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
BSc in Data Science, Minor in Mathematics *Aug 2017 – May 2021*

Experience

Client Engineering Data Scientist Mar 2022 – Jun 2023
IBM *New York, NY*

- Developed machine learning assets using various programming languages (Python, Scala, Java, SQL), open-source libraries and tooling (SKLearn, Tensorflow, Pandas, Spark, etc.), and proprietary software solutions to address individual business use cases and increase software sales.
- Contributed to requirements gathering, design, and implementation of regression models in Scala Spark API to transition a major banking client's overdraft system that mediates millions of commercial and personal transactions per day from rule-based architecture to AI.
- Utilized Python Jupyter notebooks, open-source libraries, and proprietary software to create a document analysis asset that performed document sectioning, entity extraction, and section classification using BERT and rules-based algorithms.
- Completed multiple client MVPs to help achieve the most successful quarter for software sales in IBM Financial Services Market history.

Data Quality Assurance Analyst Jun 2021 – 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

Japanese Language Student Jul 2023 – Present
ISI Language School *Tokyo, JP*

- Studying Japanese language at a foreign language school in preparation for taking the JLPT N2 exam.

Projects

Album Art Generator | *PyTorch, OpenCV, Matplotlib* Nov 2020 – Dec 2020

- Implemented 3 different generative adversarial neural networks based on computer vision research publications and individual design to generate album covers from random noise tensors designed to fit aesthetics of 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

Yelp Review Classifier | *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
- Engineered an NLP data corpus, created an automated framework for training and optimizing hyperparameters, evaluating support vector machines and deep learning models, and reviewing validation set accuracy 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, statistical models, Regression, Principal Components Regression, Random Forests, Bagging, and Boosting

Technical Skills

Languages: Python, C++, C#, Scala, SQL, R, JavaScript, HTML/CSS
Libraries: Apache Spark, SKLearn, PyTorch, Tensorflow, Pandas, NumPy, Matplotlib, C++ Standard Template Library
Developer Tools: Git, Xcode, Visual Studio, VS Code, Jupyter Notebooks, Google Colab, IntelliJ **Relevant Coursework**: Linear algebra, multivariate calculus, statistics, data structures and algorithms