Tiago dos Santos ▽ Machine Learning Engineer

Portugal, Lisbon • santos.tmd@gmail.com

Experience (9+ years)

2022/01-Present   **Machine Learning Engineer; Cleo**

Ownership of Machine Learning Lifecycle for the Salary Advance App Feature. Experimentation of new ML Problems to optimize business KPIs. Providing in-house tools for scaling up computation. A/B Test deployment and analysis, Shadow Deployment and Monitoring.

Tasks: Exploratory Data Analysis; Drift Exploration; Model Explaination; Model Deployment using AWS Sagemaker; Model Monitoring; On-line Feature Computation using Flink and Feast.

2021/07-2021/11   **Machine Learning Engineer; NannyML**

Support Machine Learning Research with frameworks (evaluation, model factory and search). MVP implementation of models, pipelines and infrastructure. Data Exploration and drift / anomaly detection. Implementation of an MVP monitoring pipeline. Product Design Research.

Tasks: Drift Exploration using PCA/Autoencoders; Implementation of pipelines using kedro; Implementation of APIs (FastAPI) and event-driven services (NATS-stream);

2021/01-2021/06   **Data Science Engineer; DefinedCrowd**

Infrastructure setup for model serving and serverless functions. Implementation of pipelines for model monitoring and artifact tracking. Implementing integration pipelines for model service documentation. Adapting model serving solutions for internal teams' usage.

Tasks: Deployment of KFServing, Nuclio; Model deployment with CI/CD; K8s logging to elastic;

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 (vehicle battery discharged - Random Forest, tire pressure - Variational Auto Encoder). Pipelines definitions for model deployment. Infrastructure creation for the Stratio AI project. Dealt with Gigabyte size data (json/parquet).

Tasks: Batch and stream infrastructure implementation; Create Ops Framework to unify project structure and deployment; Development and iteration of models; Development of dashboards for human feedback; Model monitoring and performance evaluation; Label Management;

2017/11-2019/03   **Data Science Engineer; Feedzai**

Development and production maintenance of a model (Random Forest) for *Non-Disclosure Client* to detect online payment fraud within 5 countries. Dealt with datasets having tens of terabytes size (csv/json, parquet). Dealt with fast data (circa 300 transactions p/s peak) Model comparison between Random Forest, XGBoost and LightGBM. Creation of Model Performance Reports for client.

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 (recall, fpr, auc, others) & retrain; Model Monitoring; Data analysis to identify fraud attacks; Pipeline optimization; Label Management;

2016/06-2017/06   **Data Science Consultant; Altran Portugal**

Development of a ML solution for short-circuit fault detection in Induction Motors. Dashboard for inference monitoring (R/Shiny). Model Inference module; Model Scoring service.

Tasks: ETL implementation for new data; EDA; Feature Extraction; Model Comparison (compared more than 10 models); Model Hyperparameter Tunning

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 frontent 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**, **k9s**, **bash**, Java, C#, Node.js, react.js, Scala
Data Store & Query Engine	**HDFS**, **PostgresSQL**, **Kafka**, **Parquet**, **RedShift**, **Spark**, Flink, Feast, Hive, Druid, Drill, Presto/Trino, KSQLDB, MySQL, Cassandra, Redis, Microsoft SQL
Data Viz	**Plotly**, **Shiny**, **Ggplot2**, echarts4r, Redash, Apache Superset, Periscope, Metastore, Tableau, Kibana
DS Tasks	**Classification**, **Labeled Evaluation**, **Blind Evaluation**, **Model Monitoring**, **EDA**, **Model Retraining**, Regression, Clustering
DS Models	**Random Forests**, **XGBoost**, **LightGBM**, SVM, Cubist, Variational AutoEncoder, Logit, LDA, K-NN, K-Means, NN
DS Lifecycle	**Airflow**, AWS Sagemaker, Mlflow, KFServing, Nuclio, OpenFaaS, BentoML, 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 2016/03-2016/05	**Undergraduate Teaching Assistance; FCT NOVA**
	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; FCT NOVA**
	Multivariate Inference; Cluster analysis; Resampling and Confidence intervals; Optimization
2009/09-2013/07 2014/09-2017/05	**BSc and MSc, Computer Science (17/20); FCT NOVA**
	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**