Yuechen Zhang, Joe

Tel: +1 4374210264 | **Website**: https://joe-zyc.github.io/web-portfolio | **Email:** y2799zha@uwaterloo.ca

SUMMARY OF QUALIFICATIONS

- ✓ Candidate for Honors Computer Science from University of Waterloo, with solid knowledge of computer science, software engineering, data science and statistics.
- ✓ Skilled in Python, R, SQL, TypeScript, JavaScript, DBT, Tableau, Power BI and other programming languages and software. Practical working experience with top front-end and back-end frameworks.
- ✓ Rich internship experience in software engineering, data engineering, data analysis, data mining and machine learning algorithms. Have good data acquisition, capture and cleaning capabilities. Sensitive to data, strong logical thinking, able to effectively implement business requirements and complete model building.
- ✓ Excellent team player, strong sense of cooperation, good at inter-departmental communication and collaboration, strong self-motivation and sense of goal, high work efficiency and strong resistance to pressure.
- ✓ Quality written and communication skills. Quick Learner, effective time management, and quick adaptability to environments.

EDUCATION

University of Waterloo **Sept 2018-Present**

Candidate for Honors Computer Science Ontario, Canada

- ♦ 4B Data Science student with a 4.0 GPA; Won the Third Place in the class Kaggle competition.
- ♦ Core courses: Operating System, Data Structure, Algorithm, Statistical Inference, Linear Regression, Assembly Language, Data Analysis, Machine Learning, Time-Series Analysis, Classification, Database Management, Distributed System, DCT and JPEG Compression, OOP.

TECHNICAL CAPACITY

- ♦ Programming languages: Python, R, TypeScript, JavaScript, Java and C/ C++.
- ♦ Data engineering and analysis: Pandas, SQL, DBT, Spark, Hadoop, Tableau, MS Power BI, MS Excel, R.
- ♦ Machine learning and NLP models: TensorFlow, Scikit-Learn, NLTK, Spacy.
- ♦ Front-end frameworks: React, Angular.
- ♦ Back-end frameworks: Django, Flask, Express.
- ♦ Professional knowledge: Computer science, Statistical inference, Machine learning and Time-series Analysis.
- ♦ Familiar with classification, clustering and regression algorithms. Deep understanding of SVM, Random Forests, XGBoost, MLP and NLP algorithm.

PROFESSIONAL EXPERIENCES

Babylist **2022.05-2022.08**

Intern in Data Analysis and Data Engineering Oakland, USA

- ♦ Sorted out business data links and indicator frameworks, combined business development needs, deeply dug potential problems and possible directions from the perspective of data, and linked relevant departments to give reasonable practical suggestions.
- ♦ Carried out research on self-service BI tools, including Apache Superset, Metabase, ThoughtSpot, Mode, Google Looker and Sigma. Conducted differentiated comparative analysis of research results and output statistical analysis report to provide decision support for the team to select BI tools.
- ♦ Demonstrated the operation methods of two BI tools to help the department master the use method of the platform smoothly.
- ♦ Understood the business logic, application scenarios and underlying data of the demand side, built a reasonable indicator system and data reports; Used DBT, Snowflake and SQL to build dimensional models; Designed and implemented dimension tables and fact tables of products, warehouses, sales, etc.
- ♦ Based on business scenarios and requirements, used Mode, Thoughtspot and Sigma to build multi-dimensional data visualization dashboards.

KPMG LLP **2022.01-2022.04**

Data Science Intern Toronto, Canada

- ♦ Used Regex, NLP and other text recognition algorithms to extract, classify and statistics information from regulatory documents.
- ♦ Built a web platform to display the data after text recognition under various aggregation dimensions, and realized time-series visualization, quick keyword query, data export to CSV, user role management and other features.
- ♦ Rebuilt UI/UX front-end pages, developed new features and pages based on customer service requirements, built API for features, detected and fixed bugs, generated queries for data extraction.
- ♦ Conducted statistical analysis program design and model development, tested and validated prediction models. Completed part of the back-end refactoring and modularized the code according to OOP principles. Drew out a flow chart that clearly depict the execution logic of the algorithm.
- ♦ Used JavaScript, TypeScript, Angular, Express, NPM, Python, Node and other technologies to complete development and model building; Used Azure DevOps to carry out effective team collaboration.

