

# SHIYU SONG

daxuyy@163.com | +86 18562738150 | Qingdao, China

## EDUCATION

---

### Australian National University, Canberra

02/2018 – 05/2020

*M.S. in Statistics | GPA: 6.1/7.0*

- Excellence Certification for AIG International Internship Program
- Courses: Applied Statistics, Graphical Data Analysis (R), Generalised Linear Models (R), Regression Modeling (R), Applied Investments, Introduction to Machine Learning (Python)

### University of Illinois at Urbana-Champaign, Champaign

08/2013 – 05/2017

*B.S. in Actuarial Science | GPA: 3.4/4.0*

- Dean's List Academic Achievement Award Recipient (Semester GPA ranks in the upper 20%)
- Courses: Actuarial Statistics I & II, Calculus I & II & III, Economic Statistics, Methods of Applied Statistics, Introduction to Computer Science (Java), Linear Algebra & Financial Applications (Python)

## WORK EXPERIENCE (FULL-TIME)

---

### Xiaohongshu, Qingdao

06/2023 – Present

*Data Science Influencer*

- Published over 200 data science video posts, accumulating 175,000+ likes and saves.
- Amassed more than 20,000 followers within six months, earning recognition as one of the annual top data science influencers by Xiaohongshu Official.
- Developed a series of personalized professional development courses and mentored more than 120 students in data science career development and job-seeking skills.
- Conducted bi-monthly live-streamed data science seminars, engaging audiences in statistics, machine learning, and programming.

### SeeWorld, Shanghai

06/2023 – Present

*Founder*

- Leading a team to build several learner portraits and develop data science curriculum and training, including data camp, live classes, interview training, job-seeking skills, and one-on-one tutoring.
- Managing social media platforms which aim to build data science channels and publish data science content posts, impacting 20,000+ learners.
- Built a dashboard to analyze learners' learning behavior and performance, predicting the weekly topic trend and providing more personalized training for learners.
- Analyzed the underlying logic of popular posts and built the matrix of the platforms to improve user acquisition and retention rate from 10% to 16%.

### Yuanfudao, Beijing

02/2021 – 06/2023

*Senior Data Scientist, Tutor Analytics Team*

- Designed and developed the business metric system through Tableau, including 200+ indicators for performance tracking; generated 70+ analytical reports to provide strategic insights and solutions across 15 divisions.
- Created an automated attribution analysis tool based on the LMDI method and Delta method to track and elucidate core metric dynamics, enhancing operational efficiency by 30% and conserving two days of manual labor per analysis.
- Analyzed the correlation between tutor vocal attributes and conversion rates by GLM and residual diagnostics, providing insights for developing the tutor platform's 10+ new features.

### Project: End-to-End 'Pentagon' User Profiling System

- Initiated and delivered a first-of-its-kind user profiling system by developing a 'Pentagon' scoring model for scoring and predicting user behavior metrics, contributing to the company's first positive Unit Economics in four years.
- Conducted A/B testing to improve user segmentation tactics, increasing course completion rates by 10%, payment conversion rates by 5%, and customer lifetime value by 2,000 RMB.
- Provided system insights in inter-departmental forums to lead a broad-based adoption and get a Net Promoter Score (NPS) of 4.97 out of 5 post-implementation.

#### Project: User Stratification Probability Prediction Model

- Developed and trained logistic regression and XGBoost models to implement an automated pipeline using PySpark for daily key metrics updates and maintain the AUC score between 0.85 and 0.9.
- Managed the entire cycle of data analytics, coordinated A/B testing, and provided results and analytical insights to cross-functional teams.

#### Project: Target Setting on Online Tutors Conversion Rate

- Delivered granular conversion rate predictions and assisted tutors in anomaly detection and target setting by using the Prophet algorithm to build a hierarchical time series forecasting model, achieving an RMSE of less than 0.5% and contributing to a 20% reduction in labor costs.
- Generated a report to predict seasonal patterns, developing automated prediction workflows to reduce tutor labor costs by 15%.

#### **Pingan Health Insurance, Shanghai**

08/2020 – 01/2021

*Data Scientist, Actuarial Product Team*

#### Project: Recommendation Model for Medical Insurance Add-On Packages

- Developed a GBDT/XGBoost predictive model to forecast user propensity for renewing and opting into additional overseas insurance packages, validating an AUC above 90% and increasing the renewal rate by 5%.
- Developed a data processing and feature engineering pipeline, extracting 1,000+ variables using HiveSQL.
- Promoted the machine learning models to enhance user profiling and model accuracy, increasing the recall rate by up to 90%.
- Employed Random Forest algorithms on large-scale datasets, summarizing 100+ significant predictors for the strategic planning of medical insurance products.
- Facilitated personalized customer strategies by analyzing over 30 features and executing customer segmentation by K-Means and DBSCAN clustering methods.

### **ADDITIONAL INDUSTRY WORK EXPERIENCE**

#### **Rice Warner, Sydney**

12/2019 – 02/2020

*Actuarial Intern*

- Polished the 2019 datacube dashboard in Tableau and evaluated 2019 individual investment prediction model results.
- Worked on the research of the regulatory report 639 – AustralianSuper and drafted key elements for customers.

#### **AIG, Shanghai**

11/2018 – 02/2019

*Actuarial Intern*

- Created an automated tool by VBA to convert BondEdge's 500+ portfolio data into GGY-Axis data, using Prophet to predict liability-side cash flow.
- Conducted research to analyze the rating methodologies of the Big 4 (S&P, Fitch, AM Best, Moody's) and wrote a report to provide insights and key findings to the team.

### **SKILLS**

**Programming Languages:** Python, R, SQL, Java, PHP, Hive, PySpark

**Software:** RStudio, Tableau, Anaconda, MySQL, MS Offices (Word, Excel & PowerPoint)

**Data Science Model:** Decision Tree, Random Forest, Linear Regression, Generalized

Linear Regression, GLMM, PCA, Cluster Analysis, Time Series Analysis, GBDT, XGBoost, SVM, GMM-LSTM

**Spoken Languages:** Mandarin, English

### **CERTIFICATION**

Data Scientist with Python, DataCamp, 12/2020

### **WEBSITE**

**LinkedIn:** <https://www.linkedin.com/in/shiyu-song-aa7609a6/>

**Portfolio:** <https://daxxxxxx.github.io>