Diogo Ribeiro

Curriculum Vitae

PERSONAL DATA

Birth Date	July 23, 1975
Adress	Rua Principal 958, 4505-375 Fiães, Portugal
Phone	+351934385554

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
Medium	https://medium.com/@diogoribeiro_94486
Facebook	https://www.facebook.com/Data-Science-Mathematics-261729694435409/

SUMMARY

I'm a data scientist with 10 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. With experience with Big Data in Hadoop using MapReduce, Pig, Hive, Cassandra, ElasticSearch and MongoDB. Also experience in Docker, Kubernetes, Flask and Django.

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

Data Scientist/Data Engineer 2016-2019

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.
Performed data mining, data cleaning & explored data visualization techniques on a variety of data stored in spreadsheets and text files using R/Python. Sufficient exposure to designing and developing Tableau reports and dashboards for data visualization using R, Python & Tableau. Hadoop big data ecosystems (MapReduce, HDFS, HBase, Zookeeper, Hive, Pig, Sqoop, Cassandra, Oozie, Talend). Proficient in writing Cloud Formation Templates (CFT) in YAML and JSON format to build the AWS services with the paradigm of Infrastructure as a Code. Experienced with event-driven and scheduled AWS Lambda functions to trigger various AWSresources. Worked with Docker container infrastructure to encapsulate code into a file system with abstraction and automation.

Freelancer/Obvious Capital 2016-2019
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. 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. Designed calibration of all exchange simulators.
To better understand coin correlations, we deployed an Affinity Propagation algorithm and found three distinct clusters of crypto assets using k means. I also used cluster analysis to find the best days to trade crypto coins.
Write Map Reduce Jobs, HIVEQL, Pig, Spark. Import data using Sqoop into Hive and Hbase from existing SQL Server. Support code/design analysis, strategy development and project planning. Create reports for the BI team using Sqoop to export data into HDFS and Hive. Develop multiple MapReduce jobs in Java for data cleaning and preprocessing. Involve in Requirement Analysis, Design, and Development. Export and Import data into HDFS, HBase and Hive using Sqoop. Involve in create Hive tables, loading with data and writing Hive queries which will run internally in MapReduce way. Work closely with the business and analytics team in gathering the system requirements. Load and transform large sets of structured and semi structured data. Load data into HBase tables using Java MapReduce.

Freelancer/Real Estate Company January 2017 – December 2017
Write Map Reduce Jobs, HIVEQL, Pig, Spark. Import data using Sqoop into Hive and Hbase from existing SQL Server. Support code/design analysis, strategy development

and project planning. Create reports for the BI team using Sqoop to export data into HDFS and Hive. Develop multiple MapReduce jobs in Java for data cleaning and preprocessing. Involve in Requirement Analysis, Design, and Development. Export and Import data into HDFS, HBase and Hive using Sqoop. Involve in create Hive tables, loading with data and writing Hive queries which will run internally in MapReduce way. Work closely with the business and analytics team in gathering the system requirements. Load and transform large sets of structured and semi structured data. Load data into HBase tables using Java MapReduce.

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.

Freelancer/Loan Company – 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.

Freelancer/ Auction House 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.setup, installed, and monitored 3-node enterprise Hadoop cluster on Ubuntu Linux. Regularly tune performance of Hive and Pig queries to improve data processing and retrieving.

Data Scientist 2015-2017

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. This task was accomplished using machine learning technics, like Linear Regression, Non-Linear Regression, Markov Chain Monte Carlo Models, LASSO Regression, Ridge Regression, Kernel Regression, Cluster Analysis, Neural Networks, Decision Trees, Evolutionary Algorithms, Genetic Algorithms or Artificial Neural Networks and other 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.

Managed direct reports in 4 functional departments: demand planning, materials management, warehousing/transportation, and customer service/supply chain.

Interfaced with data miners and analysts to extract, transform and load data from a wide variety of data sources using SQL, Python and R. Fine-tuned and improved query performance using profiling tools and SQL.

Owned the design, development maintenance of ongoing metrics, reports, analyses, dashboards, etc., to drive key business decisions and communicate key concepts to readers using Tableau, QlikView, ggplot2, Matplotlib, Plotly and Bokeh. 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 2014

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 2005-2015

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 desinfection of cork, later on upgraded to stochastic models.

EDUCATION

Modelação de propriedades de pastas de betão autocompactável com comportamento heterocedástico **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.