

Tiago dos Santos ▽ Data Science Engineer

Portugal, Lisbon • santos.tmd@gmail.com

Experience (6+ years)

- 2020/12-Present** **Data Science Engineer; DefinedCrowd**
- Implementation of pipelines for model serving with api for model monitoring and model tracking.
- Tasks:** Model deployment with CI/CD; Comparison of Model Signatures from lab and production; Comparison of features from lab and production;
- 2019/11-2020-12** **Data Infrastructure Engineer; DefinedCrowd**
- Development, deployment and monitoring of infrastructure and services for internal Data clients.
- Tasks:** Event Sourcing Infrastructure implementation; Data Observability processes; Data Democratization processes; Deployment and Monitoring of Kafka, Kafka Connect, JupyterHub, Druid, KSQLDB; Migration from Azure Data Factory to Airflow;
- 2019/03-2019/11** **Data Science Engineer; Stratio**
- Development of models and infrastructure creation for the Stratio AI project.
- Tasks:** Batch and stream infrastructure implementation; Framework creation for project unification and ease to deploy on production; Development and iteration of models with human feedback; Development of web apps for human feedback; Model monitoring;
- 2017/11-2019/03** **Data Science Engineer; Feedzai**
- Development of a model for *Non-Disclosure Client* to detect online payment fraud within 5 countries.
- Tasks:** Model Stability and Model variability Analysis; Developing in-house tools (faster data validation, exploration and model production monitoring); Feature Engineering; Benchmark model in production & retrain; Model Monitoring; Data analysis to identify fraud attacks; Pipeline optimization;
- 2016/06-2017/06** **Data Science Consultant; Altran Portugal**
- Development of a ML solution for short-circuit fault detection in Induction Motors.
- Tasks:** ETL; EDA; Feature Extraction; Model Comparison; Model Hyperparameter Tuning
- 2015/07-2016/03** **Software Engineering & Project Manager; Freelancer - Fluxphera**
- Development of a scalable Web App to access and manage of the online multiplayer Be First Game.
- Tasks:** Node.js fullstack; Scalable and auto deploy (AWS); Distributed state protocol for online game
- 2013/02-2014/09** **Software Developer; Safira**
- Development of a BPM unified platform for insurance claim management with legacy integration.
- Tasks:** Masterdata modeling; SQL Server Query Optimization; Backend and frontend dev (IBM BPM);
- 2011/07-2013/02** **Software Developer (part-time); GoBusiness Finance**
- Financial Fetcher of Stock Values to train a Neural Network (Maximus-AI).
- Tasks:** ETL of financial data with preprocess hooks to train the neural network;

Skills

- Lang & Tools** **Python, R, Git, Docker, k8s, bash, Java, C#, Node.js, react.js, Scala**
- Data Store & Query Engine** **HDFS, PostgreSQL, Kafka, Parquet, Hive, Spark, Druid, Drill, Presto/Trino, KSQLDB, MySQL, Cassandra, Redis, Microsoft SQL**
- Data Viz** **Plotly, Shiny, Ggplot2, echarts4r, Redash, Apache Superset, Metastore, Tableau, Kibana**
- DS Tasks** **Classification, Labeled Evaluation, Blind Evaluation, Model Monitoring, EDA, Regression, Clustering**

DS Lifecycle **Airflow**, Mlflow, BentoML, Kubeflow, Seldon-Core, Kedro, DVC

Publications

2017-04 **Stator Winding Short-Circuit Fault Diagnosis in Induction Motors using Random Forest**; **IEMDC**

2017-05 **Severity Estimation of Stator Winding Short-Circuit Faults using Cubist**; **EPIA**

Teaching Experience

2019/09-2020/07 **Teaching Assistance**; **Lusófona Universidade**
Teaching Assistance for the subjects of Multimedia Data Science and Artificial Intelligence.

2015/03-2015/05 **Undergraduate Teaching Assistance**; **Nova Lisbon University, FCT**

2016/03-2016/05 Scholarship in which I've been teaching programming principles with Matlab.

Education

2018/09-2019/01 **Partial Post Grad, Analysis and Engineering of Big Data**; **Nova Lisbon University, FCT**
Multivariate Inference; Cluster analysis; Resampling and Confidence intervals; Optimization

2009/09-2013/07 **BSc and MSc, Computer Science (17/20)**; **Nova Lisbon University, FCT**

2014/09-2017/05 **Thesis (20/20)**: Machine Learning for On-line Detection of Stator Short-Circuit in Induction Motors

2020/10-2021/02 **Apache Kafka Series**(**Kafka**, **Streams/KSQL**, **Connect**, **Schema**); Udemy

2019/09-2019/10 **Big Data for Data Engineers Specialization**; **Coursera**

2019/09-2019/10 **Apache Airflow**; **Udemy**

2017/03-2017/08 **Machine Learning Specialization**; **Coursera**