ALEXANDRE HENRIQUE S. DIAS

I'm a Full-Time Data Scientist and a MSc student in Electrical and Computer Engineering at UFRN. My research is focused on social network analysis, graph theory, and Natural Language Processing. Additionaly, the main programming languages I use are Python, R, C++, and SQL. Besides, my favorite ML FrameWorks are Scikit-Learn and TensorFlow. Lastly, I also have skills in MLOps using GKE, Kubeflow, Kubernetes, and Docker.

INDUSTRY EXPERIENCE

Present | 2021

Data Scientist

Americanas S.A. — São Paulo, SP

- Responsible for building ML models using: Python, Scikit-Learn, and Tensorflow. Apply ML to a wide range of topics, such as Complex Network Analysis, Social Networks, NLP, and HR Analytics.
- Create ML pipelines using KubeFlow Pipelines from Google Cloud AI Platform, and participate in the design of CI/CD operations of ML models.

2021 | 2019

Data Scientist

Looqbox — São Paulo, SP

- Development of BI reports and dashboards using R, Python, and SQL.
- Maintainer of the Looqbox R Package used to build R objects and data structures compatible with the Looqbox Application.

EDUCATION

Present | 2021

M. Sc., Electrical and Computer Engineering

UFRN - Federal University of Rio Grande do Norte — Natal, RN

- Research in complex network analysis, social networks, graph theory, and NLP.
- Tools: python, networkX, gephi, TensorFlow, WandB, Git.

2022 | 2020

MITx Micromaster Program in Statistics and Data Science

MITx on EdX — EdX

- The MITx MicroMaster Program in Statistics and Data Science covers the fundamentals of data science, statistics, and machine learning.

2019 | 2018

B. Sc., Computer Engineering

UFRN - Federal University of Rio Grande do Norte — Natal, RN

- Researcher and member of the Modeling and Scientific Data Analysis team.

Download a PDF of this CV

CONTACT

alexandre.hsd@gmail.com

linkedin.me

github.com/alexandrehsd

medium.com/@alexandre.hsd

alehsdias

TECH SKILLS

- Tensorflow 2.0
- Python 3 (pandas, numpy, scipy, matplotlib, Scikit-Learn, NetworkX)
- R (tidyverse, base R, package development)
- Kubernetes and Docker
- SQL
- git

*Made w/ R **pagedown** package.*

The source code is available at github.com/alexandrehsd/cv.

Last updated on 2022-05-13.

2017 | 2015

B. Sc., Sciences & Technology

UFRN - Federal University of Rio Grande do Norte — Natal, RN

- Linear Algebra and Analytical Geometry Teacher Assistant.
- Calculus II Teacher Assistant.

CERTIFICATES & COURSES

2022 | 2020

MicroMasters in Statistics and Data Science

MITx on EdX

- 6.431x: Probability - The Science of Uncertainty and Data.
- 18.6501x: Fundamentals of Statistics.
- 6.86x: Machine Learning with Python - From Linear Models to Deep Learning.
- 14.310x/Fx: Data Analysis in Social Science.
- DS.CFx: Capstone Exam for Statistics and Data Science.

2021

MLOps (Machine Learning Operations) Fundamentals

Coursera

2019 | 2018

DataCamp completed tracks

DataCamp

- Data Scientist with Python.
- Data Analyst with Python.
- Data Manipulation with Python.
- Machine Learning with Python.
- Importing & Cleaning Data with Python.
- Python Programming.
- Python Programmer.

RESEARCH EXPERIENCE

2019 | 2018

Undergraduate Researcher

Digital Metropolis Institute — UFRN

- Developed a traffic monitoring system using image recognition techniques.

2017 | 2016

Undergraduate Researcher

Department of Informatics and Applied Mathematics — UFRN

- Developed an interactive theorem prover based on Linear Logic using the Maude programming language.

ACADEMIC PUBLICATIONS

Paper published in the 2019 II Workshop on Metrology for Industry 4.0 and IoT (MetroInd4.0&IoT). Naples, Italy.

2019

Performance Evaluation of an Edge OBD-II Device for Industry 4.0

Institute of Electrical and Electronics Engineers — IEEE

- Performance evaluation of an Edge OBD-II device that collects data from vehicles in an autonomous way in order to provide customer feedback and tracking

SELECTED DATA SCIENCE WRITING

2021

Dimensionality Reduction with Factor Analysis on Student Performance Data

Geek Culture

- A dimensionality reduction technique with interpretable outputs.

2021

Stop Using the Elbow Method

Geek Culture

- Silhouette Analysis: A more precise approach to finding the optimal number of clusters using K-Means.

2021

Scikit-Learn 1.0 - A true milestone

Medium

- An overview of the design principles of Scikit-Learn and how the famous ML library became so popular.

2021

The Expectation-Maximization (EM) Algorithm

B2W Engineering

- Understanding the motivations and how the EM Algorithm works.

2020

A mathematical derivation of the Law of Total Variance

The Startup

- Understanding what is and when to apply the Law of Total Variance.

2019

Clustering with K-means: simple yet powerful

Medium

- Explain what is Cluster Analysis, and how the K-means algorithm work providing its pros and cons.

2019

An introduction to Linear Regression

Medium

- Explain all assumptions behind Linear Regression, how to measure its performance, and how to implement it in Python.

I enjoy reading about productivity, lifestyle, data science/AI, and statistics.