DANA ROCHA

858-353-4055 dbridgette.rocha@gmail.com linkedin.com/in/rochadana

EDUCATION

Northeastern University **Boston, MA**
Master of Science in Bioinformatics and Graduate Certificate in Data Analytics (3.66/4.0) *Dec 2020*
Relevant Coursework: Algorithms, Bioinformatics Programming, Data Mining/Machine Learning

Northeastern University **Boston, MA**
Bachelor of Science in Biochemistry *Aug 2018*

TECHNICAL SKILLS

Languages: Python, R, SQL, HTML5, CSS3, JavaScript
Tools: PyCharm, Git, RStudio, IntelliJ, Spotfire, Pipeline Pilot, DBeaver, Oracle SQL Developer
Databases: Oracle, MySQL

TECHNICAL EXPERIENCE

Merck & Co. **Boston, MA**
Informatics Co-op – MRL Computational and Structural Chemistry *Jan – Aug 2020*

- Trained machine learning models to identify relationships between chemical compounds, genes, and disease
- Built workflows to process chemical data from public and internal databases using Pipeline Pilot and SQL
- Delivered knowledge-based compound sets to three drug discovery research programs in Chemical Biology
- Created interactive data visualization dashboards using Spotfire for over 140,000 compounds and genes
- Led presentation of data insights to chemists and data scientists within Chemical Biology and Informatics

RESEARCH EXPERIENCE

Orig3n Inc. **Boston, MA**
Research Assistant Co-op *Jul – Dec 2017*

- Maintained stem cell cultures for internal assays and animal studies
- Characterized stem cell cultures with fluorescence microscopy and flow cytometry
- Represented Orig3n as a brand ambassador at events in New York City, Baltimore, and Lake Tahoe

Brigham and Women's Hospital **Boston, MA**
Lab Technician Co-op – CCI Specimen Processing Lab *Jun – Dec 2016*

- Processed specimens for clinical studies conducted by the CCI Units and Emergency Department
- Completed all documentation required for individual specimens according to research protocol requirements
- Maintained laboratory instrumentation to performance standards

PROJECTS

Heart Failure Prediction *Dec 2020*

- Created classification models to predict survival of heart failure patients using R with over 81% accuracy
- Trained, tested, tuned classification models using k-NN, Support Vector Machines, and ANN algorithms
- Built a bootstrapped ensemble model with 86.66% prediction accuracy

Mapping Manhattan *Jun 2020*

- Created Python script to compute distance to nearest subway stations using Pandas, Scikit-learn, and Numpy
- Visualized data with an interactive choropleth map using the Plotly library in Python
- Processed data from NYC MapPLUTO dataset and New York City Capital Planning Platform