Work Experience

Dec 2019–Present: Senior Data Scientist - ECOM trading

At ECOM I am leading the data science team within the research department where we have created different data products (dashboards, automatic reports, web apps) and machine learning models to understand and predict key questions in the industry. For example, forecasting supply and demand of coffee, cotton and cocoa; forecasting productivity of cocoa farms in West Africa; forecasting the effects of climate change in the medium term on plantations of coffee and cocoa around the world; forecasting futures price dynamics in the market based on climatic and other market variables; and estimating crop yields using machine learning and satellite images. Most recently I developed a webapp that combines the historical deforestation and georeference location of farms into a a risk deforestation model for farms around the world. I have also been involved in the digital transformation of ECOM, promoting the use of data science approaches into different areas of the business.

Oct 2018–Nov 2019: Product Developer - Data Science, Decoded

At Decoded I designed and implemented more than 40% of the curriculum of the key product: the Data Academy. Our clients were blue-chip companies that were interested in integrating data science into their day-to-day activities. The programme focused on empowering and teaching their employees the latest techniques in data science. This programme was successfully delivered to banks such as Societe General, UBS and retailers such as Unilever, Nike, and M&S. I designed the materials to teach basic coding in Python or R, advanced statistics, implementation of supervised and unsupervised machine learning , time series analysis, and text analysis. The content contained detailed explanations of a great variety of techniques and general concepts in data science and data analytics such as supervised machine learning, bias-variance trade-off, and feature engineering. I wrote the R and Python modules of (1) Introduction to data analysis (using Pandas or Tidyverse); (2) Introduction to statistics (frequentist and Bayesian statistics); (3) Regression analysis (generalized linear models, general additive models, Lasso regression, Ridge regression, and elastic-net regression, Random Forest); (4) Classification methods (decision trees, Random Forest, XGBoost, Naive Bayes, Support Vector Machine); (5) Neural networks (using TensorFlow and Keras); (6) Big Data

(using Apache Spark); (7) Time series analysis, and (8) SQL. I regularly delivered workshops where the above modules are taught to students at the companies, both in-person and in webinars. I was also responsible for training new facilitators to deliver the modules that I have created.

2018 Jan–Oct : Research Associate - Data Science, Institute for Risk and Disaster Reduction, University College London

At UCL, I was responsible for creating a machine learning model to predict in real-time the location of potential mosquito breeding points in four different cities across Brazil (at ultra-fine resolutions). I used Python, TensorFlow and Keras to develop recurrent neural networks to model geospatially the occurrence of the virus and the mosquitoes in near real time. The results were displayed in a bokeh web app, (which I developed) used by health professionals in Brazil. During this position, I also developed a mobile app to gather data from the health professionals in Brazil.