LUIS ANTONIO RODRIGUES

Principal Data Scientist

Languages: Portuguese and English
Contact: luisrodriguesphd@gmail.com
LinkedIn: www.linkedin.com/in/luis-rodrigues-phd
GitHub: www.github.com/luisrodriguesphd

ABOUT

Luis Antonio Rodrigues is an accomplished Data Scientist and Machine Learning Engineer with over eight years of experience developing innovative machine learning products and services. He holds a BSc in Mathematics, an MSc, and a PhD in Mechanical Engineering from the University of Campinas, one of the most renowned universities in Latin America. Luis's expertise spans across various domains, including Natural Language Processing (NLP), Recommender Systems, Time-Series Forecasting, and Customer Behavior Analysis, with significant contributions across Banking, Consumer Packaged Goods, Retail, Pharmaceutical and Telecommunications industries.

Currently serving as a Principal Data Scientist at DEUS, an AI firm dedicated to human-centered solutions, Luis plays a crucial role in the development of a cutting-edge Retrieval-Augmented Generation (RAG) solution. His responsibilities include improving the knowledge-to-text module, optimizing information retrieval for efficiency and precision, and enhancing text generation for real-time accuracy, showcasing his skills in RAG, IR, LLM, NLP, and several tools and platforms. Additionally, he has contributed as a Data Architect in designing a medallion architecture for a Databricks lakehouse on AWS.

Previously, Luis held the position of Principal Data Consultant at Aubay Portugal, where he led an NLP project for Banco de Portugal, focusing on AI services such as Summarization, Information Extraction, Complaint Text Classification, and Financial Sentiment Analysis. At CI&T, as Lead Data Scientist, he was instrumental in developing a Recommender System for Nestlé, resulting in a 6% sales increase. During his time at Propz, he developed a Recommender System for Carrefour, which boosted revenue by 3%.

His earlier roles include researcher at I.Systems, focusing on water distribution systems, and at the University of Campinas, where his work centered on system and control theory. Luis's proficiency is further demonstrated by his certifications in MLOps with Azure Machine Learning, TensorFlow 2.0, and Python for Time Series Data Analysis. Luis combines his deep technical knowledge with strong communication skills to lead teams and projects towards achieving significant business impacts.

EDUCATION

University of Campinas

2012 – 2016 · **Doctor of Philosophy**, Mechanical Engineering
Area: Modeling, control, and optimization of dynamic systems

2009 – 2011 · **Master of Science**, Mechanical Engineering
Area: Control and optimization of dynamic systems

2005 – 2008 · **Bachelor of Science**, Mathematics

EMPLOYMENT OVERVIEW

Period	Position	Company
12-2023 – Present	Principal Data Scientist	DEUS
08-2022 – 11-2023	Principal Data Consultant	Aubay Portugal
01-2021 – 08-2022	Lead Data Scientist	CI&T
04-2020 – 12-2020	Senior Data Scientist	CI&T
03-2019 – 04-2020	Lead Data Scientist	Propz
03-2018 – 02-2019	Data Scientist	Propz
06-2016 – 03-2018	Researcher	I.Systems

TECH STACK

ProgrammingLanguages	
Matlab	7 yrs
LaTeX	7 yrs
SQL	6 yrs
Python	5 yrs
Scala	2 yrs
C	1 yr
Frameworks & Libraries	
NumPy	5 yrs
Pandas	5 yrs
Scikit-learn	5 yrs
Spark MLlib (ALS, FP-growth)	4 yrs
Statsmodels	4 yrs
PyTest	3 yrs
XGBoost	3 yrs
SciPy	3 yrs
NLTK	3 yrs
spaCy	3 yrs
FastAPI	3 yrs
TensorFlow	3 yrs
PyTorch	2 yrs
Hugging Face Transformers	2 yrs
Sentence Transformers	2 yrs
PyMuPDF	1 yr
Gensim	1 yr
BERTopic	1 yr
keyBERT	1 yr
LangChain	1 yr
Ragas	1 yr
Tools	
Git	1 dec
Jupyter Notebook	7 yrs
Visual Studio Code	4 yrs
MLflow	4 yrs
Apache Spark	4 yrs
Docker	3 yrs
Databricks	2 yrs
Kedro	2 yrs
GitHub Actions	2 yrs
Jira/TFS/Azure DevOps	2 yrs
DVC	1 yr
Faiss/Chroma/Qdrant	1 yr
Databases	
SQL Server	3 yrs
PostgreSQL	2 yrs
MySQL	1 yr
Apache Cassandra (for NoSQL)	1 yr
Cloud Technologies	
Microsoft Azure	4 yrs
Amazon Web Services - AWS	2 yrs
Google Cloud	1 yr
Operating Systems	
Windows	4 yrs
Linux	3 yrs
Mac	1 yr
Development Methods	
Agile (SAFe, Scrum, Kanban)	5 yrs
Waterfall	3 yrs

