Guilherme Pires

ML+Data Engineer / Wannabe World-changer

Contact

Email
mail@gpir.es

Phone
+351 915049434

Website
https://gpir.es

About

I usually say my life goal is to save the World. It might be too ambitious, so I'd already be happy if I could just make it a better place. I plan to do that via my very particular set of skills. I'm hardworking and resilient. I play for the team. I don't give up. When I'm not coding, hacking a project together, training a machine learning model, reading a book, cooking, running, travelling, playing some instrument, or spending time with my friends and family, I'm either sleeping or trying to.

Profiles

Twitter
colobas

GitHub
colobas

Medium
colobas

Work

Jungle
2017-09 — 2018-07

http://jungle.ai
ML+Data Engineer

Since Sep-2017 I'm a full time ML+Data Engineer. Here I've worked with very diverse multivariate time-series datasets, with data sources spanning from heavy industry, to utilities and renewables. Here I've participated in projects that entailed finding causes of quality problems in factories, segmenting and aggregating sensor data by regimes of operation, among others. At Jungle I've also been working on my MSc thesis, on the topic of (deep) Generative Models for unsupervised learning of multi variate time series data.

Highlights

- Worked with diverse data sources, most commonly with very large multivariate time series datasets
- Developed custom algorithms/solutions to deal with said datasets

- Applied SoA techniques in DL and Gradient Boosting algorithms to our customers datasets
- Worked on unsupervised learning for multivariate time series, using (deep) generative models, as my MSc thesis

Jungle

2017-07 — 2017-09

http://jungle.ai

Data Engineering Intern

I started at Jungle as a Data Engineering intern. During my internship I worked mostly on assembling an infrastructure that allows us to do Machine Learning at scale. Mesos, Marathon, Kubernetes, Dask, Airflow, Docker, Zookeeper, are some of the technologies we're leveraging.

Highlights

- Gained experience with Kubernetes, Mesos and Docker
- Gained experience with Azure
- Assembled a prototype of an auto-scalable cluster for Data Science development environments

Snowplow Analytics

2017-01 — 2017-03

https://snowplowanalytics.com

Data Engineering Intern

I was a remote "wintern" at Snowplow for circa 3 months in 2017. They're a über-cool data analytics company. I helped them extend some of their tech to Google Cloud Platform, and along the way I learned Scala and gained experience with cloud technologies.

Highlights

- Learned Scala
- Developed an example project leveraging GCP's Dataflow. Blog post here: https://snowplowanalytics.com/blog/2017/03/30/google-cloud-dataflow-example-project-released/
- Extended key components of Snowplow, enabling them to run on GCP

Quidgest

2013-08 — 2013-10

https://quidgest.com

Software Engineering Intern

Code review and refactoring on an internal tool / R&D / C, C#, C++, SQL,

Personal projects

These are some of the projects I've worked on throughout my life

Highlights

- Politifact scraper
- My wedding website
- Automated Betfair betting strategy
- Dynamic bayesian network constructor (Java)
- SATPLAN system (Python)

- Skeleton for a distributed e-mail system (Go)
- Skeleton for a distributed hashtable (C)
- Parallelization of a MAXSAT solver (C)

Education

Instituto Superior Técnico

2012-09 — 2017-09

🎓 **Electrical and Computer Engineering**
MSc

Karlsruhe Institut für Technologie

2015-09 — 2016-02

🎓 **Electrical and Computer Engineering**
ERASMUS / Semester abroad

Skills

Data Science

- python
- scikit-learn
- pandas
- dask
- numpy
- pytorch

SW development

- Experience with C, C++, Scala, Java, Go
- Web scraping
- Bash scripting

Cloud technologies

- Experience with Azure, Kubernetes, Mesos, Docker

Web Development

- Basic back-end development in Python (Django, Flask, Bottle)
- Basic front-end development with CSS, HTML5, JS

Machine Learning / AI

- Formal training in Machine Learning, Statistical Learning, Artificial Intelligence and Optimization
- Developing my MSc thesis in probabilistic ML, using ANNs for (Amortised) Variational Inference in multivariate time-series models
- Constant studying and keeping up with the developments in the field

Languages

English
Fluent

Portuguese
Native

Spanish
Fluent

German
Beginner

Interests

Track and Field

Music

Travelling

Philosophy

Economics

Politics

Books