High Hope Wisdom Investment Management Company **2021.04-2021.07**

Data Analyst Intern Nanjing, China

- ♦ Communicated data requirements with all businesses, designed and developed data visualization dashboards based on business understanding.
- ♦ Assisted in finding data outliers. Analyzed the causes of data outliers, and provided data support for decision-makers and other stakeholders.
- ♦ Managed financial data, updated fund net asset values on daily basis, monitored data accuracy and consistency.
- ♦ Maintained 3 databases with more than 100 tables to ensure data integrity. Used Python, SQL to clear more than 1 million rows of errors and redundant data.
- ♦ Wrote Python scripts to automatically generate financial reports for production team and operation team, automatically obtained periodic financial metrics.
- ♦ Assisted in the development and maintenance of data crawling procedures, and used Regex, NLP and other technologies to synchronize the net value for all mutual fund data recorded in the email contents to the database.

Region of Peel **2020.09-2020.12**

Business Intelligence Analyst Intern Brampton, Canada

- Built automation tools using MS Access, MS Flow and SharePoint APIs to increase team productivity.
- Cooperated with team members to complete data migration, and gradually migrated the data deployed locally to the cloud.
- Used MS Power App to develop electronic intelligent inspection questionnaire, which automatically generated HTML and PDF files to display the questionnaire results. Used MS Flow to generate PDF summaries and sent automatic emails.
- Used Power BI to set up a monitoring page for the statistics of questionnaire results.
- Trained nearly 40 managers on the electronic intelligent inspection questionnaire.
- Proposed and established an automated contract document management system to manage contract status and expiration dates.

Overbond **2020.01-2020.04**

Data Science Intern Toronto, Canada

- Established data visualization reports, assisted in deploying scripts on AWS servers, and developed front-end functions with React.
- Designed the error measurement scheme for the calculation program of predicting bond price to construct the accuracy index, created visual reports using Tableau and Python to show errors.
- Utilized Google Cloud API to update and sent weekly precision reports for further collaboration with the Data Science team for algorithm precision improvements.
- Used SQL, Pandas to retrieve data for predictive model development. Tested the prediction model on AWS cloud server, recorded the error.
- Cooperated with the front-end development team and used JS and React to add new pages and functions to the company's website.

PROJECT EXPERIENCES

Personal Website **2022.04**

- **Background:** Build a personal website using Angular. Host the website with GitHub Pages.
- **Technology involved:** Angular, CSS, HTML, TypeScript.
- **Development environment:** Node
- **Key content:**
  - Completed the user design and page development of the front end of the website; Deployed the website with GitHub Pages.
  - Developed the language translation feature between English and Chinese based on user's default browser setting, and completed the development of response sidebar, word cloud (using library) and other modules.

Course Evaluation Sentiment Analysis **2021.05**

- **Background:** Build a sentiment analysis model that predicts whether a given comment about a course has a negative/neutral/positive feeling.
- **Technology involved:** Python, TensorFlow, Scikit-learn, NLTK, Jieba, CNN, Naïve-bayes, Web Scrapping.
- **Development environment:** Anaconda
- **Key content:**
  - Used Python to develop web crawling scripts. Collected sampling data. used web scarp technology, download the course evaluation with comments and ratings from https://www.imooc.com.
  - Completed the data preprocessing, marked the comments, used N-gram algorithm to construct bag-of-word matrix for the sample comments.
  - Predicted the ratings by fitting a Naïve Bayes model and a CNN model to the bag-of-word matrix.
- **Achievements:** Achieved an accuracy of 0.675 on the test set.

Protein Location Prediction Project, Kaggle Competition **2021.04**

- **Background:** Use machine learning algorithms to classify the location of protein into 1 out of 14 categories giving data about protein classes, interactions, and essence.
- **Technology involved:** Python, TensorFlow, Scikit-Learn, Pandas, SVM, Neural Network, Decision Tree, XGBoost, Random Forest.
- **Development environment:** Anaconda
- **Key content:**
  - Cleaned the raw data, performed feature engineering including missing value imputation, one-hot encoding to categorical values and standardization.
  - Fitted different machine learning models to the cleaned data. Employed train-test split and cross validation for selecting the model parameters.
  - Trained SVMs, Random Forest, XGBoost and MLP models.
- **Achievements:** Achieved a best accuracy of 0.725 on the test set. Won the third place in the in-class Kaggle competition.

Waterpoly, Waterloo version of Monopoly game **2020.09**

- **Background:** Collaborate in groups, build a waterloo version of monopoly game with C++ that players can interact using std input/output.
- **Technology involved:** C++, CMake, GitHub, Git, OOP, Design Pattern.
- **Development environment:** Linux
- **Key content:**
  - Designed the class objects, chose design patterns used for the game.
  - Drew out the UML diagram of the class relationships.
  - Implemented the game in C++.
  - Collaborated with group members with Git and GitHub.
- **Achievements:** Scored 90/100 on the assignment.