

Qi Dong

Jersey City, New Jersey • 6468661808 • qd2046@nyu.edu
www.linkedin.com/in/qi-dong-6a8b15170/

EDUCATION

New York University

May 2023

Master of Science in Data Science, GPA: 4.0/4.0

Relevant Coursework: Machine Learning, Big Data, Introduction to Data Science, Programming for Data Science, Optimization and Computational Linear Algebra, Computational Cognitive Modelling

Rensselaer Polytechnic Institute

Dec 2020

Bachelor of Science in Industrial Engineering | Minor in Psychology, GPA: 3.97 / 4.0

Relevant Coursework: Computer Science, Information System, Modeling and Analysis of Uncertainty, Statistical Analysis, Introduction to C++, Operations Research Method, Discrete Event Simulation Modelling

PROFESSIONAL EXPERIENCE

Load Forecasting & Analytics Intern - National Grid, Hicksville, NY

May 2022 – Aug 2022

- Performed Exploratory Data Analysis on day-ahead gas load forecasting data using Pandas, cleansed & transformed the raw data and visualized trends in data with Matplotlib, Seaborn, and Plotly in Python.
- Detected data anomalies and coordinated with data source provider to propose & implement anomaly corrections, corrected 2 aberrations in gas load data and 3 irregularities in weather data.
- Built neural network machine learning models using Keras and PyTorch in Python on gas load time-series data to forecast short-term gas loads, applied hyper-parameter tuning, compared model performance with existing regression models and concluded that neural network outperformed regression models in most of the regions.

Logistics Intern - BMW Group, Beijing, China

Sep 2019 – Dec 2019

- Facilitated arrival & delivery of items inside the warehouse, organized inventories & monitored ~80% of incoming deliveries daily, maintained a log of all receipts and disbursements.
- Conducted cost benefit analysis by estimating flow-in and flow-out units, full-time equivalent labor cost, space utilization and facility cost to calculate net present value of the new warehouse design.

RESEARCH PROJECT

Generalized Wind Power Prediction Modelling | R, Python, Regression

Jan 2020 – May 2020

- Cleansed & transformed raw windfarm data, imputed missing values & treated outliers using R in RStudio to create a data asset for predicting wind power output using machine learning algorithms.
- Explored the data & identified 11 target windfarms for power prediction study, built a data ETL & engineering pipeline in R to gather and process the raw data into metric like hourly average, curtail, etc. for windfarms.
- Build a benchmark model with variables like air density, wind speed, and swept area of turbine blades using Linear Regression as a foundation for further development, achieved MAPE of 11%.

ACADEMIC PROJECTS

Movie Recommender System | Python, Spark, Recommender

Jan 2022 – May 2022

- Built a Baseline Popularity, Spark Alternating Least Square, and LightFM models for movie recommendations using 1.2M records from the MovieLens dataset.
- Implemented hyperparameter tuning & cross-validation, evaluated models using Mean Average Precision, achieved best MAP of 0.182 with LightFM.
- Assessed model performances on small & large datasets & recommended models suited to different situations.

Movie Ratings Analysis | Python, PCA, Clustering, Classification, Hypothesis Tests

Sep 2021 – Dec 2021

- Analyzed the impact of movie categories on movie ratings with hypothesis testing in Python.
- Performed correlation analysis & linear regression to determine factors that influence movie recommendations with a testing error of 1.2%.
- Leveraged Principal Component Analysis for feature reduction & kMeans to identify 4 clusters in the data.
- Constructed Nearest Neighbor & Neural Network classifiers to predict movie ratings using the clusters, evaluated model performance with confusion matrix, achieved 98% precision and recall.

SKILLS & CERTIFICATIONS

Programming Languages: Python (NumPy, Pandas, Matplotlib, SciPy, Scikit-Learn), SQL, R, C++

Analytical Tools: MySQL, RStudio, Jupyter, MS Excel, MS Access, MATLAB, Minitab, Visual Studio, Arena, CAD

Certifications: Machine Learning | Coursera, Advanced SQL for Data Scientist | LinkedIn, Six Sigma Green Belt | IISE