MARCO RUSSO

Machine Learning Engineer & Professor

☎ **+34 603 220 001 // +39 338 1038 912**   @ **mrussorb@gmail.com**

🔗 **www.marcusrb.com · linkedin/in/marcusrb**   📍 **Europe**

EXPERIENCE

Senior Data Scientist

Paradigma Digital by Indra Group · Plexus · Bosonit

📅 06/2018 - Ongoing   📍 Madrid, Spain

As a data specialist I carry out statistically valid data analysis to solve specific business problems. As a data engineer I build statistical and machine learning models and automate them by designing and building secure, reliable and scalable data processing systems.

- Worked in 8+ **different projects and industries** in AWS, GCP, Azure.
- Developed a Data pipeline and ML workflow Python and Spark.
- Managed 15 data specialist agile autonomous teams got the data which they need to forecast sales and demand. ETL created in Azure involved Synapse/Azure DWS, SQL Server, Data Factory and Databricks) in two different projects for bank and financial services and retail company. Programming in R for statistical inference and data mining; Python to developing and create ML applications; Scala into Databricks or Jupyter or Zeppelin notebooks.
- **Data Architecture** for ATM Predictive Maintenance (NodeJS > Kafka on Confluent > Spark Streaming > MongoDB > Delta Lake > Batch and Power BI. Proof of Concept about fraud detection, customer churn, etc.
- Other Machine Learning models time series based.

Machine Learning Engineer & in-company Trainer as freelancer

Dataschool.Tech for banking · insurance · energy&power co.

📅 02/2012 - Ongoing   📍 Europe

As freelance I design, develop, deploy and manage differents statistical models focused:

- Data Mining & EDA - Data Visualization
- Machine Learning modeling, used deep statistic techniques (SVM, Gradient, Clustering, PCA, LDA, Decision Tree, KNN, K-Means, Logistic Regression and Linear Regression, ANOVA, time series, neural network.
- Used machine learning algorithms to analyze adoption phase retention that resulted in a 10x increase in long term retention
- Trained and assessed over 40 individuals in over 70 total sessions and consultations

Associate Professor · Tutor of Data Science program and master's degree

Universitat Oberta Catalunya · The Bridge Talent · other BS

📅 01/2005 - Ongoing   📍 Spain, Europe, Remote

Associate professor and main instructor for the below courses:

✓ Create, coordinate and tutoring of the Introduction to Big Data platform, Statistics, Maths, Algorithms fundamentals and **Machine Learning**, Data Mining, Cloud programming, Data pipeline and MLOps.

✓ **Business Intelligence** program from Data Integration ETL , SSIS, SSAS SSRS and Visualization Power BI or Tableau. Use Cases with real dataset or open data datasets.

- **Cloud computing**: Big Data Essentials, Streaming and Batching data processing and Machine Learning (Hadoop, Spark, Mahout, Hive, HBase, Storm, Zookeeper, Ambari, Flume, Kafka) on-premise and in the cloud (AWS, GCP, Azure and Databricks). Developing ETL pipeline, ML workflow and created in Apache NiFi tasks to collect, transform and put the data in the data lakes (Blob Storage, Cloud Storage, S3, Elastic).

SUMMARY

Involved durting 10+ years experience as a Data specialist for **Machine Learning**, Business Analytics and Data Engineering tasks for intl. and national large and midsize companies (CEPSA, Santander Bank, BBVA, Naturgy, Auchan, Mapfre and so on). Lead **in-company trainer'courses** and as associate professor in BI, Data Analytics and **Data Science specialization** program for some universities and business'schools, focused in AWS, GCP and Azure - Databricks platform, Data Integration/Wrangling and visualization services. **Certifications**: Associate Data Engineer AutoML on Microsoft Azure, GCP and AWS - Spark Analysis, SparkML; Databricks and Confluent Streaming.

SKILLS

Spark & Databricks	AWS · GCP · Azure

ML Engineering & MLOps

Data mining & Data modeling

Lambda & Kappa architecture

SQL & NoSQL databases

Python & R programming	Training

Data analysis & Dataviz	PowerBI

LANGUAGES

English
Proficient   ●●●●○

Italian
Native   ●●●●●

Spanish
Native   ●●●●●

French
Intermediate   ●●○○○

PROJECTS

Maintenance ATM's & Fraud predictive model

Design, Deploy and Analyze complete architecture on-premises and cloud scenario for important banking company.

- Deploy microservices on Docker and K8S cluster to create lambda architecture (Kafka / Flink and Storm), Spark, Storage, Visualization.
- Deploy Kappa architecture Confluent Kafka base with KSQL, Kstream for processing, aggregation and Spark to apply Machine Learning models. MLflow to log and serve versioning.
- Fraud predictive model based Confluent use case.
- Working in Batch and Streaming flow, understanding and train different machine models in both scenarios, analyze and follow main KPIs to Apache Superset, Grafana for understanding performance and compare the metrics from collect, storage and processing phase.

Advanced Analytics - CRM & on-premises to GCP BigQuery / Looker & PowerBI

Migrated CRM data to data lake in GCP and BigQuery:

- Data collect from BigQuery, apply Machine Learning methods for analyze customer behaviours. AutoML and custom ML models.
- Migrated organizational files, excel, data marts to Azure to data analysis and visualization in Power BI - Tableau - Data Studio, SQL Datawarehouse, Synapse, Databricks are some tech stacks used.
- Digital transformation processing, A/B testing, Analytics tools

Convolutional Neural Network Based Emotions Recognition System

A CNN based emotions recognition system, built with Scikit-Learn, Tensorflow and OpenCV

- High degree of accuracy 86% and confidence for real-time emotions recognition. PoC deployed on GCP and Azure.
- Potential applications in - Behavioural analysis, real time analysis of customer satisfaction in retail settings, automatic photography.

PASSIONS

My family · cycling · play guitar piano & e-drum, hiking · travel books/movies · cook
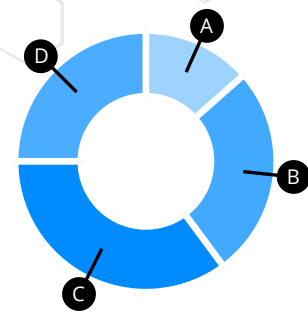
EDUCATION

Data Engineering & Data Science & Business Analytics Master degree · Economics degree spec. financial markets

Universitat Oberta Catalunya

MY TIME

- **A** PoC, Investigation & Papers
- **B** Data mining & Stats
- **C** Machine Learning & Deep Learning cases
- **D** Data architecture & Cloud engineering & MLops - Project management