

# DIEGO PEREZ

## DATA SCIENTIST

✉ hi@diegopersas.com  
🌐 www.diegopersas.com  
☎ +34 686 18 47 65  
📍 Madrid  
🐦 diegopersas  
📘 diego-perez-sastre/  
👤 persas

## Summary

I am a data scientist with a software engineering background, tech-lover and amateur athlete. I began playing around with computer science when I was 13, I studied two associate degree's and a Bachelor's degree in computer science (Software Engineering track). Afterwards, my interest in the Artificial Intelligence world started growing and I studied a Master's Degree in Artificial Intelligence. Meanwhile, I have performed an internship at BBVA in the innovation team and I have been working at Clarity as Data Scientist since 2017.

## Skills

### TECHNICAL

TensorFlow  
Python  
Pandas/Numpy  
Scikit-Learn  
PySpark  
Java  
Research

### PERSONAL

Sport Lover  
Communication  
Team Work  
Fast Learning  
Problem Solving

## Education

Technical University of Madrid Current  
M.S. Artificial Intelligence 2020

Technical University of Madrid  
B.S. Software Engineering 2018

Maria Auxiliadora, Leon  
Associate Degree. Administration of Computer System in Network 2014  
-----

Coursera | DeepLearning.ai  
On-Line course Deep Learning Specialization 2018

## Employment

Clarity Madrid  
Data Scientist | Machine Learning Engineer Oct. 2017 to Current  
At Clarity AI we aim to help the decision-makers to better allocate their capital based not only in the financial performance but also on a societal point of view.

Since Clarity is a fast-growing startup within the Data Science team we perform a very wide range of tasks, from Analytics to Engineering.

Some examples of the tooling that we use:

- Python, Pytorch, TensorFlow, Sci-kit learn, Jupyter, ...
- Airflow, mlflow, domino...
- Redshift, MongoDB, S3...

(check personal web for more info)

BBVA Madrid  
Data Scientist Oct. 2016 to June 2017  
I was part of the innovation lab at BBVA. We developed an Automated ETL framework from end to end—currently in production internally—able to handle very big datasets. Some of the tooling and task we performed during this time:

- Development of a protocol to analyze streaming data based on previous transformations used in a real fraud detection project.
- Tools: PySpark, Pandas, Numpy, Matplotlib, HDFS, Scikit-Learn, Jupyter, Zeppelin

## Projects

Object detection in small data Feb. 2020 to Current  
Using transfer learning from YOLO v3 models trained on big data for object detection in small data

Recommendation System Mar. 2018 to July 2018  
I have developed a movie recommender system content-base—actually based on movie's genders and its synopsis (NLP)—from scratch.

LSTM to generate old Spanish Dec. 2017 to Jan. 2018  
LSTM neural network able to write old Spanish trained with the classic Spanish book "El ingenioso hidalgo Don Quijote de La Mancha"