MAIN PROFESSIONAL EXPERIENCES

DEUS: human(ity)-centered AI
Dec 2023 – Present · 5 months

Principal Data Scientist
Dec 2023 – Present · 5 months

Principal Data Scientist at DEUS, an AI company focused on human-centered solutions. Integrating the development team of a state-of-the-art Retrieval-Augmented Generation (RAG) solution, the work mainly involves the following activities:
- Improve the knowledge-to-text module to provide contextually relevant chunks;
- Optimize the information retrieval (IR) module for precision, diversity and efficiency;
- Enhance the text generation module for more accurate responses in real-time;
- Create evaluation protocol and define metrics for modules and end-to-end solution.

Contributing as a Data Architect in a project to design and implement a cutting-edge medallion architecture for a Databricks lakehouse, my key contributions include:
- Craft the data model that underpins each layer of the proposed lakehouse structure;
- Design raw and business data vaults for the silver layer to facilitate a seamless flow and transformation of data;
- Designing a star schema for the gold layer to perform complex queries with high performance for reporting.

Skills: RAG, IR, LLM, NLP, Real-Time Predictions, MLflow, Docker, Github Actions, Azure, Databricks Lakehouse, Medallion Architecture, Data Modeling, Data Vault 2.0, Star Schema, VS Code, Python, Scrum.

Aubay Portugal
Aug 2022 – Nov 2023 · 1 year and 3 months

Principal Data Consultant
Aug 2022 – Nov 2023 · 1 year and 3 months
Principal Data Consultant at an information technology consultancy present in 7 countries in Europe. Responsible for the technical leadership of an Natural Language Processing (NLP) project for Banco de Portugal, the central bank of the Portuguese Republic, the work mainly involves the following activities:
- Use heuristics, classic Machine Learning, Deep Learning and Large Language Models to solve NLP problems;
- Architect NLP services taking into account the integration between them and the customer's NLP platform;
- Design NLP services for Text Classification, Text Summarization, Topic Modeling and Information Extraction;
- Optimize performance of services for Financial Sentiment Analysis and Text Extraction and Cleaning;
- Application of MLOps principles to build automated (CI/CD/CT) and monitored Machine Learning services;
- Use text vectorization (TF-IDF with Scikit-Learn) and text embedding (transformer-based encoders with Hanging Face Transformers, Sentence Transformers - SBERT, and PyTorch) for feature extraction;
- Use XGBoost and Neural Networks (Tensorflow) in multilabel / multiclass-multioutput classification problems;
- Create the Work Breakdown Structure (WBS) Gantt chart and the product backlog of machine learning projects;
- Conduct SCRUM ceremonies, remove technical blocks from the team, review codes and approve deliveries.

Skills: NLP, Real-Time Predictions, FastAPI, Docker, Github Actions, MLflow, SQL Server, Amazon S3, Kedro, Pytest, Visual Studio Code, Python, SQL, TFS, Agile Methodologies (Scrum).

CI&T
Apr 2020 – Aug 2022 · 2 years and 5 months

Lead Data Scientist
Jan 2021 – Ago 2022 · 1 year and 8 months
Lead Data Scientist in a digital transformation company that provides strategy, research, design and engineering services. Responsible for the technical leadership of the Data Science area in a big data and AI project for Nestlé, the largest CPG company in the world, the work mainly involves the following activities:
- Development of a product recommender system integrated with the sales process;
- Maintenance of purchase recurrence of customers' product assortments through time series analysis (SARIMAX from Statsmodels) for quantity forecast and statical modeling (probability distributions from SciPy) for offer time forecast;
- Expansion of the customers' product portfolio via cross-selling using collaborative filtering (ALS from MLlib);
- Optimization of market basket, aiming to leverage incremental revenue and operational efficiency;
- Development of methodologies for measuring results via randomized controlled trial and A/B testing;
- Development of a methodology of target audience selection for tests considering business rules;
- Enrichment of geolocated data using supervised (K-NN) and semi-supervised (Label Propagation) learning;
- Application of Machine Learning Operations (MLOps) principles and tools for productization and scale.

The recommender system for Nestlé helped small retailers, sellers, and Nestlé, increasing their revenue by 6%.
Skills: Recommender Systems, Batch Predictions, Azure Data Factory, MLflow, PySpark, Delta Lake, Databricks Notebooks, Python, SQL, A/B testing, Miro, Azure DevOps, Agile Methodologies (Scrum).

