JIANGYUE MAO

PHONE: 7343538253

EMAIL: maojy@umich.edu, GitHub Portfolio: https://github.com/alisamao09/Portfolio

SKILLS AND QUALIFICATIONS

- SQL, R, Python(PyTorch), C++, Tableau, Microsoft Office(Word, Excel, PPT etc.)
- 3 years experience in statistical analysis, machine learning, and data visualization, capable of forecasting, time series, experimentation, hypothesis testing, classification, clustering, anomaly detection or regression analysis, and data mining
- Leadership, communication, teamwork, innovative and active independent learner, time management, detail oriented

EDUCATION

Masters of Science in Data Science, University of Michigan, Ann Arbor **GPA: 4.0/4.0 Sept 2021 – Dec 2022**

- Course highlights: Machine Learning, Applied Data Mining, Programming for Scientists and Engineers (C++ on Linux system), Data Structures and Algorithms, Time Series, Deep Learning for Computer Vision

Honours Bachelor of Arts with High Distinction GPA: 3.62/4.00 2017 – 2021
Specialist in Architecture, Major in Statistics, Minor in Mathematics, University of Toronto
Dean's List Scholar 2020 & 2021

- Course highlights: Methods of Data Analysis I & II, Design and Analysis of Experiments, Data Visualization, Advanced Topics in the Technology of Architecture: Designing [with] Inherited Data, Surveys, Sampling and Observational Data

DATA SCIENCE EXPERIENCE

Deep Learning Assistant (Intern), School of Information, University of Michigan Sept 2021 – Jan 2022

- Create nodes on deep learning and related research papers on 1cademy (a collaborative learning platform)
- Present nodes in weekly team meetings to update research progress

Map Operation Assistant (Intern), MyH2O(NGO), Remote (Beijing) Jul – Aug 2021

- Developed web-based and phone-based interactive maps using Tableau to display the water condition in each village
- Presented the interface based on filtered survey results for each village
- Worked with around 10 team members to investigate in user interface design to present to policy makers and the public

Data Scientist (Intern), CPPEI, Beijing Jun – Jul 2021

- Collected and cleaned electricity consumption, electrical installed capacity, and GDP data by months, quarters, and years
- Built multiple regression and time series models to preliminarily forecast electricity consumption in China
- Visualized and analyzed the dataset using R and Tableau

Consultant, Data Analytics Consulting Virtual Internship, KPMG, Online May 2021

- Performed data analysis and data quality assessment based on customers' demographic information and the transaction data to suggest potential customers
- Visualized data trend, customer value, and model outcomes in Tableau
- Presented team planning goals, results, and supplied growth strategies and external dataset suggestions to the client company to maximize their profits

Undergraduate Researcher, Department of Physics, University of Toronto Oct 2020 – Feb 2021
Supervised by Prof. Carolyn Sealfon

- Trained NLP models such as GMM in Python to study students' open-ended questions for course materials
- Classified responses using Naïve Bayes and SGD classifier and applied K-Means, Affinity Propagation, Agglomerative, and Birch Clustering algorithms
- Applied PCA and t-SNE to visualize students' common questions
- Tested multiple training and test filters to filter out non-answers, with a success rate above 90%
- Improved the performance of machine learning classifiers drastically by around 90%, measured by the 2D visualized results

CAPSTONE PROJECTS 2020 – 2021

- Applied machine learning classifiers on traditional and modern tobacco use and the prediction of Cervical Cancer
  - Constructed machine learning models including Support Vector Classifier, Kernel Ridge regression, Logistic Regression, Lasso Regression, Random Forest Classifier, Gradient Boosting Classifier, and Multilayer Perceptron
  - Applied SMOTE method to handle the imbalanced dataset
  - Achieved high predictability with an AUC of approximately 90% and a 0.2 MSE
- Constructed a generalized linear mixed regression project using R: "How is diabetes readmission rate related to patients' pathologic conditions and medications?"
- Developed a 3D data visualization model using Grasshopper and its built-in Python to associate Covid-19 death cases in Toronto with social media statistics