Nishan Mann | Data Scientist

New York, NY

+1 (929) 204 8317 • nishan.singh.mann@gmail.com
http://www.linkedin.com/in/nishanmann • http://www.github.com/NahsiN

Experience

automotiveMastermind
Data Scientist *May 2017–Present*

- Used constrained logistic regression (implemented via projected gradient descent using tensorflow) to devise a ranking method based on purchase likelihood for retention customers of dealerships (customers present in the dealership's database). Demonstrated $\sim 135-175\%$ lift in sales for the top 10% of ranked customers in the dealership's finance portfolio. Dataset dimensions: 1 million customers and 20 features.
- Performed univariate data analysis of vehicle maintenance related features to examine their potential for increasing sales in a dealership's finance and cash portfolios.
- Used logistic regression with L1 regularization for ranking conquest customers for BMW dealerships (customers not present in the dealership's CRM). Demonstrated 150% lift in sales for the top 10% of ranked customers. Imbalanced dataset with 70 million customers and 600 features.
- Using test driven development, collaboratively stood up a microservice on Kubernetes responsible for batch ingesting $500GB$ of data from multiple Pub/Sub message streams into to a PostgreSQL database.
- Using Prometheus metrics, created a dashboard in Grafana for performance monitoring of microservice.

Insight Data Science
Fellow *Fall 2016*

- Built SafeWalk (`safewalk.ddns.net`): a routing app that uses NYPD's major crimes data from 2005–2016 to suggest safe walking routes for any time of day.
- Used pgRouting and PostGIS in PostgreSQL to build an undirected graph of every road in NYC ($700,000$ roads from Open Street Maps), then used k-nearest neighbours to associate each of the $1,000,000$ major crimes with a road.
- Developed a cost function for Dijkstra's routing algorithm that accounts for crime intensity and the user's crime avoidance preferences.

Department of Physics, Queen's University
Doctoral Researcher *2011–2017*

- Used numerical integration methods to compute approximations of analytical expressions derived from first principles (Maxwell's equations).
- Used Monte Carlo methods and Gaussian processes to numerically generate disordered waveguide samples having an Ornstein-Uhlenbeck covariance function.
- Developed and implemented an implicit finite difference scheme using Python for solving coupled nonlinear partial differential equations with non-standard initial conditions.
- Used numpy's memory mapped files for optimized disk caching of large datasets (200-600GB) generated during a simulation.
- Achieved a 45% reduction in runtimes by exploiting multithreading capabilities of Intel's Math Kernel Library.

Tools

Statistical Methods: Logistic Regression, Random Forest, Kernel Density Estimation, Clustering, Hypothesis Testing, Principal Component Analysis

Data Wrangling: Scipy, Pandas, Dask, Scikit-learn, Matlab, PySpark, BigQuery, Hive, Tensorflow, AzureSQL, PostgresSQL

Development Languages: Python, R, Javascript, CUDA, C

Cloud Tools: Docker, Kubernetes, Prometheus, Google Cloud, Azure, AWS

Education

Ph.D. in Physics **2011–2017**
Queen's University *Kingston, Canada*

MSc in Mathematical Modelling in Engineering **2009–2011**
University of L'Aquila, University of Nice and Politechnika Gdańska *European Union*

HBSc in Physics *with distinction* **2005–2009**
University of Toronto *Toronto, Canada*