Yang Han

+(1)734 882 7937 | marshan000824@gmail.com
linkedin.com/in/yang-han-50b8b823b

EDUCATION

University of Michigan   Sep 2023 - May 2025
Master of Applied Statistics GPA: 3.7   Ann Arbor, US

Relevant Courses: Statistical Modelling: Regression, Statistical Modelling: Multivariate, Data Science & Analytics Using Python, Computational Methods & Tools in Statistics(R, SQL, etc.), Bayesian Modeling & Computation

University College London   Sep 2019 - Jul 2022
BSc Statistics, Economics and Finance GPA: 3.7   London, UK

Relevant Courses: Probability & Inference, Statistical Inference, Linear Models & The Analysis of Variance, Forecasting using Time Series Analysis, Applied Statistics: Stochastic Process

PROFESSIONAL EXPERIENCE

Ford Motor Company   May 2024 - Jul 2024
Intern: Data Scientist   Dearborn, US

- Utilized **GCP** to write and execute data extraction queries, enhancing data accessibility for analysis.
- Analyzed vehicle battery drain issues using machine learning algorithms in **Python**, leading to a projected 20% reduction in battery warranty expenses and an improvement in customer satisfaction.
- Developed a **Python** script utilizing clustering algorithms such as **K-Means, DBSCAN,** and deep learning **Autoencoders** to automate the analysis process, increasing efficiency and scalability of the project.
- Attended the Hackathon with the team, aiming to set up a more efficient method for the battery triage project. Presented the presentation individually and successfully helped the team achieving the aim.

AD Services Co.   Jan 2023 - Jun 2023
Data Analyzing Consultant   Birmingham, UK

- Utilized **Tableau** to create dynamic dashboards for social media metrics, enabling data-driven decision-making and significantly increasing company visibility, resulting in more potential customer engagement for the team.
- Conducted an analysis of client behavior by developing reports based on limited social media metrics from up to 30 student organizations, successfully developed cooperation relations, achieving a 50% increase in social media following.
- Generated daily dashboards to monitor the rental market availability and consumer preference metrics, helping the team conduct consumer analytics, leading to a 40% increase in GMV in the leasing business during the off-season.

Beijing Xiaomi Technology Co.   May 2021 - Sep 2021
Intern: Data Analyst   Beijing, China

- Automated manual reporting using **Excel** index & match functions, resulting in a 50% reduction in weekly log writing time.
- Performed **ETL** operation, extracted metrics related to user growth, carried out the cleansing process, and loaded the data to the cluster by writing **SQL** queries and using **Python**, helping the team with the pre-analysis preparations.
- Performed user growth funnel analysis utilizing **Logistic Regression** in **Python** to pinpoint high-conversion touchpoints, delivering actionable data-driven recommendations to the team, improving the new user conversion rate by 15%.
- Detected abnormal click PV fluctuations at a particular node, found a scam by a click farm after investigating the data, and reported it promptly, preventing the team from losing a budget of 30,000 CNY.
- Designed and implemented **A/B testing** to optimize interface interaction, resulting in a 28% reduction in user churn to the second-level page and an 80% increase in banner exposure PV.

PROJECT EXPERIENCE

Trends in the Prevalence of Major Chronic Diseases Across the U.S. Over Time | Python   Apr 2024 - May 2024
Machine Learning Engineering   Ann Arbor, US

- Conducted data preprocessing, including data cleaning and exploratory data analysis, using **Pandas** and **Numpy** to ensure data quality and uncover key insights from the data.
- Performed feature engineering tasks such as feature interaction generation, normalization, and label encoding using **Scikit-learn** preprocessing to enhance model performance and data representation.
- Applied classification techniques such as **Logistic Regression**, **Support Vector Machines**, and **XGBoosting** to develop predictive models for common chronic diseases in the U.S.
- Performed hyperparameter tuning using grid search and applied K-fold cross-validation to optimize model performance, resulting in a 35% improvement in predicting accuracy.
- Generated interactive visualizations using **Plotly** and **Matplotlib** and compiled a comprehensive report. Delivered a successful presentation to non-technical audiences, receiving positive feedback for clarity and engagement.

SKILLS

- **Programming:** Python(Numpy, Pandas, Matplotlib, Sci-kit learn, etc.), Spark, TensorFlow SQL, R(dplyr, etc.), SAS
- **Tools:** Microsoft Office, Tableau, AWS, GCP, ETL Process, A/B Testing, Statisitcal Modelling, ML Algorithms, Gen AI Prompting