First Name: Arturo
Last Name: Del Valle

Principal Data Scientist at Mastercard |B.S, M.S Mathematics, Ph.D Industrial Engineering

ar2rodelvalle@gmail.com

This resume was automatically generated from: https://www.linkedin.com/in/arturo-del-valle-7b557b6a

---

Summary

Arturo is a data scientist who has a passion for telling stories through data.

Arturo is experienced in mathematical, statistical, and Machine Learning techniques with an ability to implement Complex theoretical knowledge into applied programming languages such as R, python, SQL, SPL using the latest big data technologies such as Hadoop, Hive, Impala, Splunk, Spark

Billions of observations (Terabytes) is a usual sized dataset in a regular workday.

Arturo is involved in all areas of data science: from data pipeline automation to modeling to deployment

---

Experience

Principal Data Scientist at Mastercard
March 2022 -

Lead Data Scientist at Mastercard
December 2017 - March 2022
Created a near-real-time data stream pipeline (~400Gb/day) that PREPARED raw transactional data by leveraging streaming capabilities in the pyspark.sql engine.

Created, tested, and validated pyspark models for anomaly detection with the objective to detect functional network anomalies in near real time with the objective to alert our Fast Emergency Response Center. Placed in production

Conduct data experimentation in R to arrive at deployable ML models at scale with limited access to R-libraries.

Evaluated new data science technologies on a regular basis as they are being considered for integration into Mastercard’s data science workbench.

Data quality exploration and modeling of massive IoT time series (1,000,000 rows per minute) by leveraging Splunk SPL language and its graphing capabilities.

Created a Splunk-to-Hadoop live pipeline to augment modeling capabilities available in our Hadoop Workbench.

Applied NLP approaches to IT-helpdesk tickets generated by our internal enterprise to return a prioritization and segmentation of ticket clusters.

Generated detailed technical documentation of modeling techniques by using the CrispDM workflow for reproducibility and replicability of the modeling process.

Participated with varied stake holders in order to acquire domain knowledge part of the project development

Participated and led meetings with numerous business partners for acquisition of domain knowledge and prioritization of assigned projects.

Training and mentoring of incoming junior data scientists to successfully onboard them in our proprietary Data Science workbench.

Experienced with tools Big Data tools: Hive, Impala, Hadoop, Spark, R, Pyspark, along with heavy bash scripting to automate their use.

Applied Statistician at Monsanto Company
July 2016 - December 2017
Used a variety of statistical tools as steps for data mining and insight process: Linear regressions, logistic regressions, ANOVAs, ANCOVAs, Mixed Effects models, and Spatial Statistics.

Answered experimental design questions by using linear mixed modeling from satellite imagery data with a focus in detection of diseased soybean plants for yield loss mitigation of up to 22%.

Applied machine learning techniques such as decision trees, boosting methods (gbm, bagged, RFs xgboost) to create prediction models applied in farms to spot zones which discover opportunities for +50% production.

Acquired experience with relational databases by procuring data from a variety of big data tools (i.e. APIs, Hadoop, S3) using querying methods such as: SQL, PostgREST, R scripting.

Acquired expertise in writing automated R-code scripts which generate data visuals in ggplot for data mining, analysis and data-output in experiments resulting in accelerated and standardized analytical workflows.

Performed aggregation/structuring of geo-spatial data to generate unified datasets ready for analysis.

Acquired experience in Unix commands by preparing servers by loading python, R, and git repositories using the command line / terminal to enable accelerated script execution.

Performed succinct summarization of workflows and analytical strategies to be presented to diverse leadership teams.

Research Assistant at New Mexico State University
June 2012 - May 2015
• Developed scripts in Matlab, and R for ad-hoc statistical analysis and Machine Learning using regression analysis, GLMs, Random Forests, Naïve Bayes, Bagged and Boosting Methods.

• Statistical modeling of wind energy outputs in turbine blade design experiments with a resulting +25% theoretical improvement in efficiency (results published)

• Implemented Matlab scripts for analytical modeling of wind energy production, and utilized a combination of gradient-based methods (SQP) with evolutionary algorithms (GAs) for design optimization with a resulting +35% increased efficiency.

• Created an interface of Visual Basic/Matlab/Siemens software for automatic mechanical test simulations of device components resulting in an accelerated analytical workflow.

• Implemented C++ code in Matlab for faster script execution using MEX programming resulting in a 10-fold speed performance improvement.

• Conducted a Matlab automation of Fortran code ‘XFOIL’ for automatic data gathering of airfoil properties.

Mathematics Instructor at El Paso Community College
May 2007 - May 2015
Preparation and presentations of mathematics undergraduate classes such as: Mathematics for Business I and II, Pre-Calculus, Descriptive and Inferential Statistics.

Classes lectured in a variety of universities including: El Paso Community College, University of Texas at El Paso, Universidad Autonoma de Cd. Juarez, Universidad Tec. Milenio.

Developed excellent communication skills with non-technical audiences for motivational purposes aimed towards student success.

Education

New Mexico State University
Doctor of Philosophy (Ph.D.), Industrial Engineering, 2012 - 2016

The University of Texas at El Paso
Master of Science (M.S.), Mathematics, 2005 - 2007

The University of Texas at El Paso
Bachelor of Science (B.S.), Mathematics, 1998 - 2005