Jingyi He

jingyi.he@mail.mcgill.ca
514-661-0316
kylie-box
kylie-box.github.io

Education

BSc | Statistics & Computer Science

McGill University | June 2017 - May 2020(expected)

- CGPA: 3.86
- Jacqueline Johnson Desoer Science Undergraduate Research Award

BSc |Dietetics

MacDonald Campus, McGill University | September 2015 - May 2017

- Eliza M Jones in course scholarship
- Dean's Honour List

Developer Skills

Programming Languages

- Proficient: Python, Java, R
- Capable: C, SQL, OCaml, Matlab, bash

Utilities

- Git, LaTeX, PyCharm, Eclipse

Selected Coursework

Computation

Algorithms and Data Structures
Algorithm Design
Numerical Computing
Database Systems
Applied Machine Learning
Natural Language Processing
Probabilistic Graphical Models

Stat & Math

Probability
Statistics
Regression and analysis of variance
Sampling Theory
Linear algebra
Abstract Algebra
Analysis
multivariate Calculus
Mathematics in Machine Learning

Misc

Languages

- Bilingual or Native Proficiency:
  English
  Mandarin
  Cantonese

Volunteer

- Kids Code Jeunesse - Taught 30 young kids to write python programs.
- People's Potato - Food prep at the soup kitchen.

Experience

Mila(Quebec Artificial Intelligence Institute) | Research Assistant

May 2019 - Present

- Research Project: Sampling Methods for Training Word Embeddings (Supervisor: Prof. Jackie C. K. Cheung)
- Investigated, derived and implemented three novel sampling based machine learning algorithms to learn word representations that capture linguistic features of words on 5.4 million tokens corpus using Pytorch library.
- Substantially improved the efficiency, scalability and stability of existing word embedding training algorithms.

China United Network Communications Group | Data Analyst Intern

summer 2018

- Predicted the loss of customers up to 97% of accuracy on held out test sets by training SVM, XGBoost and Neural Network etc. models using Scikit-learn library.
- Applied statistic and machine learning knowledge to efficiently analyze and select user usage features of high correlation among more than 300 raw features using Pandas library.
- Enhanced data visualization using Seaborn and Matplotlib libraries.

Projects

Semi-supervised learning of Tweets Sentiment Analysis During Californian Campfire Fall 2018

- Scraped over 5,000 raw tweets data related to the 2018 Californian Campfire using Twitter API.
- Well-documented and published the clean dataset for the community interested in low resource natural disaster corpus research.
- Implemented semi-supervised learning algorithms that overcome the low resource caveat to analyze the sentiment of the population towards 2018 Californian Campfire over time.

Google "Quick! Draw!" Hand-drawn picture recognition

Fall 2018

- Implemented CNN models to classify hand-drawn pictures of 31 classes.
- Achieved 78.2% of accuracy comparing to baseline of 3% accuracy.
- Ranked top 15 among 50 groups in class.

Private VPN server with OpenVPN on AWS May 2018

- Hosted a secure private VPN server using OpenVPN on AWS EC2 from scratch.
- Familiarized myself with fundamental network routing concepts and basic AWS cloud services.
- Have been providing and actively maintaining the VPN service to a small group of users for more than a year.

Smart City IoT: Pothole Detection | McHack(24 hrs Hackathon) project

February 2018

- Awarded the Telus IoT Prize
- Built a web app prototype to improve the efficiency of collecting road construction information in Montreal by providing simulating crowd source data aggregation from sensors in individual cars.