Dr Juan D. Beltran

3 Dignum Street | London | N1 0FJ
jd.beltran43@gmail.com
LinkedIn | Github | Personal Website
Mobile number (+44) 07922 85 77 88

Summary

Since I started my data science career—spanning over seven years in diverse sectors, including the global soft commodities trading industry—I have developed an extensive understanding of machine learning, time series analysis, and Geographic Information Systems (GIS). These skills have been crucial in navigating intricate data patterns, thereby facilitating the provision of actionable insights in a complex trading/market ecosystem. My technical acumen, underscored by my proficiency in Python and R, has been instrumental in overcoming data-driven challenges in the commodities market.

My adept project management skills, refined through tangible experience with Agile methodologies and CI/CD principles, have allowed me to spearhead and manage cross-functional teams, driving business value through innovative strategies.

Furthermore, my leadership is rooted in my commitment to fostering a data-centric culture across diverse teams and mentoring upcoming talent. My problem-solving approach is characterised by a blend of intense curiosity and meticulous rigour, equipping me to tackle complex issues effectively. Coupled with my ability to articulate intricate concepts into comprehensible narratives, I ensure clear and empathetic communication with stakeholders. This fusion of leadership, problem-solving acuity, and communication prowess has been pivotal in my evolution as a data scientist.

Skills

Programming Python	Advanced proficiency in Pandas, GeoPandas, SciKit-Learn, TensorFlow, Pytorch, keras, Matplotlib, NumPy, SciPy, SQLAlchemy, Seaborn, Bokeh, SciKit-Image
Programming R	Advanced proficiency in RStudio, Shiny, tidyverse, Parsnip, bayestestR, sparklyr, keras, dygraph, forecast
Modelling	Advanced experience in statistical modelling, time series analysis, supervised and unsupervised machine learning algorithms, currently implementing solutions in reinforcement learning (intermediate)
Geospatial Analysis	Advanced skills in geospatial machine learning, professional proficiency in ArcGIS, QGIS and OpenGeos (Python)
Big Data & Databases	Proficient in SQL, MongoDB (NoSQL), Spark (pyspark and sparklyr)
Experimental Design	Strong background in designing and implementing A/B tests and Bayesian statistics
Languages	Fluent in English and Spanish (native)

Work experience

Lead Data Scientist | ECOM Trading | Nov 2022 - Present

- Currently leading the data science team at the Research department in ECOM Trading, one of the world's largest traders of coffee, cotton and cocoa. I started at ECOM Trading in 2019 as a Senior Data Scientist, I was promoted to the role of Lead Data Scientist in November 2022.

- My career at ECOM has been marked by pivotal contributions to our core product offerings. These contributions span the realms of data infrastructure and data centralisation as well as the creation, deployment and maintenance of machine learning models, which I describe in further detail below.

- I've played a significant role in the development and maintenance of advanced trading signal models that leverage Supply & Demand dynamics, weather data, and historical price trends. This unique blend of information allows us to predict market behaviour with high precision, creating a strategic advantage in our trading operations.

- Recognizing the critical importance of sustainability in today's market, I've been instrumental in developing a new framework for measuring deforestation on global farms using machine learning and satellite images. This initiative highlights our commitment to responsible trading and provides invaluable insights into environmental impact.

- To further streamline our processes and ensure data integrity, I've implemented an automated alert system that identifies potential anomalies in cotton, coffee and cocoa price data. This system proactively sends email alerts when irregularities are detected, allowing us to swiftly mitigate potential risks and maintain market-leading accuracy in our data-driven decisions.

Senior Data Scientist | ECOM Trading | Dec 2019 - Nov 2022

- Led the development of a comprehensive suite of data products, encompassing dashboards, automatic reports, and web applications in the research department as a senior data scientist.

- Developed an array of machine learning models to tackle key industry questions, from predicting the supply and demand of commodities like coffee, cotton, and cocoa to estimating crop yields using satellite imagery.

- Utilised my predictive models to forecast a range of pivotal factors, including the productivity of West African cocoa farms, the medium-term climate change effects on global plantations, and future price dynamics based on climatic variables.

- Conceived an innovative web application that fused historical deforestation data with georeferenced farm locations, forming a unique deforestation risk model for global farms.

Product Developer - Data Science | Decoded | Oct 2018 - Nov 2019

- Created and delivered over 40% of the curriculum for Decoded's flagship offering, the Data Academy, teaching data science integration to blue-chip clients, including Societe General, UBS, Unilever, Nike, and M&S.

- Crafted comprehensive course material covering fundamental coding in Python and R, advanced statistics, and a wide array of machine learning techniques.

- Authored specialised R and Python modules for an extensive range of topics such as regression analysis, classification methods, neural networks, Big Data handling, time series analysis, and SQL.

- Regularly hosted workshops to impart these modules to client employees, both in-person and via webinars, and trained new facilitators to carry the torch forward.

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

- Developed a machine learning model capable of real-time prediction of potential mosquito breeding points to prevent the spread of the Zika virus across four cities in Brazil, employing Python, TensorFlow, and Keras to create recurrent neural networks that geospatially modelled virus and mosquito occurrences.

- The results of my work were manifested in a user-friendly bokeh web application, which I built and was actively used by health professionals in Brazil. Complimenting this, I also created a mobile application to streamline data collection from health professionals on the ground.

Education

Degree	Institution	Duration	Details
PhD (DPhil) Plant Sciences	University of Oxford	2013 - 2017	Applied machine learning, computational modelling, and advanced statistics to analyse plant behaviour under drought stress and the impact of climate change on plant species distribution. Used various techniques including Neural Networks, XGBoost, Random Forest, Clustering analysis, and A/B tests using bootstrapping and Bayesian statistics.
MSc, Biology (1st place)	Universidad de los Andes, Colombia	2012	Utilised R to explore hypotheses about biochemical changes in juvenile plants under drought conditions.
BSc, Biology	Universidad de los Andes, Colombia	2010	Gained proficiency in R during biostatistics coursework and applied it extensively in my dissertation and subsequent work.

Awards

2013-2016:
Weidenfeld Scholarship and Leadership Program for DPhil. funding at the University of Oxford
2013-2017:
Scholarship from the Administrative Department of Science, Technology, and Innovation (Colombia) for DPhil. funding at the University of Oxford