DAVID A. CAÑONES

96 Jorge Juan, 28009 Madrid, Madrid, Spain

davidarcano@gmail.com

+34 637 539 220

ABOUT

Data Scientist & Machine Learning Engineer with 4+ years of experience helping companies and institutions solving complex problems using data analytics solutions. I have successfully completed projects ranging from predictive modeling to data pipelines designing for both enterprises and startups.

I have a broad experience using Python Data Science toolkit (pandas, scikit-learn, Tensorflow, Keras, PySpark, etc.)and work with large amounts of data ($\sim$GB, TB) in a daily basis, so I am also experienced with Big Data and distributed computing tools (Hadoop Ecosystem: Spark, Hive, Impala, etc.) as well as parallel computing ones (GPU accelerated computing).

I am creator of products like The Moderator Guru, a Natural Language Processing tool that automatically spots and classifies offensive text content.

I am member and supporter of the startup community and Indie Hackers Community Ambassador for Madrid.

If you are interested in my job, please visit my personal web and my blog.

EXPERIENCE

Pragsis Bidoop   July 2018 - Present
Senior Data Scientist

· 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)

- Orchestrating data pipelines using Apache Airflow

- Leveraging Python toolkit for Data Science to extract valuable information from Data and explain it through visualizations and reports using among other tools pandas, NumPy, Matplotlib, Plotly, Bokeh, Seaborn, etc.

- Development of Computer Vision solutions using Google Edge TPUs, Nvidia GPUs, Tensorflow and OpenCV

· ACHIEVEMENTS:

- Improved power production forecasting error in 10 percentage points for a cluster of 13 wind farms (about 1GW total managed power) located in Washington, USA, resulting in important savings for our client

- Developed a Reinforcement Learning algorithm for Amazon Web Services DeepRacer League and became member of the winning team (Gold, Silver and Copper positions)

- Developed a live tracking system for detecting people using Computer Vision techniques optimized for low consumption hardware requirements (Google Edge TPU, Raspberry Pi)

Kernel Analytics (now BCG Gamma) October 2017 - July 2018
Data Scientist

· TASKS:

- Development of Machine Learning models using traditional (scikit-learn) and Big Data (Spark MLlib) frameworks

- Designing custom KPIs based on customers needs and data availability

- Designing data pipelines able to ingest data from heterogeneous inputs into Hadoop Distributed File System (HDFS)

- Orchestrating data pipeline executions using Apache Airflow

- Extraction of insights from customers data and creation of meaningful visualizations using Matplotlib, ggplot2, Seaborn, Plotly and Bokeh

- Creating interactive dashboards using Plotly Dash and Microsoft PowerBI

· ACHIEVEMENTS:

- Developed a Customer Experience Management framework for a successful Mobile Operator. Designed pipeline from ingesting 3G/4G antennas data to creating a model to relate Customer Experience with Churn and Complaints, resulting into our client being able to monitor its mobile network infrastructure impact in Customer Experience

- Developed a predictive model for a well known Mobile Operator able to predict users complaints based on consumption patterns and user personal profile, resulting into our client being able to automate part of support process

The Moderator Guru January 2019 - Present
Founder

· TASKS:

- Development of a fast NLP engine capable of detecting and classifying offensive text contents

- Development of the web app around the NLP model, with Bootstrap and jQuery front-end and back-end based on Django

- Management of infrastructure based on Linux servers and networking

· ACHIEVEMENTS:

- Developed a NLP based product working on production with an available REST API able to serve thousands of requests per minute

- Full project is bootstrapped and self-hosted in on- premise home infrastructure, allowing me to become Indie Hackers Community Ambassador for Madrid

Grupo Servinform February 2015 - October 2017
Junior Data Scientist

· TASKS:

- Designing data pipelines able to automatically clean and ingest data from heterogeneous inputs into relational databases using pandas and NumPy

- Developing Natural Language Processing models using NLTK

- Developing a data product (web app) to make data exploration easier for pharmaceutical and healthcare researchers allowing them to establish complex relationships between data from different sources

- Identifying potential automation opportunities internally and for our clients and validating technical feasibility

· ACHIEVEMENTS:

- Developed a Natural Language Processing Model able to interpret user queries (written as natural language) and translate into queries to our database

- Developed a framework able to automate parts of hand-made back-office processes and integrate seamlessly with human workers, resulting in a new business line for our company and the ability to tackle projects that otherwise would be discarded

- Collaborated with my manager in the writing of a proposal for H2020 R&D program with an excellent result and public research funds granted to our company

EDUCATION

EOI Business School *September 2014 - September 2015*
MBA, Corporate Finance

Universidad de Sevilla *September 2007 - September 2014*
MS Industrial Engineering, Energy

HONORS & AWARDS

DeepRacer League, 3rd position *Amazon Web Services (AWS)*
Machine Learning competition organized by AWS which consisted on developing a Reinforcement Learning model for an autonomous car. I got the 3rd position in Spanish competition and was member of the team that made the top 3 positions (Gold, Silver and Copper)

COURSES AND CERTIFICATIONS

- **Machine Learning**
  Stanford University

- **Advanced Machine Learning Specialization, 2/7**
  Higher School of Economics Moscow

- **Deep Learning Specialization, 2/5**
  deeplearning.ai

- **Learning From Data**
  Caltech

- **The Analytics Edge**
  MIT

- **Computational Thinking and Data Science, 2/2**
  MIT

- **Data Manipulation at Scale: Systems and Algorithms**
  U. Washington

- **Data Scientist Path**
  dataquest.io

TECHNICAL STACK

Machine Learning	scikit-learn, XGBoost, lightGBM, H2O, MLlib
Deep Learning	TensorFlow, Keras, PyTorch
Computer Vision	OpenCV
NLP	NLTK, spaCy
In Memory	pandas, NumPy, Apache Arrow
Relational	PostgreSQL, SQLite, Oracle
Big Data	Apache Spark, Hive, Impala
Cloud Computing	AWS (Amazon Web Services)
Orchestration	Apache Airflow, Apache Oozie
Web Development	HTML, CSS, Javascript, Django, Flask
Languages	Python, R, SQL