



Javier Guzmán

Machine Learning Engineer

Education

- 2019 – Present **PhD in Artificial Intelligence.**
University of A Coruña, Spain
- 2017 – 2019 **MSc in Artificial Intelligence Research.**
Specialty in Data Science and Machine Learning
UIMP – AEPIA, Spain
- 2012 – 2016 **BSc in Computer Science Engineering.**
University of Santiago de Compostela, Spain

Experience

- Feb. 2021 – Present **Sr ML Engineer, DefinedCrowd, Porto, Portugal - Remote.**

Additionally to the maintenance and improvement of previously developed services and tools, focused on building a solution that oversees the **quality of speech data delivered to clients** and eases the delivery process.

Among other details, this system includes the development of:

- a micro-service and a REST API (FastAPI),
- a binary **classification model for audio** (deployed in Kubernetes using MLFlow),
- an **on-demand metrics computation** using Druid.

As well, cooperated with stakeholders to understand the impact of our solutions on their needs and guide them on its usage. And, helped new team members in onboarding processes and first steps inside the team.

- Feb. 2020 – Jan. 2021 **ML Engineer, DefinedCrowd, Porto, Portugal - Remote.**

In the context of a data crowdsourcing company, worked in the squad responsible for ensuring the quality of the data submitted by the crowd.

To achieve this goal, worked towards the **productization of metrics and models**

which allowed taking **dynamic actions** as part of the data flow and **building analytical tools** for monitoring purposes.

Other tasks performed:

- Improve CI/CD and packaging of artifacts from data scientists and internal tools.
- Development and maintenance of end-to-end tests.
- Helping data scientists in their day-to-day routine.
- Maintenance of micro-services and throughput enhancement.
- Monitoring and alerting of infrastructure.

Used a tech stack with Python as the main programming language. Additionally, worked with .NET/C# to maintain some parts of the system. Some of the related tools and frameworks used in the development:

- Flask/FastAPI
- Pandas, scikit-learn, PyTorch, MLFlow
- PySpark over YARN (with ADF as orchestrator)

Our infrastructure includes the next platforms:

- Kafka, RabbitMQ
- PostgreSQL, SQL Service, Apache Druid, Apache Hive, CosmosDB
- Superset
- Docker, Kubernetes
- Kibana, Grafana

Oct. 2016 – **ML Engineer**, *Gradient*, Vigo, Spain.
Jan. 2020

Participated in different projects to solve the emerging industry challenges making tasks as:

- Project definition and team coordination.
- Exploratory data analysis and data visualization using code notebooks.
- Development of proofs of concept and data pipelines.
- Support to design the machine learning architectures.
- Development and exploitation of ML/DL models using both classical tools, big data and streaming platforms.
- Deployment of the developed data pipelines and fitted ML/DL models to production and monitoring of its performance.

Designed, implemented and deployed systems capable to provide **recommendations** or **detect anomalies** through serving the trained models **in real-time** with *Apache Spark Streaming* or through an API REST with *Flask*. Applied techniques to build the models:

- Matrix factorization techniques (SVD, ALS) and AutoEncoders models.
- Wide&Deep models using embedding neural networks.
- Clustering techniques (GMMs, DBSCAN) and statistical techniques.

Achieved the implementation of a system designed to capture, transform and store user actions to provide behaviour metrics, activity **predictions** over **temporal series, frequent patterns** and **clustering** of users.

- Pattern mining and frequent itemsets mining.
- ARIMA, Hidden Markov Models and LSTM Networks.
- K-means and statistical techniques.

Technology stack used:

- Languages: Python, R, Scala, TypeScript.
- Platforms: Apache Spark, Apache Kafka, RabbitMQ, Metabase.
- Data libraries: Pandas, Numpy, Matplotlib, ggplot2.
- ML/DL Libraries: Scikit-Learn, SparkML, Keras, TensorFlow, Caret.
- Databases: MySQL, PostgreSQL, MongoDB, InfluxDB.
- Tools: Jupyter, Zeppelin, Airflow, MLflow, Docker, Kubernetes, CI/CD, HDFS, Git.
- Other: Scrum, Kanban, TDD, Clean Code, Clean Architecture.

2015 – 2017 **Co-founder and FS Developer**, *Gradox*, Santiago de C., Spain.

Co-creation of a collaborative platform for students, oriented to the early-creation of a professional profile. To afford this, designed and implemented a system based on a microservice-based architecture.

Tech stack used:

- Languages: Java, TypeScript, Python.
- Frameworks: Spring Boot, ReactJS, NodeJS.
- Databases: MySQL, PostgreSQL, MongoDB.
- Other: Scrum, Docker, Apache Kafka, Git.

2016 **FullStack Developer Intern**, *2Mares*, Vigo, Spain.

Worked with a management application broadly used on call centers. Performed tasks like:

- Maintenance of the system.
- Development of new features.
- Bug fixing

Tech stack used:

- Languages: C#
- Frameworks: .NET
- Databases: SQLServer

2014–2015 **Android Developer**, *Chattyhive*, Vigo, Spain.

Collaborated with a multidisciplinary development team by implementing UX designs in an Android application.

Side projects

2018–2019 **Feature selection in big image datasets.**

Researched **feature selection methods over Apache Spark** from features extracted from images of the Imagenet dataset. Performed feature extraction steps through CNNs (TensorFlow) and classical **artificial vision** methods like LBP and keypoint detection with BoW (OpenCV).

2016 **MetaDW.**

Created a management and presentation system for a data warehouse metadata related to a **Business Intelligence** environment.

Skills

Main skills	Machine Learning, Deep Learning
Relevant skills	Architecture Design, DevOps, Computer Vision
Big Data	Apache Spark, Apache Kafka, RabbitMQ
Analytical Tools	Druid, Superset
Languages	Python, R, Scala, Typescript, Java, C/C++
Other tech	MLFlow, Docker, Kubernetes
Soft skills	Teamwork, Critical Thinking, Problem solving, Research, Mentoring

Community Activities

October 2019 **DSPTDay 2019.**

Speaker on the DS@Academy activity.

2017–Present **Blockchain.gal.**

Co-organizer of a meetup group related to **blockchain** technologies.