ALYSSA ZHANG

aspiring data scientist

alyssa.zhang@uwaterloo.ca | (416)-436-5885 | alyssaazhang | in/alyssaazhang/ | alyssaazhang.github.io/

SKILLS

Languages & Formats: Python, R, SQL, C, C++, HTML/CSS, JavaScript, JSON, CSV
Tools & Software: Oracle SQL Developer, Jupyter Notebook, Google Colaboratory, RStudio, Git & GitHub, Apache Hadoop, Keras & TensorFlow, MS Word/Excel/Access/SharePoint/Power BI, Libraries (Pandas, NumPy, Matplotlib, Scikit-Learn, SciPy), Linux

EXPERIENCE

DATA SCIENTIST

Intact Financial Corporation | May-Aug 2022

- Investigated proof of concept gradient boosting model to predict claims that were likely to incur larger expenses
- Communicated with 2 other teams and analyzed 3 data sources to find and evaluate metadata to be linked to existing features
- Supported department POC pitch by extracting model statistics and exploring model correlation, leading re-assessment of pitch through disproving business hypothesis
- Conducted data wrangling in Python and SQL to explore possible sources of new features and implemented 24 new features

MACHINE LEARNING SYSTEMS ENGINEER

Toyota Motor Manufacturing Corporation (TMMC) – Innovation Lab | Sep-Dec 2021

- Trained convolutional neural networks (CNN) with YOLOv3 for 2 live visual systems; used image classification and object detection to inspect vehicle parts in production and identify defects
- Introduced new workflow for training deep learning model with Mask-RCNN to accommodate for polygon labeling; reduced model training time by 50%
- Proposed ideas to General Managers to prevent live freight issues (unused vehicle parts that are accidentally returned to manufacturers)
- Assisted with media production and photography for 40+ engineers and the TMMC President during the annual Innovation Showcase

PROJECTS

TWITTER STATISTICS (GitHub)

- Performed **data mining** on Twitter for trending topics via **Python Tweepy** and the **Twitter API**
- Conducted **data munging** and used **natural language processing** (NLP) techniques to create a word cloud
- Implemented **object-oriented programming** to collect live tweets related to the COVID-19 vaccine; displayed their locations on an interactive map via **MapQuest Geocoding API**
- Retrieved tweets as **JSON** objects and stored in a **MongoDB** database to display user information and activity on an interactive choropleth map

FORECASTING WEATHER AND HOUSING PRICES (GitHub)

- Estimated past and predicted future temperatures using **time series forecasting** on NOAA's weather data
- Implemented **simple linear regression** two ways: **Python SciPy** and a **Scikit-learn machine learning estimator**
- Generated house price predictions with 8-variable **multiple linear regression** on Scikit-learn's California Housing data

MNIST DIGITS (Github)

- Trained and tested a model for predicting handwritten digits (from MNIST database) using **k-nearest neighbors algorithm**
- Measured and evaluated model accuracy with a **confusion matrix**, **classification report**, and **k-fold cross-validation**
- Performed **dimensionality reduction** and visualization with **Python Scikit-learn** and **Seaborn**
- Built a **convolutional neural network** (CNN) that executed probabilistic classification of digits using **Keras** and **TensorFlow**
- Visualized CNN with **TensorBoard** and evaluated incorrect predictions

COVID-19 DASHBOARD (Github)

- Visualized COVID-19 statistics around the world on an interactive **Power BI** dashboard

EDUCATION

UNIVERSITY OF WATERLOO

BCS, Honours Computer Science (Co-op) | 2020-2025

- Varsity Swim Team – 3x Athletic Financial Award recipient, Ontario University Athletics (OUA) medalist, volunteer swim instructor for non-profit Adventure 4 Change and anti-racism group The Alliance
- Relevant Courses: Data Structures, Computational Statistics & Data Analysis, Database Management, Applied Linear Models