Joel Pinho Lucas

Machine Learning Engineer

São Paulo, Brazil
+55 19996625347
joelpl@gmail.com
linkedin.com/in/joelplucas
github.com/joelplucas

PROFESSIONAL EXPERIENCE

Indeorum, Remote — *Machine Learning Engineer*

MAR 2020 - PRESENT

Tail, Remote — *Data Engineer*

MAY 2014 - FEB 2020

Tail provides data science services for digital publishing partners. I'm member of the Data team, which has already designed and implemented the following deliverables:

- Pipelines, within a Hadoop cluster and employing Apache Crunch, for processing 4 billion log entries per day and store aggregated data from 400 million anonymous profiles in NoSql databases, mainly MongoDB, Redis and ElasticSearch.
- Extended text processing and supervised learning algorithms, from Apache Mahout (Java), for classifying Web page content in subject segments.
- Implemented an internationalization component for classifying Spanish language Web page content from other Latin American countries too.
- Periodically analyze statistical data from digital audiences and update, using Jenkins and AWS, classification models from every country.
- Built a Lookalike feature able to uncover the most relevant patterns by means of unsupervised learning algorithms. Lookalike runs on EC2 instances and was developed on the top of Spark, using Python (Pandas and sklearn) and Spark MLlib.
- Before committing to production, deliverables must pass functional tests (besides unit tests) run on a local environment using Spark and NoSql databases Docker images.

Mobjoy, Campinas (Brazil) — *Recommendation Engines Developer*

SEPTEMBER 2012 - APRIL 2014

- Designed and developed, in Java, the backend engine of the Gameverse Recommender System. The engine was composed of two elements: content-based algorithms responsible for crawling and preprocessing text content from Apple Store game descriptions and ratings (data was stored in MongoDB and MySQL); collaborative filtering (CF) approach, based on MapReduce on the top of Amazon EMR, identifying users with similar ratings.
- Built and maintained the Amazon EC2 instances for running the engine.
- Implemented the engine integration with Amazon EMR for running the CF algorithm on the top of Hadoop.
- Implemented the API routes (RESTFul) for providing recommendations data.

HP R&D, Porto Alegre (Brazil) — *Software Developer*

APRIL 2011 - SEPTEMBER 2012

LIRMM Research Lab, Montpellier (France) — *Researcher*

FEBRUARY 2009 - JULY 2009

EDUCATION

Universidad de Salamanca, Spain — *PhD in Computer Science*

JANUARY 2007 - DECEMBER 2010

Thesis: Using Classification Based on Association Methods in Recommender Systems

Univ. Federal de Pelotas, Brazil — Undergrad. in Computer Science

MARCH 2001 - SEPTEMBER 2006

SKILLS

- Python
- Java
- MongoDB
- Spark
- MLlib
- Hadoop
- Pandas
- Scikit-Learn
- Redis
- MySQL
- NumPy
- PostgreSQL
- Docker
- Git
- AWS
- Jenkins
- RESTFul

EXTRAS

MOOCs from Coursera and edX, 2012-2015: Startup Engineering,, Developing Innovative Ideas for New Companies, Data Analysis, Foundations of Business Strategy, Scalable Machine Learning.

Banco Santander Scholarship, 2007-2010: scholarship program for 100 students selected from Latin America for PhD studies in Spain.

Certificate in Advanced English from Cambridge University (CAE) – 2007

LANGUAGES

Portuguese, Spanish, English, French

CONFERENCES TALKS

2017 – PAPIs Connect Latam — Building ML applications locally with Spark.

2016 – JavaOne Brazil — Applying Machine Learning and Java for Web Content Classification.

2018 – The Developers Conf. (Brazil) – Discovering Lookalike Audiences at Scale for Digital Publishing with Spark MLlib.