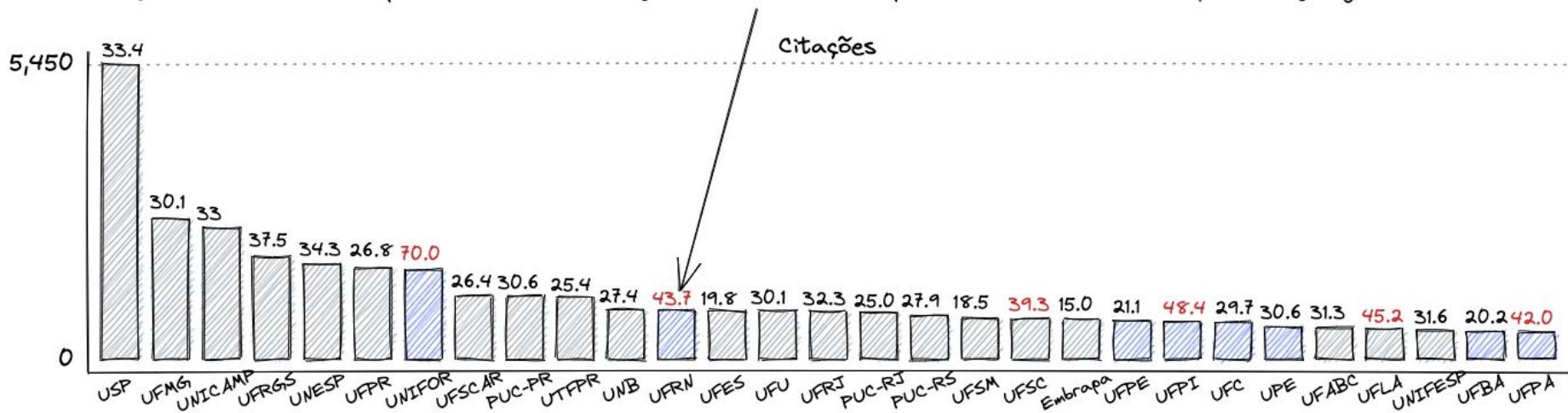
+265k
publicações

Internet of Thing Long Short-term Memory Transfer of Learning
Support Vector Machine Artificial Neural Network
Collaborative Filtering Classifier Recommendation System
Recommende System Deep Learning
Learning System Semantic Facial Recognition Image Recognition
Malware Embedding Object Detection
Embedding Segmentation Generative
Segmentation Learning Method Big Data
Learning Method Recurrent Neural Network Computer Vision Decision Tree
Machine Learning Feature Selection
Neural Network Feature Extraction
Reinforcement Learning CNN Autonomous Vehicle
Data Mining Learning Algorithm Extreme Learning Machine
Learning Algorithm Multi-agent System Natural Language Processing System Convolution
Extreme Learning Machine Learning Model
Natural Language Processing System Image Classification