Senior Data Scientist
Apr 2020 – Dec 2020 · 9 months
Data Scientist in a digital transformation consulting. Responsible for the Advanced Analytics part of innovation projects for Vivo and Centauro players, the work mainly involves the following activities:
- Translate business context into data-oriented hypotheses to drive impact;
- Conducting exploratory data analysis (EDA) to generate insights and support the experiments;
- Creation of AI-based personas for product prioritization (PCA, KMeans/DBSCAN, DecisionTree from Scikit-Learn);
- Proof of concept for content creation automation (Google Trends API to identify hot topics, summarize news automatically via TextRank from Gensim, use Azure NLP API to generate captions for images);
- Design experiments for hypothesis testing and validation of candidate solutions via A/B testing;
- Help the Design team by summarizing interviews using NLP (text cleaning via NLTK and Regex, Named Entity

Recognition via spaCy, key sentence extraction via TF-IDF, and visualization via word cloud);
- Prediction of Key Performance Indicators (KPI) to define Objectives and Key Results (OKR).

Skills: EDA, Clustering, NLP, AWS, Amazon Athena, Github, Visual Studio Code, Google Colab, Python, SQL, A/B testing, Miro, Trello/Jira.

Propz
Mar 2018 – Apr 2020 · 2 years and 2 months

Lead Data Scientist
Mar 2019 – Apr 2020 · 1 year and 2 months

Lead Data Scientist in an artificial intelligence company for marketing and CRM. The scope of the position has varied over time, having started as a people and process management position, then a Product Owner of visualization products, and becoming a technical leader of the Personalization squad to deliver value based on Scaled Agile Framework – SAFe. The personalization squad is responsible for product modules, such as the Recommendation Engine, Contact Time Engine, and Price Virtualization Engine. As the leader, the work mainly involves the following activities:

- Translate the customer's business needs into technical requirements of the company's products;
- Product specification, creating of the Product backlog, and planning the Data Science roadmap;
- Eliminate technical obstacles, review codes and approve deliveries made by the team;
- Conduct Scrum's ceremonies with the team and report the progress to the manager and director;
- Communication with customers about the squad's products and company value deliveries.

Skills: Communication, Product Management, Asana/Jira, Visual Gantt Team, Prototyping, Spark, Scala, SQL, SAFe.

Data Scientist
Mar 2018 – Feb 2019 · 1 year

Data Scientist in a CRM and marketing company for retail. Responsible for monitoring existing products and developing new ones, the work mainly involves the following activities:

- Developing the recommender systems for Carrefour Brazil from scratch that delivered 3% incremental revenue;
- Training and deployment of descriptive models for product price elasticity of demand (Linear Regression and RANSAC);
- Training and deploying reinforcement learning models (Thompson Sampling) for next best channel identification;
- Training and deploying of machine learning models to predict customer's engagement (Random Forest Classification);
- Automation of target audience selection (segmentation) and performance reporting creation for marketing campaigns;
- Development of KPIs to evaluate marketing campaign results and communication with partners in a non-technical way;
- Conduct exploratory data analysis to generate actionable insights and write reports of the developed solutions.

Skills: Recommender Systems, Batch Predictions, AWS/GCP, Big Data, Data Lake, Hadoop, Spark, Git (Bitbucket), Apache Zeppelin, Scala, SQL, Randomized controlled trial, A/B testing.

I.Systems
Jun 2016 – Mar 2018 · 1 year and 10 months

Researcher in Process Control and Optimization
Jun 2016 – Mar 2018 · 1 years and 10 months

Researcher in an advanced industrial process control company, integrating the team of a research project in Smart Cities for the development of a software for modeling, control and optimization of water distribution networks. As responsible for developing the theoretical solution and rapid prototyping, the work mainly involves the following topics:

- Mathematical modeling by phenomenological equations and parametric identification of discrete-time dynamic models;
- Process control and optimization using multivariable Model Predictive Control - MPC;
- Optimization using interior-point methods (nonlinear) and Particle Swarm Optimization - PSO (metaheuristic);
- Water demand forecasting by means of time series analysis (Autoregressive - AR and SARIMA models).

In a case study with a Brazilian company to validate the proposed approach, the software reduced approximately 4,7% of the electrical costs with hydraulic pumps.

Skills: Research, Water Distribution Networks, Predictive Control, Demand Forecasting, Matlab, Python, EPANET, Latex.

MAIN CERTIFICATIONS

End-to-end machine learning operations (MLOps) with Azure Machine Learning
Microsoft
Aug 2023

Tensorflow 2.0: A Complete Guide On The Brand New Tensorflow
SuperDataScience
Jul 2023

Python for Time Series Data AnalysisPython for Time Series Data Analysis
Udemy
Aug 2019