Wen Zhang

Email: wzhang59@buffalo.edu | Mob: (716)-907-2682 | LinkedIn: Wen (Vivien) Zhang | GitHub: WenZhang-Vivien

EDUCATION

University at Buffalo, The State University of New York, Buffalo, NY, United States
- PhD in Computer Science and Engineering — February 2023 - January 2027
- Master of Science, Data Science, GPA: 3.927/4.000 — August 2021 - February 2023

Anhui Normal University, Wuhu, Anhui, China
- Bachelor of Science, Geographical Information System — August 2012 - July 2016

SKILLS

Proficiencies: Data science, Machine learning, Deep learning, Probability, Statistics, Optimization, Dynamic pricing
Languages: Python (Libraries: Pandas, PySpark, Scikit-learn, TensorFlow, Pytorch, NumPy, Plotly, Gurobipy), Java
Data Management & Analytics: Spark, MapReduce, Hadoop, PostgreSQL, MySQL, SQLite, Redis, Minitab, XML, JSON
Tools: Jupyter Notebook, Visual Studio Code, Eclipse, Google Colab, MATLAB, Atlassian JIRA

WORK EXPERIENCE

Research Assistant, University at Buffalo, Buffalo, United States — July 2022 - Present
- Analyzed trillions of traffic vehicle movement data, driver event data, and incident data, conducted in-depth analysis of traffic conditions in the Buffalo area. Employed data analysis models such as anomaly detection and time series analysis to identify traffic accidents and assess their impact on traffic
- Utilized Aimsun, a traffic simulation software, to design and execute simulations of real-life traffic accidents, simulating accident scenarios to evaluate their impacts on traffic. Analyzed accident simulation results, including the number of affected vehicles, level of traffic congestion, and changes in vehicle behavior

Intern, National Renewable Energy Laboratory, Denver, United States — May 2023 – August 2023
- Delved into the construction of various confusion matrices to compute travel downstream metrics and their uncertainty. The research serves as the foundation for a research paper detailing the findings and conclusions
- Utilized a labeled travel dataset consisting of 90,021 trips contributed by 270 users. Established three distinct confusion matrices using trip count, trip distance, and trip duration as the defining parameters
- Applied the confusion matrices to calculate values and uncertainties of energy consumption, carbon emission, mode counts, and mode shared by distance. Conducted comprehensive analysis of the results

Software Engineer, TravelFusion, Shanghai, China — March 2018 - August 2021
- Integrated 10+ Low Cost Carriers APIs and 2 Full Service Carriers APIs to realize aggregation, extraction and transmission of flight booking information including normal booking flow and post booking flow using Java
- Led and programmed ANA NDC project which passed IATA Level 4 certification, achieved the direct sales of air tickets between Google and All Nippon Airways on Google via “Book on Google with ANA”
- Solved 1200+ JIRAs including implementing new requirements and maintaining airline ticket distribution platform
- Analyzed air ticket datasets and logs, saved 30% in system caching by restructuring booking process
- Maintained close communication with airlines and customers, and provided fast technical support

Intern, Twenty First Century Aerospace Technology, Beijing, China — March 2015 - July 2015
- Programmed WebGIS applications and mobile phone applications by applying C# and JavaScript
- Processed and examined geographical information data and remote sensing image including data extraction, analysis, supervised and unsupervised classification and data visualization by ArcGIS and ENVI
- Designed databases for storing geographic data and carried out the edit processing in Oracle database

PROJECTS

Backtesting Trading Strategies Database (Link) — October 2022 - December 2022
- Processed a year's worth of data for 100 stocks pulled from an API, designed and created a database in PostgreSQL to store the data, and constructed 20 different backtesting strategies accordingly
- Built an Rshiny app and deployed on Heroku, helping users to run backtesting on custom strategies

Improving Super-Resolution WGAN for Image Restoration (Link) — March 2022 - May 2022
- Implemented Wasserstein GAN-Gradient Penalty into SRGAN to improve unstable training processes
- Deployed total variation loss and structural similarity index in the generator objective function to improve the results, obtained a Peak signal-to-noise ratio 25.37

Analyzing Earthquake Data Using Machine Learning and Deep learning — October 2021 - December 2021
- Performed exploratory data analysis on an earthquake data set of 4,000+ records with 39 attributes
- Applied Logistic regression, Decision Tree, Random Forest, Boosting and Neural Network to train and test data set to segregate a tsunami generating earthquakes from non-tsunami generating earthquakes
- Predicted the severity of the earthquake with ~90% accuracy utilizing Gaussian Process Regression and Support Vector Regression and reported the findings in a tableau dashboard

Predicting Health Insurance Medical Bills Costs Using Machine Learning — October 2021 - December 2021
- Cleaned a Medical Cost data set of 1300+ records with 7 attributes and normalized it to an SQLite database
- Produced figures and plots based on the database and inspected the relationship between the features
- Predicted the individual cost of insurance with ~80% accuracy by Multiple Linear Regression and Decision Tree