RAUL G. MARTINEZ

@ gio.mtz3@gmail.com | +1 619-734-6231 | San Diego, CA | raul-giovanny-martinez-688b1bb4 | github.com/raulgiovannymartinez

WORK EXPERIENCE

Data Engineer, Illumina

July 2018 – Present | San Diego, CA

- Developed data integration, ETL/data pipelines, and data modeling for instrument data (semi-structured, unstructured) across all Next-Generation Sequencing (NGS) platforms with both programming (i.e. python) and application (i.e. alteryx, SQL server) based approaches.
- Enabled ad hoc analytics for system troubleshooting and root cause investigations for the different NGS core components including optics, hardware, fluidics, software, and chemistry.

Pharma Technical Development Intern, Genentech/Roche

February 2018 – July 2018 | South San Francisco, CA

- Collaborated in process development viral clearance studies (i.e. virus removal, inactivation) for the production of therapeutic monoclonal antibodies.
- Investigated and optimized virus titer infectivity assays by implementing design of experiments (using SAS JMP), inferential statistics, and regression analysis.

Software Support Specialist, Dotmatics

December 2017 – February 2018 | San Diego, CA

- Collaborated with Software Developers and Application Scientists to deliver scientific informatics product suite solutions by testing software front-end, software architecture, and Oracle relational databases.

Data Analyst, Retrovirox

May 2016 – July 2017 | San Diego, CA

- Implemented data analysis and processing workflows for statistical analysis (i.e. linear regression, Z-factor) in high-throughput screening of hundreds of novel small-molecules, using FACS Flow Cytometry techniques, for measuring antiviral activities.
- Executed data management and documentation for R&D experiments in support to data integrity, reporting, visualization, and experimental design.

PROJECTS AND RESEARCH

PyTorch Deep-Learning Time Series Forecasting Library (2021)

- Develop an efficient and user-friendly deep learning open-source library that would benefit the entire PyTorch research community.
- We will integrate and unify some research code from Dr. Qi (Rose) Yu's lab, and others, in order to include both state of the art research and benchmark reports.

Drug Epidemic Tracking Through Social Media (2020)

- Built knowledge graph with Data Integration, NLP, and Entity-Matching strategies such as Blocking, Similarity Functions, Feature Vectors, and ML-Matchers.
- Consolidated a graph database schema (in Neo4j) to explore drug abuse and overdose patterns, with Cypher declarative queries, across the data ingested from Reddit posts and US News articles.

Predicting Helpfulness of Amazon Reviews (2020)

- Analyzed numerical and text data from 200,000 reviews in order to extract useful information with feature engineering and NLP techniques. Feature selection was done with an ablation study.
- Developed a Random Forest Regressor Model to predict the ratio between the number of helpful votes with total votes, performance metrics used include MAE, MSE, precision, and recall.

Identifying Oncogene-Specific Essential Genes (2016)

- Evaluated a bioinformatics prototype using R, based on distributional entropies and kernel-based density estimators to find drug sensitivity profiles that match patterns of gene expression with over 10,000 variables.

EDUCATION

Masters Degree in Data Science and Engineering, GPA 4.0/4.0

September 2019 – June 2021

University of California, San Diego

Bachelors Degree in Bioengineering Biosystems, GPA 3.4/4.0

October 2014 – June 2017

University of California, San Diego
Minor: Entrepreneurship and Innovation

PROFESSIONAL SKILLS

- Domain Related

Data Warehouse | Data Lakes | Data Models | Data Integration | Database Design | ETL | BI Reporting | Statistical Techniques | ML/DL

- Programming and Software

Python | Scala | R | C#/C++ | PyTorch | SQL/NoSQL | APIs | HPC | OOP | AWS | Hadoop/Spark | GitHub | Tableau | Alteryx | Denodo | SAS JMP

RELEV. COURSEWORK

Data Visualization | Data Integration and ETL | Beyond Relational Data Models | Machine Learning | Data Analysis Using Hadoop and Spark | Probability & Statistics Using Python | Data Management Systems | Python for Data Analysis

ACHIEVEMENTS

- Chancellor's Associate Scholar - UCSD (2014-2017)
- Community College Scholarships - "ASO Campus Involvement" and "Neva Smith" (2014)