PEDRO MUÑOZ BOTAS

C/Orense 69, 808, 28020 Madrid, Madrid

p.munozb2103@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 am the AWS DeepRacer Spanish League winner, 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 hr-bots.com.

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

EXPERIENCE

Pragsis Bidoop   April 2018 - Present
Senior Data Scientist

· TASKS:

- Design and development of different architectures in Spark on Scala and Python
- Development and productivization of electrical power prediction models in wind mill farms: LightGBM, Airflow
- Implementation of ML algorithms: Tensorflow, scikit-learn, Spark MLIB...
- Python toolkit for data science: pandas, matplotlib, plotly, numpy...
- WiFi tracking analysis to create people movement maps: scapy
- Computer vision analysis: edge TPU, tflite, 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
- Live tracking of people in buses with WiFi sniffing techniques
- Live tracking of people on street with computer vision techniques and adaptation to hardware requirements (edge TPU)

HRbots   January 2018 - Present
Tech Director

· TASKS:

- Development of chatbots applied to Human Resources tasks

- Design and development of all chat interfaces (web based and PWAs)

- 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

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...

- Development of web applications with Django framework

- 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 MLIB, 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 substituted 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	Tensorflow, Keras, OpenCV, NLP, Scikit-Learn, Pandas, Numpy
Big Data	Spark, Hive, Impala, Flume, Kafka, Cloudera, MapR
Cloud Computing	AWS (Amazon Web Services)
Orchestration	Airflow, Apache Oozie
Web Development	HTML, CSS, Javascript, PHP, Django
Computer Languages	Python, Scala, Java

LANGUAGES

Spanish   Native proficiency

English   Full professional proficiency (C1)