Publicações

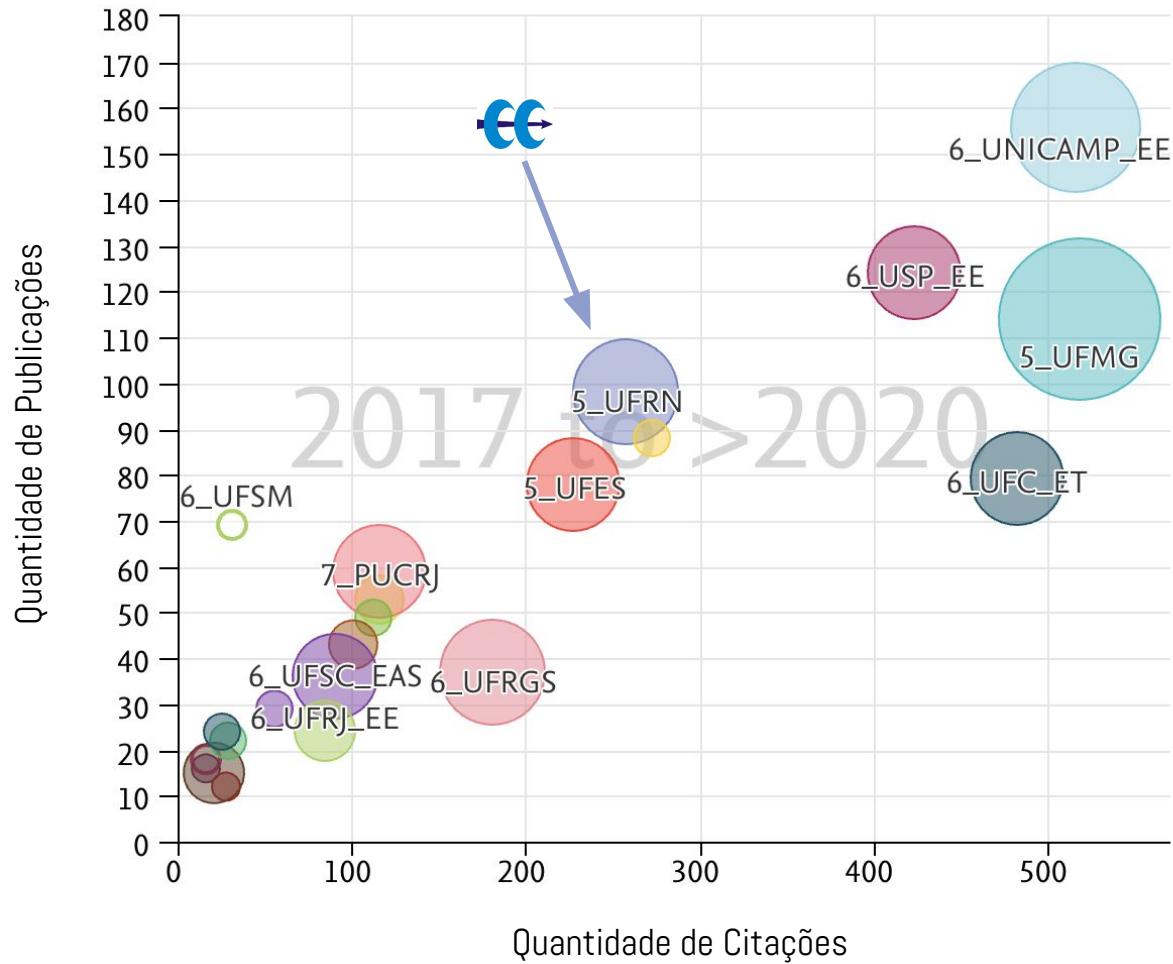


Citações



+1500

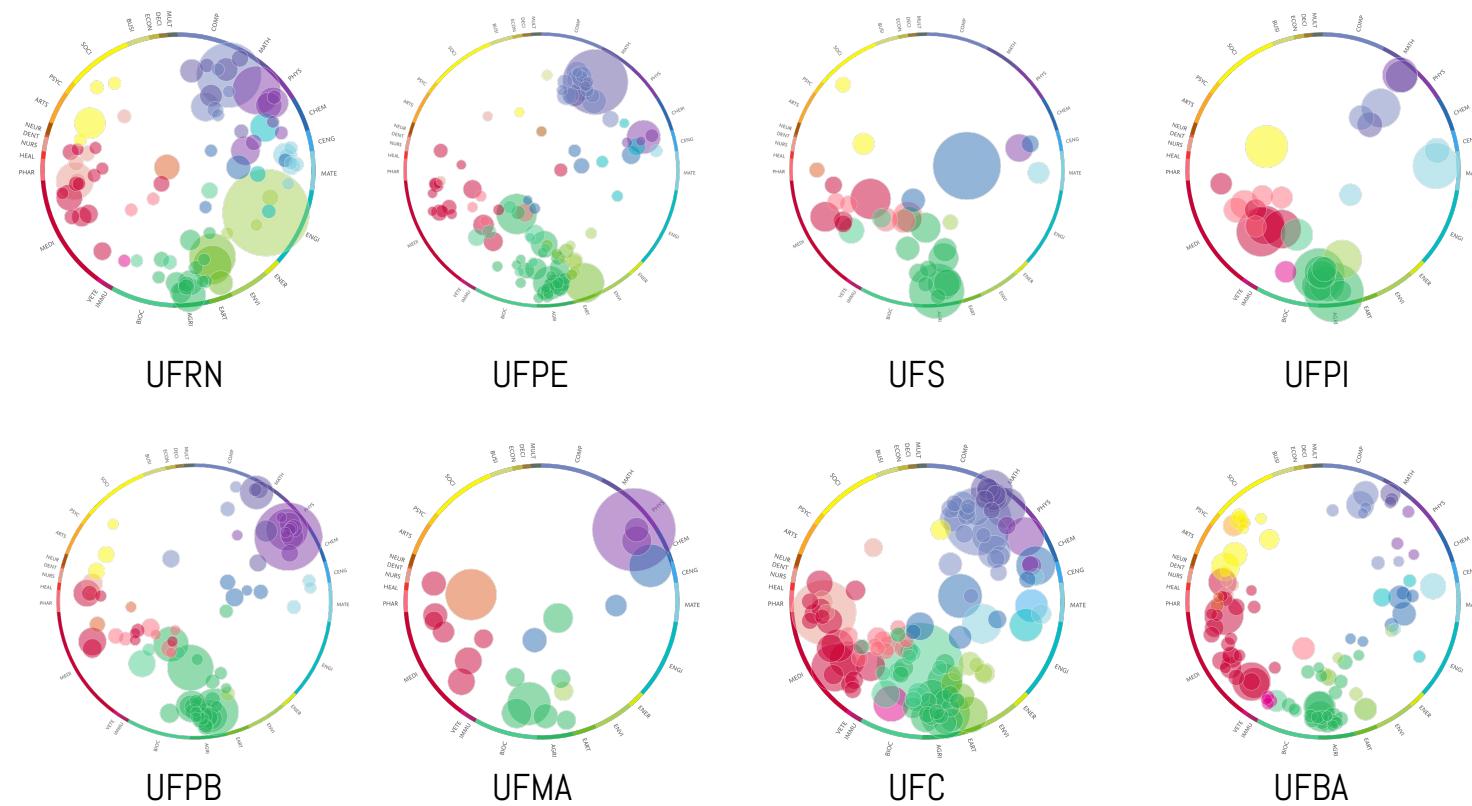




Computer Science (13)

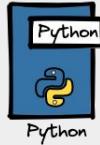
- Artificial Intelligence
- Computational Theory and Mathematics
- Computer Graphics and Computer-Aided Design
- Computer Networks and Communications
- Computer Science (miscellaneous)
- Computer Science Applications
- Computer Vision and Pattern Recognition
- General Computer Science
- Hardware and Architecture

Tópicos e Áreas de Atuação entre Instituições



Course Outline





Optional



Linear Algebra & Math

Probability & Statistics

Data Science

Machine Learning

Deep Learning



Supervised Learning

Decision Tree
Random Forest
Ensemble
XGBoost

Fundamentals of Deep Learning
Better Generalization vs Better Learning
Hyperparameter tuning
Batch normalization
Convolutional Neural Networks
Transfer Learning

Teach your students cutting-edge data skills.

The Dataquest academic program gives qualified institutions access to our data courses.

- ✓ Free access for the semester
- ✓ Hands-on coding and practice
- ✓ Real-world projects
- ✓ Paths to fill skill gaps

[Apply Now](#)

THE UNIVERSITY OF
CHICAGO

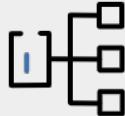
