BO WEN WEN

SENIOR DATA SCIENTIST

5 years of experience in predictive modeling and applied machine learning. Expertise in building data-driven software using statistical and machine learning models in Python and R, to support business cases. In my spare time, I enjoy hiking and managing my home server with Docker and Kubernetes.

EXPERIENCE

2020.06 – present

SENIOR MODELER – Forecasting, Research & Analytics, TransLink (SCBCTA)

- Designed procedures and developed software tools to evaluate the impacts of the COVID pandemic with over 8 risk factors on regional transportation, using uncertainty analysis based on Latin Hypercube sampling and meta-model simulations, to inform revenue projection scenarios of a $2-billion regional transportation authority.
- Trained nested multinomial logit models for choice behavior and linear regression models for trip demand in R and Python capturing the effect of transportation accessibility on travel behavior, which are used in production environment to forecast travel behaviors of Metro Vancouver residents.
- Improved model testing and deployment processes by integrating Travis CI into the GitHub repository to automate code validation and minimized manual testing.
- Built a prototype of transit passenger counting and load prediction system using transit GPS data to optimize operational crowding levels, which won the Grand prize at TransLink's 2020 Garage Day hackathon.

2018.06 – 2020.06

MODELER – Forecasting, Research & Analytics, TransLink (SCBCTA)

- Evaluated the consumer benefits and costs of infrastructural investments using TransLink's in-house decision support transportation planning model (Regional Transportation Model).
- Processed travel survey data, using random forest and k-NN in Python to identify outlier data records, and using quasi binomial glm in R to reduce sampling bias.
- Authored Tableau dashboard of travel surveys with over 450k sample trips, to report travel behavior changes to stakeholders and policy makers.
- Built a transit trip planning chatbot in Python and Docker using the Rasa NLU, integrated with Facebook Messenger and Slack, which won the "Facelift" award for TransLink's 2019 Garage Day hackathon.

2017.09 – 2018.06

DATA SCIENTIST – IBI Group Inc.

- Built a data mining application in Python that aggregated Land Development Permit data from various cities to inform development potential in land use planning.
- Produced statistical and machine learning models using SQL and Python to evaluate pump performance and detect failure events of industrial SCADA control systems.

2014.09 – 2017.09

RESEARCH ASSISTANT – University of Toronto, Engineering

- Evaluated the model performance of artificial neural networks (NNs), support vector machines, linear mixed effects, regression trees, and random forest models, using caret in R, and scikit-learn in Python, to predict travel speeds of public transit vehicles.
- Produced a tool in C# which ingests streaming data from multiple web APIs to MS SQL Server, reducing data collection time by 90% for 3 applied science research projects.
- Developed an end-to-end data mining, machine learning, and simulation software for the Toronto transit system in C# and R.
- Developed a reinforcement learning (RL) agent in Java and traffic simulation script in C in a simulated environment to minimize passenger traffic signal delays.

2015.05 – 2015.08

TRANSPORTATION ANALYST – IBI Group Inc.

- Reduced processing time for traffic data by over 25% and improved maintainability.
- Wrote technical user manual for the NITCIP 1211 protocol hardware compliance.

APPOINTMENTS

2020.05 – present

COMMITTEE MEMBER – AI and Advanced Computing Committee AED50, TRB NASEM

- Assisted with subcommittee activities such as the development of the Machine Learning Primer for Transportation and the Triennial Strategic Plan (TSP).
- Conducted peer review of over 25 papers for the Transportation Research Board

CERTIFICATIONS

2021.11	CONTINUOUS DELIVERY AND DEVOPS	University of Virginia on Coursera
2021.09	PROJECT PLANNING AND MANAGEMENT	University of Virginia on Coursera
2019.10	DISCRETE CHOICE MODELS	EPFL on edX
2018.01	APPLIED PLOTTING IN PYTHON	University of Michigan on Coursera
2018.01	DATA SCIENCE IN PYTHON	University of Michigan on Coursera
2017.07	MACHINE LEARNING	Stanford University on Coursera

EDUCATION

2015.09 – 2017.09

MASTER OF APPLIED SCIENCE – University of Toronto cGPA: 3.94/4

- Civil Engineering, Transportation (Applied Machine Learning)
- Thesis: Data-driven mesoscopic simulation of large-scale surface transit networks

2010.09 – 2015.05

BACHELOR OF APPLIED SCIENCE – University of Toronto cGPA: 3.76/4

- Civil Engineering, Minor in Business
- Thesis: Reinforcement learning-based adaptive traffic signal control system for transit