Dan Wu

+1-425-435-7683| daniwu0912@gmail.com | LinkedIn | Los Angeles

EDUCATION

University of California - Los Angeles **Los Angeles, CA**
Master of Applied Statistics and Data Science (currently pursuing) *2023.9-present*
Expected graduation: June 2025
Bachelor of Science in Statistics and Data Science and Minor in Mathematics *2021.7-2023.6*
Dean's Honors List

Bellevue College **Bellevue, WA**
Associate Degree in Arts and Science *2019.8-2021.6*

PROFESSIONAL EXPERIENCE

PwC **Shanghai, China**
Process Quality Assurance Intern *2021.10-2021.12*

- Utilized Alteryx for data collection and analysis, employed Boosted Models and Principal Component Analysis, enhanced prediction accuracy, reduced dimensionality, which led further refinements to the analytical process.
- Implemented Bayesian models for regression analysis to evaluate 16-quality control metrics, achieving an accuracy of 94%, which optimized the quality assurance process and contributed to a 15% improvement in the accuracy of quality forecasts.
- Engaged in the evaluation process with a third-party entity, Measures, to analyze the potential risks of the project and streamlined the process of testing Agile methodologies.

Shangli Information Technology (Shanghai) Co., Ltd. **Remotely**
Data Intern *2021.6-2021.9*

- Compiled and organized over 300 annual reports from executive leaders, supplemented by public recruitment data from platforms such as LinkedIn. Successfully extracted and cleansed salary data across various company positions and ranks.
- Conducted secondary research on the client's industry and utilized Salesforce to analyze employee registration data to provide insights that allowed the company to set more competitive salary standards based on industry benchmarks.

EXPERIENCE & INVOLVEMENT

Python Project: Impact of Missed Minor League Time on MLB Performance
Group leader *2023.3-2023.6*

- **Data Extraction:** Leveraged MLB's RESTful API to scrape the data of game logs spanning a decade for 1,400 players, amassing over 600k rows of raw data.
- **Data Cleaning & Transformation:** Employed JSON parsing and Regular Expressions to structure data and devised a standardized WAR (Wins Above Replacement) metric to assess variations in games played and innings pitched.
- **Statistical Analysis**: Implemented regression models and tested with statistical tests, such as the t-test, finding a low correlation between the duration of missed Minor League time due to injuries and subsequent MLB performance.

Machine Learning Project Using R and Python: Heart Disease Predictive Modeling
Group leader *2022.6-2022.8*

- **Feature Engineering**: Applied one-hot encoding to the categorical variables and scaled numeric variables using min-max scaler to address data consistency.
- **Data Visualization:** Performed comprehensive Exploratory Data Analysis using Python's Seaborn package to visualize variable correlations and distributions, via heat maps, histograms, scatter plots, and box plots.
- **Machine Learning Modeling**: Conducted feature selection and built multiple machine learning models including Logistic Regression, Support Vector Machine ("SVM"), K-Nearest Neighbors, and XGBoost, with emphasis on SVM.
- **Model Evaluation**: Utilized 10-fold Cross-Validation to validate model performances and selected SVM, both Radial Basis Function and Degree-2 Poly kernel, which achieved the highest mean F1 score of ~0.85 as the final model for prediction.

SKILLS & OTHERS

Skills: R / Python (NumPy, Pandas, SciKit-Learn, PyTorch, Matplotlib, TensorFlow) / SQL / Machine Learning / C++ /Alteryx/ Tableau / MATLAB / Java / Excel / Word / PowerPoint / GitHub repos
Languages: English, Mandarin Chinese