Jiawei Yu (Charlie)

Data Scientist

519-729-9479
Jiawei.yu@mail.utoronto.ca
jiawei-yu-97.github.io
linkedin.com/in/charlieyu7/

Education

Master of Science in Applied Computing (Data Science Concentration) *2020-present*
University of Toronto

Bachelor of Mathematics in Data Science *2015-2020*
University of Waterloo

- Cherry Statistics Course Award
- George Barnard Statistics Prize

Experience

Data Scientist Intern *05/2021 – 12/2021*
Geotab, Waterloo

- Designed, Implemented, and validated pipeline for detecting harsh driving events, using Airflow and BigQuery.

Teaching Assistant *09/2016 – 04/2021*
University of Waterloo / University of *Toronto*

- Hosted tutorials and office hours. Marked assignments, tests, and projects
- Courses: Introduction to AI; Algebra; Linear Algebra; Statistics; Computational Thinking

Data Scientist Co-op *10/2019 – 12/2019*
Carta, Waterloo

- Churn prediction using feature engineering and GLM models.
- Answered business intelligence questions with complex SQL queries.

Data Scientist Co-op *09/2019 – 10/2019*
Kik Interactive, Waterloo

- Managed data dashboard to monitor spammers.
- Data exploration and design of manual rules to detect and categorize spammers.

Software Developer (Data Engineering) Co-op *01/2019 – 04/2019*
Miovision Technologies, Waterloo

- Experimented with different datasets and techniques to improve object detection model. Examples include improving night performance, improving under-represented classes, and models fitted to specific backgrounds. Gains ranged from 15% to 30% improvement in F1 score.
- Implemented tight rotation (of images and bounding boxes) as an addition to TensorFlow's data augmentation library for object detection in Python.

- Designed protocols to expand datasets with continuous data inflow, while maintaining train-valid integrity and fair grounds for model comparisons, resulting in improved model evaluation.
- Front-end development in Python and JavaScript for data and model visualizations.
- Relational database design, database transition from MongoDB to PostgreSQL and data backfill.

Junior Data Scientist Co-op   01/2018 – 04/2018
Intelligent Mechatronic Systems, Waterloo

- Data cleaning & transformation for battery health prediction using random forest.
- Built internal python libraries for data preprocessing, quality control and sanity check.
- Designed and verified algorithms to detect and predict unlabelled events such as gear shifts.
- Applied K-Means to distinguish hybrid and traditional cars, achieved 95%+ recall and specificity.

Rail Operations Co-op   01/2017 – 04/2017
Metrolinx, Toronto

- Analyzed the on-time performance of GO trains, summarized major incidents, and provided statistics for decision-making in platform adjustments and schedule modifications.
- Improved reporting style by writing macro commands to replace pivot tables and human effort.
- Refactored old Excel VBA codes to shorten runtime by 30-50%.
- Designed new report templates and automated reporting with VBA.

Project Management Associate
PureLiving China (Shanghai) Ltd, Beijing
05/2016 - 08/2016

- Delivered audits, proposals and work reports for clients.
- Assisted engineers with AutoCAD design.
- Installed, configured, debugged and repaired air monitoring devices onsite.