John Corio

973-896-6625 | corioj@umich.edu | https://corioj.github.io/ | https://www.linkedin.com/in/john-corio-63b128196/

Education

University of Michigan: College of Literature, Arts, and Sciences — Ann Arbor, MI
Bachelor of Sciences in Data Science, Minor in Mathematics — *Aug. 2017 – May 2021*

Experience

Client Engineering Data Scientist — Mar. 2022 – Present
IBM — *New York, NY*

- Increasing IBM software sales by engaging in pre-sales data science related proofs of concept and MVPs to rapidly solve clients' business problems and create further business opportunities for IBM
- Developing machine learning assets using a variety of programming languages (Python, Scala, R), standard open-source libraries and tooling (SKLearn, Tensorflow, Pandas, etc.), as well as engineering proprietary software solutions to address individual business use cases.
- Serving as subject matter expert on ML, AI, statistical models and data science for client engagements, and how they can be applied to solve individual clients' business problems

Data Quality Assurance Analyst — Dec. 2020 – Feb. 2022
ImageCare Centers — *Newton, NJ*

- Drafted first designs of a proprietary SQL database specifically for analytics, migrated large datasets from a third-party software vendor, and ran quality assurance via SQL queries
- Created SQL queries, tables, and views to pull data and evaluate data to be used in analytics informing on key KPIs

Technical Skills

Languages: Python, C++, C#, Scala, SQL, R, JavaScript, HTML/CSS
Libraries: Apache Spark, SKLearn, PyTorch, Tensorflow, Pandas, NumPy, Matplotlib, C++ STL

Projects

Album Art Generation via Neural Networks | *PyTorch, Pillow, OpenCV, Matplotlib* — Nov. 2020 – Dec. 2020

- Generated album covers from random noise tensors designed to fit the aesthetics of albums from given genres such as rock, pop, and others
- Designed a script to query the Spotify API using the SpotiPy library, extract desired data from JSON objects, and export cleaned image dataset to a preprocessing routine using Pillow and OpenCV
- Implemented 3 different generative adversarial neural networks based on computer vision research publications and individual design, then evaluated them based on image generation clarity

Sentiment Classifier of Yelp Reviews | *SKLearn, Matplotlib, Pandas, NLTK* — Sep. 2020

- Achieved top 10 percent in class on testing dataset accuracy for an NLP-based classification model of Yelp reviews in Python-based Jupyter notebooks
- Underwent a full machine learning life cycle by iteratively engineering an NLP data corpus, created an automated framework for training, optimizing, and evaluating Support Vector Machines models, and reviewing validation set results.

Asset Return Predictor | *R, ggplot* — Sep. 2020

- Placed second out of forty five teams on final project test dataset performance for predicting the forward returns of stock market assets
- Preprocessed time series finance data using various R libraries and self-made functions, then fitted and evaluated the holdout dataset performance of a variety of models such as KNN, Regression, Principal Components Regression, Random Forests, Bagging, and Boosting

Traveling Salesperson Problem | *C++, STL* — Dec. 2019

- Created custom coordinate data structures from input data and designed a map to display the information
- Implemented optimal and fast solutions for the Traveling Salesperson utilizing branch-and-bound algorithms, minimum spanning trees, and Prim's algorithm