Brian Li

Windermere, FL | 765.413.3280 | ziyong68@gmail.com | github.com/ziyong68

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, Certification & Training

M.S. Statistics, North Carolina State University
GPA: 3.9 out of 4.0
M.S. Engineering Management, Purdue University
GPA: 3.5 out of 4.0
B.S. Chemical Engineering, Purdue University
Minor: Mathematics and Economics
GPA: 3.6 out of 4.0
Certifications: Managing Big Data with MySQL, Duke University | Python Programming Certification, University of Michigan | Lean and Six Sigma Green Belt Certificate, Purdue University | FCRA Data User Certification program through CDIA
Skillsets: R | Python | Machine Learning | SAS | SQL | Teradata | Hadoop | Spark | Docker | Airflow | GCP Vertex | AWS SageMaker | SAP Business Objects | SharePoint | Tableau | Power BI | Qlik | Jira | Confluence
Languages: English, Mandarin Chinese, Cantonese

Professional Experience

Data Scientist – Levi's, Orlando (remote) ***(Jan 2023 - Current)***
Provide data science consultation to the demand planning forecast team and recommend future product catalogue allocation.

- Apply XGBoost, LightGBM, CATBoost with lag features to build shipping demand forecasting for Europe and sizing forecasting for USA/CA using Python, which facilitate the decisions on the apparel styles and quantity to produce
- Monitor prediction accuracy by WMAPE and perform hyperparameter tuning/retraining with Optuna to upkeep models
- Maintain a CI/CD sellout data pipeline and build test cases to alarm data shift, ensuring the training data is healthy
- Migrate Europe forecast from AWS SageMaker to GCP Vertex Instances and refactor code to reduce ~15% program 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, allowing them to quickly extract the business insights
- Adopt Agile philosophy and share findings through confluence documentation to reduce other data professional's research time (e.g. remotely SSH into GCP Vertex instance while obeying company security protocol)

Principal Data Scientist - Verizon, Lake Mary ***(Aug 2021 – Jan 2023)***
Lead data science projects within the credit risk modeling and wireless subscriber involuntary churn forecasting space.

- Applied statistical models including Logistic Regression, Generalized Linear Models with Regularization (Lasso/ Ridge), Random Forests, XGBoost to develop production ready score cuts that assign optimized credit offer to consumers
- Performed wireless involuntary churn forecast using Python random forest model and Cox proportional hazard model
- Built a Qlik dashboard tool with REST API to automatically update churn forecast trending vs budget to client, eliminating manual update and trigger work
- Conducted 5G Home credit modeling and optimization to determine deposit schemes that maximize expected cash flow given write off probability, revenue, and costs.
- Translated R code of Loan Loss Forecasting to Python and migrate from Unix to GCP, which facilitates cross team collaboration and process integration
- Cross trained other team members on scorecut modeling/generation with new EFX Fraud Superscore and Neustar score
- Acted as a subject matter expert on probability scoring for accuracy monitoring, and Oxford Economics macroeconomic correlation analysis for SOX compliance

Data Scientist - Verizon, Lake Mary ***(Jan 2019 – Aug 2021)***
Provide data science consultation to the credit risk modeling space and advocate best modeling and programming practices in Business Analytics Centre of Excellence.

- Implemented Multi Adaptive Regression Spline (MARS) model to forecast performance on variable term loans in order to predict exposure at default for device monthly payment plan loans, which is a critical input to expected loan loss
- Optimized FraudIQ score threshold of red flag policies to stop identity frauds from entering credit check and prevent losses (estimated fraud saving of 700K to 1 million per month)
- Transformed Ignite credit bureau data assets using Impala SQL in Hadoop clusters, and perform quality control & data deduplication to ensure a clean set of custom attributes are ready to use by other data scientists.
- Initiated 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, reducing time of editing documentation

- Developed a new scorecut automation to generate post NITP v2 score launch adapting both volume neutral and risk neutral strategy, which balance credit quality and gross adds to align with company strategy at different times
- Created new R functions for team library. Examples include size compression of big GLM model RDS objects for more efficient storage, as well as aggregated empirical probability density/ cumulative distribution functions

Planning Analyst - Walt Disney Company, Orlando ***(Aug 2014 – Jan 2019)***
An analytical role that combines data analysis and strategy creation to support the Integrated Facility Planning (IFP) process.
- Project lead for strategic facilities planning of Disney Springs and ESPN Wide World of Sports ($15M annual expenditure)
- Created lifecycle strategies to implement maintenance activities to extend asset service life under budget constraints
- Developed BIRT user reports using SQL to query data and create custom functionalities with JavaScript under Eclipse IDE
- Performed data cleansing and migration from a project based (Maximo) to an equipment-based database application (Tririga), which enable accurate tracking of asset conditions and lifecycle activities
- 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

Professional Engineering Intern - Walt Disney Company, Orlando ***(May 2014 – Aug 2014)***
An engineering internship program within of facility asset management organization.
- Analyzed conditions of various property assets (e.g. roofing, light poles, HVAC systems) and predict the timing of next major maintenance or upgrade
- Collaborated with an architectural consulting firm to transform Disney roof survey results into quantifiable data, which were utilized to plot roof degradation curves for different materials
- Identified approximately 920 projectors/screens across locations and created an inventory with comprehensive specifications and warranty documentation

Research and Projects

R Shiny App for Forest Fire analysis, https://brianli.shinyapps.io/Forest-Fire-Investigation/
Built a web dashboard using R to enable users to perform exploratory data analysis and model fitting with dynamic options. R code is also available in my GitHub Repository, https://github.com/ziyong68/590-Shiny

Study of Biomass Torrefaction, Purdue University *(Spring 2012)*
Applied linear regression to determine the kinetic order of Fatty Acid Methyl Ester (FAME) production and optimize yield

AICHE Chem-E-Car Design, Purdue University *(Spring 2012)*
Drove a Lego built pneumatic engine with CO_2 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