Yun Han Xiao

https://www.hanxiao.ca
yhxiao@edu.uwaterloo.ca | https://linkedin.com/in/hanburger

RELEVANT SKILLS

Proficient:
Python • C/C++ • R •Tensorflow • Docker (Swarm) • Apache Kafka • SciPy/Numpy/Pandas • Javascript • MongoDB
Familiar:
SQL • PyTorch/Keras • Apache Zookeeper • Hadoop (Horton Works) • CUDA/NVCC • OpenAI

EXPERIENCE

INTACT FINANCIAL CORPORATION | Data Lab Intern
May 2018 – August 2018 | Montreal, QC

- Took ownership from design to release of a **data modeling pipeline** that standardized models to production format
- Built an internal library that reduced Cross-Validation time by more than half using **sklearn and Dask**
- Integrated a **denoising auto-encoder** for preliminary feature selection on insurance premium data with **Pytorch**
- Built fault-tolerant and **load balanced infrastructure** using Docker Swarm for deploying different environments

NUANCE COMMUNCATIONS INC. | QA Automation Intern
May 2017 – Aug 2017 | Montreal, ON

- Reduced overhead of load testing by moving away from **JMeter** to a custom multi-threaded test framework
- Automated server configurations workflow for testing using scripts to accelerate deployment cycle
- Gained significant insight on test-driven development by implemented automated testing scripts with **JUnit and Unittest**

WEDO SERVICES | Software Developer Intern
Jul 2016 – Oct 2016 | Montreal, QC

- Integrated quick quote feature that calculates surface area by coloring on a map which improved UX
- Designed and built a **chatbot** using Facebook's Messenger API that quadrupled customer convergence
- Refactored a reactive interfaces with **MeteorJS's two-way data binding's protocol (DDP)** over websockets.

PROJECTS

MACHINE LEARNING BASEBALL (M.L.B.) | Predicting MLB games' outcome
Sep 2018 – Present

Aggregated data from multiples sources on game logs and team/player's statistics from 2016-2017 regular season. Performed feature selection using **random forest** on players' statistics to best estimate player's game won. Computed each past game as a vector with n features (aggregated home/away players/team stats) and trying to perform supervised classification on home win-loose using an **support vector classifier** (SVM with Linear Kernel)
Cross validation with 60-40 data split with **Gridsearch on hyperparameters**

CORNY VOLATILITY | Implied Volatility Analysis on Corn Futures
Apr 2018 – Present

Analyze the effect of the WASDE report published monthly by the USDA on the implied volatility dynamics for corn futures. Built a **real-time data pre-processing pipeline** using Interactive Brokers' API in C++ with **Apache Kafka** for low-latency message processing to compute implied volatility.

MAGIC PORTFOLIO | Deep RL Portfolio Optimizer
Aug 2017 – Oct 2017

Attempt to reproduce a paper by Jiang et al. on financial portfolio optimizer using a **deep reinforcement learning** agent using **Keras** and data from Yahoo Finance. Outputs a vector of weights to m instruments in portfolio for optimized performance.
CNN of 3 dense layers with **rectified linear activation** and **Gradient Descent** Optimizer (SGD) and a final layer with **softmax** .

EDUCATION

UNIVERSITY OF WATERLOO | B.Sc in Software Engineering
Minor in Computational Mathematics | Expected May 2022 | Waterloo, ON
UW Data Science Club • UW Poker Studies Club • Mathematical Finance Club