

BRIAN CHI YAN LI

Detail-oriented professional skilled in statistical programming and machine learning, proficient in SQL, R, and Python. Strong problem-solving abilities to support analytics, process automation, and business enhancements.



EDUCATION

- | | | |
|-------------------------|--|----------------------|
| 05/2019

08/2017 | • Master of Science in Statistics (GPA: 3.9)
North Carolina State University | 📍 Raleigh, NC |
| 05/2014

08/2012 | • Master of Science in Engineering Management CO-OP (GPA: 3.5)
Purdue University | 📍 West Lafayette, IN |
| 05/2012

08/2007 | • Bachelor of Science in Chemical Engineering (GPA: 3.6)
Purdue University

• Minor in Mathematics and Economics | 📍 West Lafayette, IN |

WORKING EXPERIENCES

- | | | |
|-------------------------|---|------------------------|
| Current

01/2023 | • Data Scientist
Levi's

• Apply XGBoost, lightGBM, CATboost with lag features to build shipping demand forecasting for Europe and sizing forecasting for USA/CA using Python
• Monitor prediction accuracy by WMAPE and perform hyperparameter tuning/retraining with Optuna
• Maintain a CI/CD sellout data pipeline and build test cases to alarm data changes
• Migrate Europe forecast from AWS SageMaker to GCP Vertex Instances and refactor code to reduce ~15% runtime
• Research ways of model splitting and apply prophet time series on a subset of products to reduce WMAPE by ~12%
• Build an analytical chatbot using LLM to enable business planners self inquiry without writing SQL during a google x Levi's GenAI Hackathon
• Adopt Agile philosophy and share findings through confluence documentation | 📍 Orlando, FL (Remote) |
| 01/2023

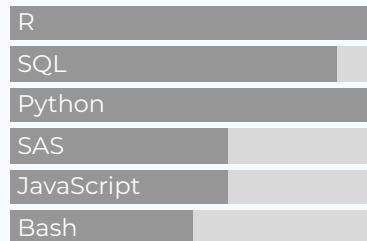
08/2021 | • Principal Data Scientist
Verizon

• Apply statistical models including Logistic Regression, Generalized Linear Models with Regularization (Lasso/ Ridge), CART, Random Forests, Extreme Gradient Boosted Tree to develop production ready score cuts in consumer credit and fraud red flag policy
• Wireless involuntary churn forecast using Python random forest model and Cox proportional hazard model
• Build a Qlik dashboard tool with REST API to automatically update churn forecast trending vs budget to client, eliminating manual update work
• 5G Home credit modeling and optimization to determine deposit schemes that maximize cash flow given variable costs and write off probability
• Translate R code of Loan Loss Forecasting to Python and migrate from Unix to GCP
• Cross train other team members on scorecut modeling/generation with new EFX Fraud Superscore and Neustar score
• Act as subject matter expert on probability scoring for accuracy monitoring, and Oxford Economics macroeconomic correlation analysis for SOX compliance | 📍 Lake Mary, FL |

CONTACT

- ✉ 8787 Lookout Pointe Dr., Windermere, FL 34786
✉ ziyong68@gmail.com
📞 765-413-3280
🔗 github.com/ziyong68

PROGRAMMING SKILLS



OTHER SKILLS

- Tools: Excel, PowerPivot, MATLAB, Tableau, Docker, GCP Vertex, AWS SageMaker, Airflow
Reporting Systems: Qlik, SAP Business objects, Eclipse BIRT, Jira, Confluence
Languages: English, Mandarin, Cantonese

CERTIFICATIONS

- Lean and Six Sigma Green Belt Certificate, Purdue University
Managing Big Data with MySQL, Duke University
Python Programming Certification, University of Michigan
FCRA Data User Certification program through CDIA

08/2021
|
01/2019

● Data Scientist

Verizon

📍 Lake Mary, FL

- Implement Multi Adaptive Regression Spline (MARS) model to forecast performance on variable term loans in order to predict exposure at default for 6-, 30- and 36-month device payment plan loans
- Optimize FraudIQ score threshold of red flag policies to stop identity frauds from entering credit check and prevent losses (estimated with 700K to 1 million per month)
- Transform Ignite credit bureau data assets using Impala SQL in Hadoop clusters and perform quality control & data deduplication
- Initiate the use of R markdown with knitr during CECL auditing to automatically generate dynamic documents that integrates model statistics, interpretation and validation plots with inline code
- Develop a new scorecut automation to generate post NITP v2 score launch adapting both volume neutral and risk neutral strategy at different time
- Create new R functions for team library. Examples include size compression of big GLM model RDS objects for more efficient storage, as well as empirical probability density/ cumulative distribution functions cast on grouped data

01/2019
|
08/2014

● Project Planner

Walt Disney Parks & Resorts

📍 Lake Buena Vista, FL

- Project lead for strategic facilities planning of Disney Springs and ESPN Wide World of Sports (\$15M annual)
- Develop BIRT user reports using SQL to query data and create custom functionalities with JavaScript under Eclipse IDE
- Perform data cleansing and migration from a project based (Maximo) to an equipment-based database application (Tririga)
- Developed an interactive PowerBI Dashboard to summarize water quality and pollutant distribution of Reedy Creek rivers
- Initiated projects to replace sport field light fixtures, with a total savings of 44% of the original energy consumption

LAB RESEARCH & PROJECTS

Current

● R Shiny App for Forest Fire analysis

Github

📍 <https://brianli.shinyapps.io/Forest-Fire-Investigation/>

05/2012
|
01/2012

● Study of Biomass Torrefaction

Purdue University

📍 West Lafayette, IN

- Applied linear regression to determine the kinetic order of Fatty Acid Methyl Ester (FAME) production and optimize yield

05/2012
|
01/2012

● AIChE Chem-E-Car Design

Purdue University

📍 West Lafayette, IN

- Drove a Lego built pneumatic engine with CO₂ released in an acid-base reaction under controllable conditions. The team was awarded a second place (2/10) recognition in the regional competition held in the University of Akron