Luis Henrique Zanandréa Paese

1623, Duque de Caxias St., Suite 618
Porto Alegre, Rio Grande do Sul, Brazil, 90010-283
+55 54 99152 0111
lhzpaese@gmail.com
https://github.com/paeselhz/
https://linkedin.com/luishzp

EDUCATION

Bachelor in Computer Engineering - Unfinished February 2014 - December 2015
Federal University of Rio Grande do Sul, UFRGS, Porto Alegre, Brazil
Bachelor in Economics February 2016 - Current
Federal University of Rio Grande do Sul, UFRGS, Porto Alegre, Brazil

HABILITIES

Programming Languages: C, C++, Python.
Mastership in Statistical Softwares: R and Stata.
Operational Systems: Linux, Windows, OSX.
Additional Softwares: Microsoft Office, Adobe Photoshop CC.
Data analysis environment:
- *Big Data:* Hadoop(Hive), Spark, PrestoDB, AWS Athena and GCP BigQuery environments;
- *Relational Data Analysis:* Microsoft SQL Server, PostgreSQL;
- *Data Visualization platforms:* Power BI, Tableau, Shiny(R);

Development environments:
- *Conteinerization of data science applications:* Docker, ECS, EKS;
- *Deployment of microservices that collect and share data through Kafka topics*;
- *Configuration and deployment of EMR Cluster with Spark, R and Python stacks available and integration with a data lake to be used by other data analysts and scientists*;
- *Development of backend microservices using Python and the FastAPI framework*;

LANGUAGES

Portuguese: Native
English: Advanced with proficiency
College Yes Cultural (Caxias do Sul, RS, Brazil)
- Two-month English as a Second Language Course at Toronto, ON, Canada Western Town College (Toronto, ON, Canada)
- TOEFL ITP 2016
  - Total Score: 633

EVENTS

First IEEE Summer School on Smart Cities
- Natal, RN, Brazil
- 60 hours
  - Data Visualization Prize

The Developer's Conference - Porto Alegre, RS, Brazil 2018
- Big Data - *8 hours*
- Machine Learning - *8 hours*

The Developer's Conference - Porto Alegre, RS, Brazil 2019
- DevOps Tools - *8 hours*

EXPERIENCE

Data Scientist I   May 2018 - June 2019
Agibank

- Deployment of data visualization reports using Power Bi and Tableau to support business decisions;
- Using on-premises APIs to extract and analyze data;
- Experience integrating the R Statistical Programming Language with Data Visualization tools and Microsoft SQL Server;
- Connecting R with big data environments such as the Hadoop/Spark stack;

Senior Data Scientist   June 2019 - September 2019
Portocred

- Development of credit models;
- Structuring the Data Science pipeline, since data ETL, analysis, modelling and deployment using APIs;
- Deployment of behaviour model using AWS Lambda and AWS SageMaker;

Mid-Senior Data Scientist   October 2019 - Current
Confederação das Cooperativas do SICREDI

- Automation of Machine Learning model routines in batch, so that they run in specific predetermined times, as well as the conteinerization of those processes to ensure stability and reproducibility of them;
- Development of standards and routines to ensure that data scientists and data analysts work following some patterns, ensuring and enforcing that this standards helps reproducibility of code and the capacity of others to understand what each person did;
- Development of shell scripts along with Terraform code to provision EMR clusters in AWS environments ensuring the installation of the tools that are required by data scientists and data analysts to explore data and develop machine learning models;
- Development with the DevOps teams of a *self-service* solution for data scientists and data analysts to provision their cloud environments according to their needs (RStudio Server Open Source or JupyterHUB), following the standards of Infrastructure as a Code and using the Hashicorp stack (Packer, Terraform, Vault and Consul);
- Automation of the deployment of Machine Learning models through Git repositories.
- Development of a recommendation engine to ensure that users receive the offer that best fits them, my main role in this project was the development of backend microservices that serve multiple ML models through APIs;

ADDITIONAL INFORMATION

Data Analyst with Python   *2019 - Datacamp*

- 47 hours

Text published in Towards Data Science   *2020 - Medium*

- Using RStudio Project Templates to help the project standardization in data science teams

Text published in Towards Data Science   *2020 - Medium*

- Using GitHub Actions to speed up CI/CD in data science projects