ZENING WANG

New York, NY | zw749@cornell.edu | (858) 344-3825 | GitHub |LinkedIn

EDUCATION

Cornell Tech (Cornell University) | *MS in Information Systems and Applied Information Science* Aug 2024 - May 2026
Jacobs Technion-Cornell Dual Master's degrees | The Related Companies Fellowship | Concentration in Urban Tech

University of California, San Diego | *BS in Mathematics - CS; Minor in Data Science* Sep 2020 - Jun 2024
Marshall College Honors Program | Provost Honors

SKILLS

Programming Skills: Python (Pandas, NumPy, Pytorch, Tensorflow, Sklearn, Matplotlib), Java, JavaScript, CSS, R, Matlab
Analytics Skills: Database (MySQL, Spark, MapReduce), A/B testing. Tableau, Power BI, MS Office, Matplotlib

EXPERIENCE

Break Through Tech AI | Course Support/ Data Scientist (Part-Time) | *New York, NY* Aug 2024 - Current

- Lead and support 3 student groups in completing ML projects for Allstate, Latitude AI, and Verizon. Coordinate with company Engineers tracking team progress, grading assignments, debugging, and providing technical guidance.
- Developed machine learning models to predict insurance claims data, aiming to improve Allstate's claims process. Provided insights on model selection, performance evaluation and optimize faster training times across large datasets. Implemented models like SGDRegressor, XGBoost, and Random Forest, achieving ~38% improvements in accuracy.
- Built an NLP model to classify software products into categories from Verizon product descriptions. Led the team through data processing and feature engineering, utilizing techniques like TF-IDF and n-gram analysis. Oversaw the development of multiple models, including XGBoost, multi-shot learning, and BERT, reaching up to 69% accuracy.

HP Tech Ventures | Corporate Venture Capital & Business Analytics Extern | *New York, NY* Dec 2024 - Current

- Acquired deep understanding of venture capital operations, including commercial acquisition process, evaluation metrics for startup success and HP Tech Ventures' investment goals and its importance to overall business strategy.
- Performed deal sourcing, analyzed 20,000+ pieces of startup data using Excel to support market mapping, investment thesis development, and deal evaluation. Conducted correlation and regression analyses to identify relationships between key variables and used statistical results like R^2 and p-value to interpret the significance of the analysis.
- Leveraging SQL, built and queried datasets to extract actionable insights on startup funding trends and performance metrics. Created visualizations to communicate key findings and risks and recommended on 5 high-potential startups.

University of Washington (STAR Lab) | Research Assistant on Re-ID Project | *Seattle, WA* Jun 2023 - Sep 2023

- Utilized FastReID codebase for Re-ID algorithm development and ported it to an Edge device to realize cross-camera vehicle tracking in collaboration with Norwegian Department of Transportation (DOT).
- Trained models on Market1501 (pedestrian) and AIC19 (vehicle) datasets in Python to realize re-identification from multiple angles, achieving 86% and 98% accuracy respectively. Evaluated classifier performance using ROC Curve.
- Implemented data annotation and wrote Python script for data splitting, loading and configuration file for new database.
- Ported the pre-trained model to work on Nvidia's hardware. Re-optimize the production model by adjusting Batch Size and Checkpoint Period of BagTricks baseline with ResNet backbone to maintain functionality and accuracy on the new device, successfully satisfying the Norwegian government's special restrictions on privacy regulations.

ITSSKY Technology Co., Ltd. | Data Science Intern | *Nanjing, China* July 2021 - Sep 2021

- Operated Sunflower application to connect to remote server to finish basic data processing, including data inquiry, data table replication, and export designated data in MySQL database.
- Utilized Kettle from Spoon to scan over 2,000,000 rows of data to complete large-scale data extraction. Discover and analyze patterns of overweight trucks in the city using Tableau, prompting a better notification system with other teams.
- Participated to investigate how to improve performance by using data structures like B+ tree for index in SQL.

PROJECTS

Analysis on a League of Legends Dataset (Python, Pandas, Sklearn, NumPy, SciPy, Matplotlib) Feb 2024 - Mar 2024

- Within group of 2, quantitatively assess, visualize and compare the carry power of different roles in League of Legends using entire professional match statistics for the throughout 2023.
- Conducted data cleaning and exploratory data analysis on the large dataset to uncover patterns in player behavior and applied statistical methods, including hypothesis testing, to identify significant trends across various roles.
- Introduced a new metric and implemented ML techniques, such as regression and random forest models, following one-hot encoding and hyperparameter tuning. Successfully reduced RMSE by 20% compared to baseline model.

Processing of An Email Spam Dataset (R, ISLR, MASS) Feb 2023 - Mar 2023

- Within group of 3, used R to extract, clean and standardize the dataset of 4601 email messages.
- Fitted a Logistic Regression model. Applied Linear and Quadratic Discriminant Analysis methods to the data.
- Performed Linear and Nonlinear Support Vector Machine classifiers and deployed a Decision Tree model. Concluded all methods and evaluated their performances. Applied Decision Tree model and achieved over 91% accuracy.