**Diogo Ribeiro**
*Curriculum Vitae*

## PERSONAL DATA

*Birth Date:* July 23, 1975
*Adress:* 20 Barrow Road, SW16 5PF, London, UK
*Phone:* +351934385554
*Phone:* +447748679620

## ONLINE PRESENCE

*E-Mail:* diogo_dj@hotmail.com
*LinkedIn:* https://www.linkedin.com/in/diogo-ribeiro-9094604a/
*GitHub:* https://github.com/DiogoRibeiro7
*Hackerrank:* https://www.hackerrank.com/diogo_dj

## SUMMARY

I'm a data scientist with 11 years of experience in research and in leading new software projects, data mining and machine learning. I enjoy working on using Big Data stream mining and large-scale data analysis to gain insights from Big Data. I design and build applications on real-time data mining, at the intersection of mathematics, data science, algorithms, and business. My work Catalog spans across Europe, UK and US.

I have two degrees in mathematics, one in Pure Mathematics and the other in Applied Mathematics. This degree gave me an solid mathematical preparation, skills and ingenuity in mathematical modeling techniques like machine learning, computer fluid dynamics, data mining, time series and regression that can suit a wide range of applications in various branches of economy and industry and facilitated the use of computational and numerical techniques applications. With excellent skills in computer simulation using Monte Carlo methods.

I have extensive experience using applied statistics in machine learning with techniques such as deep learning, natural language processing, pattern recognition, artificial intelligence and neural networks (CNN/ ANN). My programming language competency includes, Python, R, Sql, Matlab and C/C++. My skillset also includes use of Qlik Sense, SSRS, Power BI, SSIS and Tableau.

Key Skills: Machine Learning Researcher, Quantitative Researcher, Finance, Neural Networks, Python, Gaussian Process, Bayesian Inference, Time-Series, Deep Learning, Reinforcement Learning, Probabilistic Models, Approximate Inference, Neural Networks, Natural Language Processing, Markov Models, Signal Processing, Computational Statistics, Econometrics, Mathematical Modelling, Predictive Modelling, Credit Risk.

# WORK EXPERIENCE

1. MySense from September 2019 to the moment
2. Live More Capital from October 2017 to December 2017
3. Obvious Capital from January 2016 to May 2019
4. Jefferson Midstate Auto Auction from January 2018 to July 2018
5. Piedade SGPS from June 2005 to September 2017

### Data Scientist/Data Engineer – MySense - September 2019

Senior Data Scientist at MySense.ai, my main responsibility are analysis of raw data, cleaning data, advise on implementing good data practices, planning collection of data, building pipelines to extract data, build feature engineering and models. I use Python, AWS, SQL, R and Qlik.

At the moment I applied models like Markov Chain Monte Carlo to analyze user behavior and sensor activation, Bayesian Statistics and Monte Carlo to predict users mobility, reliability for sensor life span and maintenance support, survival models for users expectation of life, and visualizations to the frontend. Structured a Data Lake to collect current data points, statistics and processed data, and data from sources like weather, demographics, social networks data, genetics.

### Data Scientist/Data Engineer – Obvious Capital - January 2016/May 2019

Calculating security risk factors and suggesting necessary precautions. Developing feasible and profitable investment models based on stock trading strategies. Work with Risk Analyst on enhancing hedging strategies and performing hedge effectiveness tests. Provide management with ad-hoc market risk analysis upon request. Participate proactively in ad-hoc and monthly stress testing. On a weekly basis, perform duties as assigned that will utilize risk and trading systems to monitor and measure department activities versus Market Risk limits including but not limited to Value-at-Risk (VaR), P&L back-testing and position limits.
Assist in developing and implementing a robust risk management program. Improvement of risk models to estimate exposure of portfolios to market factors.
Various projects in the field of financial modelling, for Obvious Capital, like for example, the modelling of cryptocurrency market using GARCH models with extreme value distributions to predict price ranges and minimize risk performance.
Performed research to develop trading systems and associate technologies, like deep reinforcement learning technologies to create Bitcoin trading bots. Provided efficient interface for financial engineering and technology teams. Implemented quantitative trade through various algorithms. Gathered market data for all portfolio managers and traders. Implemented high frequency trading algorithms for deployment. Analyzed trade data and researched all relevant information.
These tasks were performed using Python, R and SQL as main tools.

### Self-employed January 2017 – December 2017

Applied a CNN to predict house prices using images and geolocation, geolocation implied the gathering of data like distances to hospitals, schools and other features that helped predict a better price. The use of images of the houses were to predict maintenance costs. The business goal was to improve the accuracy of prices and get a better view over the asset to the client.