UCI University of
California, Irvine

Northeastern
University

THE UNIVERSITY OF
SYDNEY

Northwestern
University

UFRN
UNIVERSIDADE FEDERAL DO RIO GRANDE DO NORTE



Flipped
Classroom



Active
Learning



Complementary
Material



Career Paths

Data Scientist With Python
Data Analyst With R
Data Analyst in Python
Data Engineering
Business Analyst



Dataquest
Method

Learn with real-world data
Complete exercises and get feedback
Build your portfolio with projects



Search or jump to...

Pull requests Issues Marketplace Explore

Bell + ⚙️

ivánovitchm / ppgeecmachinelearning

Public

Pin

Unwatch

5

Fork

4

Star

18

Code

Issues

Pull requests

Actions

Projects

Wiki

Security

Insights

Settings

main

1 branch

0 tags

Go to file

Add file

Code



ivanovitchm update readme

2b4c5c5 9 minutes ago 33 commits

images

Create readme.md

1 hour ago

.gitignore

configure .gitignore

1 hour ago

README.md

update readme

9 minutes ago

README.md



About

Repository for EEC1509, a graduate course on PPgEEC about Machine Learning

Readme

18 stars

5 watching

4 forks

Releases

No releases published

[Create a new release](#)

Packages

No packages published

[Publish your first package](#)