Xinning (Derrick) Chu

Data Scientist / ANALYST

2145 NE 164th St, N Miami Beach, FL, 33162 (Willing to Relocate)

+1 (202) 243-8499 | xinningchu@gmail.com | derrick23c | xinning-derrick-chu

Skills

Programming	R (**tidyverse, ranger, xgboost, Shiny**), Python (**Numpy, Pandas, Scikit-learn, Tensorflow, Keras**), SAS, SQL (MySQL, Postgre), Excel, Tableau, Hadoop, Hive, Spark, Git
Certificate	AWS Certified Cloud Practitioner; Data Analyst with R Track by DataCamp; Data Scientist Track with R by DataCamp; Statistician Track with R by DataCamp; Machine Learning Scientist Track with Python by DataCamp
Language	Mandarin, Cantonese
Other	Amazon Web Services (EC2, S3, SageMaker and etc.), Microsoft Azure

Professional Experience

71 Pounds, Inc.
Davie, FL

Data Analytics Specialist — *Jun. 2020 - Present*

- Developed / deployed predictive models using advanced statistical and ML techniques (**ensembled tree-based models / neural network / SVM / Baysian**) with **R** and **Python**, evaluated model performances, and tuned the hyperparameters, providing actionable insights which increased active customers by 30%.
- Worked with internal and external stakeholders to understand the underlying business needs, translated ambiguous problems into specific quantitative questions, and delivered data-driven solutions by storytelling or visualization.
- Applied time series methods, correlation analysis and linear models in R to detect **patterns, anomalies, relationships and trends** on a regular basis by manipulating data from varying sources, increasing revenues by 25%.
- Used **k-means / hierarchical clustering** models to perform **customer segmentation**, and provided data-driven insights to create customer-focused product positioning and marketing mix, increasing the profit of one product by 15%.

TAL Education Group
Beijing, China

Data Analyst — *Mar. 2015 - Nov. 2016*

- Worked with team members to efficiently collect, clean, and prepare **high-volume and high-dimensional** data for analysis.
- Created dashboards with Tableau, tracked key metrics by them, and generated ad-hoc reports by writing **SQL** queries.
- Implemented **A/B testing** to make product / service improvements and worked closely with the product team on improving the experiment design pipeline, increasing conversion rate by 20% and click-through rate by 15%.
- Built **Generalized Additive Models** with R to price company's products for the China market, increasing the profits by 27%.

Project Experience

Customer Benchmarking and Shipping Method Assistant
Davie, FL

Team Leader — *May 2020 - Jul. 2020*

- Implemented cluster analysis (**k-means / hierarchical**) to divide customers into groups, and benchmarked their shipping costs and discounts against a peer group of similar companies with data visualization tools, helping customers saving cost by 10%.
- Applied ensembled tree-based models, effectively predicting the discounts and helping top customers save cost by 22%.
- Built neural network/naive bayes/XG Boost models, selected the best one and predicted lower cost shipping methods that can meet parcel delivery requirements with high probability.

Lending Club Loan Status Prediction
Fort Lauderdale, FL

Team Leader — *Mar. 2020 - Apr. 2020*

- Implemented EDA to better understand the dataset, cleaned the data (**variable reduction, missing value imputation**), and performed feature engineering (**one-hot encoding, standardization**) to prepare data for training predictive ML models.
- Trained models on the training dataset using **logistic regression / random forest / gradient boosting / SVM** to predict the loan status (Default or Fully Paid) of Lending Club users, and tuned the hyperparameters, acheiving an accuracy of 96%.
- Verified the models on the testing set, evaluated them by confusion matrix, ROC curve, AIC and **cross-validation**, and selected the best to make predictions.

Beijing PM2.5 Level
Washington, DC

Statistical Consultant / Team Leader — *Dec. 2018 - Mar. 2019*

- Performed **exploratory data analysis**, mainly summarizing the characteristics with data visualization tools (**ggplot2, plotly**).
- Fit a **seasonal ARIMA** model of Beijing pm2.5 values, made diagnostics on it and decomposed it into three factors to figure out the trend of this time series with the results of stationary tests.
- Built **spatial-temporal** models on the pm2.5 values of 10 weather stations in Beijing from 2012 to 2015 with R (spBayes), effectively predicting the pm2.5 values for each location and providing our client insights about how the pm2.5 level would change in the future.
- Applied the **BP neural network** method to predict the monthly average pm2.5 values in Beijing, found that coal is the main cause of air pollution and came up with a data-driven solution for our client to improve the air quality of Beijing.

Education

The George Washington University
Washington, DC

M.S. in Statistics GPA: 3.70 /4.00 — *Aug. 2017 - May 2019*

- Relevant Courses: Mathematical Statistics, Applied Linear Models, Advanced Time Series, Machine Learning, Data Analysis, Statistical Consulting, Sample Survey, Applied Multivariate

Beijing Institute of Technology
Beijing, China

B.S. in Statistics — *Aug. 2011 - June 2015*

- Relevant Courses: Probability Theory, SAS Programming, Computational Statistics, Experiment Design, Monte Carlo Experiments, Reliability Statistics, Applied Regression Analysis, Statistical Analysis Using SPSS, Data Structure