### Self-employed//Live More Capital – October 2017 – December 2017

Essentially has a problem of adaptions to the new regulations of loans to clients over 55 years old. The problem was to predict the probability of a client given is expenses and future expenses with health, predict time of death and if the surviving member of the couple will be capable of paying the mortgage. This involved in the first stage a statistical approach and when the model was tuned, it was improved using Neural Networks to improve accuracy in some of the features of the model, like the capacity of savings of the couple.

### Self-employed/Jefferson Midstate Auto Auction -January 2018 – July 2018

Implemented Hadoop data pipeline to identify customer behavioral patterns. Develop MapReduce jobs in Python for log analysis, analytics, and data cleaning. Perform big data processing using Hadoop, MapReduce, Pig, Hive, and Impala.
Import data from MySQL to HDFS, using Sqoop to load data. Regularly tune performance of Hive and Pig queries to improve data processing and retrieving.
The goal was to transform the business management with on time dashboards to get incites in to customers behavior and the cars that existed for sale.

### Data Scientist - December 2015/September 2017

*Piedade SGPS*

Responsible for production management and data analysis. My responsibilities consisted in developing data models to predict cycles of production capacity, production requirements to sale cycles and response times to clients purchase orders and production and expedition planning. Managed direct reports in 4 functional departments: demand planning, materials management, warehousing/transportation, and customer service/supply chain.

This task was accomplished using machine learning technics, statistical and probabilistic models. This algorithm were implemented using R, Python, C/C++ and Fortran. Also was required to transform a Excel sheets management in an SQL database management system. Fine-tuned and improved query performance using profiling tools and SQL.

Responsible for implementing forecasting methods and models. Responsible for forecasting 60 packaged products sold to 80 customers worldwide. Increased forecast accuracy by 12% over the last two years. Reduced days of inventory 40% over one year. Success attributed to improved controls, leveraging the ERP system. Reduced total inventory value by €500K by implementing inventory management tools. Reduced obsolete inventory by 25% in the last 2 years. Analyzed statistical data and reports to ascertain trends in performance, resulting in maximum effective and efficient use of logistical resources. Applied lean concepts using machine learning models to the distribution center to improve daily routes and reduce logistical spending by 18%.

**Internship - September 2014/December 2015**

*Faculty of Engineering of Oporto University, Civil Engineering Department, Investigation Unit Construct, with an graduation of 18 points.*

This internship give an experience on applications of my knowledge in real world and integration into the labor market. In this internship I applied linear regression technics, nonlinear regression, Lasso regression, Ridge regression and Neural Networks to model the heteroscedastic behavior of an self-compacting concrete. That implied also the use of classifications algorithms in the analysis of the residuals of the mean model. The main programming tools used were R and Python.

**Department Chief – Data Scientist – June 2005/December 2015**

*Piedade SGPS*

Chief of a cork crushed unit. Responsible for production management, quality control and staff manager. Responsible for quality production analysis to meet the requirements of this industry. That implied the creation of time series models to have a model to compare the real production against the model and calculate the seasonal variation of the production. At the same time was built some deterministic models using partial differential equation to model the thermic systems of disinfection of cork, later on upgraded to stochastic models.

## EDUCATION

Modela¸c˜ao de propriedades de pastas de bet˜ao autocompact´avel com comportamento heteroced´astico **Faculty of Sciences of Oporto University** 2015

This thesis was developed within CONSTRUCT Research Unit and carried out in the Structural Concrete Laboratory of the Faculty of Engineering of the Porto University. The aim of this study is to understand the influence of cement paste mix-proportions on the results Marsh Cone test and its relationship with the free water content measured by the centrifuge test.

**Master in Mathematical Engineering** 2015

*Faculty of Sciences of Oporto University*

Master with average of 14 points. This degree gave me an solid mathematical preparation, skills and ingenuity in mathematical modeling techniques like machine learning, computer fluid dynamics, data mining, time series and regression that can suit a wide range of applications in various branches of economy and facilitated the use of computational and numerical techniques applications.

**Specializations** 2014

*Faculty of Sciences of Oporto University*

Statistical Analysis in Chaotic Dynamics, 17 points. Mathematical Models in Economics and Finance, 10 points. Mathematical Models in Transport Phenomena, 16 points. Seminar in Mathematical Modeling, 15 points.

**Master in Mathematics** 2003

*Faculty of Sciences of Oporto University*

Master in Pure Mathematics with an average of 13 points.