

PEDRO MUÑOZ BOTAS

C/Orense 69, 808, 28020 Madrid, Madrid

pedro.munoz.botas@gmail.com

+34 691 101 949

ABOUT

I help companies and organizations to solve their problems through data analysis. I like to do a careful and very organized work so that my solutions can stand out. I have a broad experience in the Big Data ecosystem (Spark, Hive, Impala, Flume...) and in the Machine Learning one (Tensorflow, OpenCV, Scikit-Learn...).

I won the AWS DeepRacer Spanish League, where I developed an autonomous driven car model using Reinforcement Learning. Thus, I will be in Las Vegas competing in the World League.

I am the creator of The Moderator Guru, an NLP tool that is capable of detecting and classifying toxic comments. I've also developed all web and chat interfaces, along with the analytics stack on HRbots.

I am the maintainer of DataTau, the reference newsboard for Data Scientists and Data Engineers inspired by the popular Hacker News.

I am the Indie Hackers Madrid Community Ambassador. You can follow what I'm currently doing at The Gurus and a detailed CV at my personal website.

EXPERIENCE

Pragsis Bidoop

Senior Data Scientist

April 2018 - Present

· TASKS:

- Development of Machine Learning models using traditional (scikit-learn, XGBoost, lightGBM) and Deep Learning (TensorFlow, Keras) frameworks
- Scaling Machine Learning models from prototyping to production using distributed and parallel computing (Spark, Dask, Celery)
- Orchestration of data pipelines using Apache Airflow
- Python toolkit for data science: pandas, matplotlib, plotly, numpy...
- Live tracking of people in buses with WiFi sniffing techniques: Scapy
- Computer vision analysis: edge TPU, tf-lite, OpenCV

· ACHIEVEMENTS:

- Improvement of 10 points in the electrical power prediction of Iberdrola wind mill farms
- Reduction of Spark computation times in processing chains: from days (or not finishing at all) to few minutes
- Developed a live tracking system for detecting people using Computer Vision techniques optimized for low consumption hardware requirements (Google Edge TPU, Raspberry Pi)

DataTau

Maintainer

June 2019 - Present

· TASKS:

- Web development with Django (Python)
- Management and maintenance of Open Source project (issues, roadmap, pull requests, etc.)
- Promoting the site in tech forums and communities
- Hosting management in personal datacenter

· ACHIEVEMENTS:

- Development of a fully functional clone of Hacker News in a record time (3 days)
- Viral relaunching of the site, reaching Hacker News front page and a very relevant position in Reddit
- Hosting in self-owned infrastructure, supporting major traffic spikes
- Revitalizing DataTau community and creation of open source project that ensures continuity and independence of DataTau forever

HRbots

January 2018 - Present

Tech Director

· TASKS:

- Development of chatbots (engine and chat interfaces) applied to Human Resources tasks
- Use of Google AI engine for improving the chat experience: Dialogflow
- Development of a custom training system based on NLP: spaCy
- Extraction of insights from employee-chatbot conversations through data analysis to generate value in business processes: pandas

· ACHIEVEMENTS:

- Successful integration of a full featured PWA (push notification system, app-like interfaces...) in Top-Employers Institute
- Creation of engagement in bot conversations through contests and analytics

The Moderator Guru

January 2018 - Present

Founder

· TASKS:

- Development of a NLP engine capable of detecting and classifying toxic comments
- Development of the Frontend stack (HTML, CSS, Javascript) using Bootstrap and jQuery along with a Django backend

· ACHIEVEMENTS:

- An NLP project working on production with an available API, able to server thousands of requests per minute
- Full project is bootstrapped and self-hosted in on- premise home infrastructure

P3 Group

June 2017 - April 2018

Big Data Developer - Data Scientist

· TASKS:

- Development of software with the Cloudera Environment: Spark (Java API), Hadoop, Impala, Hive...
- Use of AWS infrastructure (EC2, EMR, S3...)
- Creation and automatization of Cloudera clusters over EC2 instances
- Data analytics using simple correlations and data processing: Spark MLlib, pandas, scikit-learn

· ACHIEVEMENTS:

- Fully automatization of Cloudera clusters in AWS (launching, installation, processing and shut down)

GMV

May 2016 - June 2017

Software Developer

· TASKS:

- Development of flight dynamics software in Java and Python languages
- SST operator specialised in collision, re-entry and fragmentation risk analysis

Etulos Solute

July 2015 - May 2016

Software Developer

· TASKS:

- Development of a full-featured fatigue computation tool from scratch in Python and Fortran languages
- Management and cleanup of high amount of fatigue data with Python data science toolkit: numpy, pandas
- Teacher of a 10h course about fatigue analysis: It comprised theoretical fatigue concepts along with a description of Solydro, a software for fatigue analysis which I have developed at Solute

· ACHIEVEMENTS:

- Huge savings for Solute in terms of SW licenses (30K), as my code replaced the comercial SW used for fatigue analysis at the company

EDUCATION

Formación Hadoop

March 2018 - April 2018

Master in Big Data

Overall GPA: N/A

Universidad Politécnica de Madrid

September 2014 - February 2016

Master in Applied Mathematics

Overall GPA: 9.3/10

Universidad Politécnica de Madrid

September 2010 - June 2014

Degree in Aerospace Engineering

Overall GPA: 9.0/10

HONORS & AWARDS

DeepRacer Winner

Amazon Web Services (AWS)

Machine Learning competition organized by AWS which consisted in developing a Reinforcement Learning model for an autonomous driven car. I won the Spanish competition so I will be competing in Las Vegas in the World championship

Excellent Performance

Universidad Politécnica de Madrid

Prize awarded to the highest marks (5 % of students) in Aerospace Engineering

CERFITICATIONS

Spark and Hadoop Developer CCA-175

Cloudera

Credential ID: 100-020-312

Certificate in Advanced English (C1)

Cambridge University

Grade C

TECHNICAL STACK

Machine Learning	scikit-learn, lightGBM, MLlib
Deep Learning	TensorFlow, Keras, PyTorch
Computer Vision	OpenCV
NLP	NLTK, spaCy
In Memory	pandas, NumPy, Apache Arrow
Relational	PostgreSQL, SQLite, MySQL
Big Data	Apache Spark, Hive, Impala, Flume, Kafka, Cloudera, MapR
Cloud Computing	AWS (Amazon Web Services)
Orchestration	Apache Airflow, Apache Oozie
Web Development	HTML, CSS, Javascript, Django
Languages	Python, Scala, Java