Xiao (Seanny) Song

(240)564-1321 songxiao@umd.edu songxxiao xsong.ltd seannysong

EDUCATION

Robert H. Smith School of Business, University of Maryland, College Park, MD — Aug 2021 ~ Dec 2022
Master of Science in Business Analytics (STEM), Terrapin Scholarship

East China Normal University, Shanghai — Sep 2016 ~ Jun 2020
Bachelor of Sociology, Shanghai Scholarship

WORK EXPERIENCE

FancyDigital, Shanghai — Sep 2020 ~ Aug 2021
Data Analyst(Full-Time)

- Selected metrics and collaborated with a team to built multi-dimensional public BI dashboards; Visualized and evaluated health of a SaaS system resulting in a 5% growth of customers
- Mentored new co-workers to perform user churn analysis based on users databases and provided user service suggestions to operators resulting in a decrease of user churn rate to 10%
- Cooperated with product managers to optimize product design and user experience using error logs data leading to a reduction of bug rate by 7%; Upgraded and tested data product in a SaaS system during developing schedule
- Designed ETL shell scripts to aggregate original detailed data into intermediate tables to optimize query performance by 85% and successfully automate daily reports; Reduced human-power input and time cost

PROJECT EXPERIENCE

Click-Through Rate Prediction — Jul 2022

- Applied XGBoost, LightGBM and Sklearn package in Python to build Predictive Models; found 11 variables associated with the expected results from a large dataset with 25 different variables.
- Calculated click-through rate on high-dimensional features to separate data into bins, make every thousands-class feature into 5-class categorical factor, avoided sparse data problems and reduced program runtime by 15%
- Sub-sampled data to balance positive and negative cases, constructed stratified 3 fold cross validation and used grid-search method to find best hyper-parameters, improved accuracy of 5% on test score
- Drill-down every feature to clarify importance using exploratory data analysis, including data visualization and grouped Statistics; Selected features based on results of EDA to reduce complexity of training models

Legal Provision Machine Learning — Dec 2021

- Preformed word frequency method and several NLP methods to construct feature matrix and trained a 5 folds cross validation to classify legal text's category and obtained improvement of accuracy to 0.8
- Using reverse-translation method to add 25% noise to the dataset and augment sample size, made models more generalize on unknown samples, increased test accuracy by 5%

AWARDS AND HONORS

Kaggle M5 Forecasting Competition 103rd/5558 Top2% Silver Medal — Mar 2020 ~ Jun 2020
Estimate unit sales of Walmart retail goods

- Created features on tabular data using rolling-window method and trained a LightGBM algorithm to forecast daily sales for next 28 days; Built time series split to optimize hyper-parameters and acquired 0.58 of RMSE
- Stacked and averaged single LightGBM model predictions to improve the generalization ability of forecasts promoted 10% of ranking on leaderboard; Successfully avoided severe shake-down in privte leaderboard

TECHNICAL SKILLS

- Python (numpy, pandas, scikit-learn, tensorflow/keras), R (data.table, dplyr, shiny, tidyr)
- SQL(MySQL, SQL Server, TiDB, Clickhouse), Tableau, Linux/Shell, SPSS, Stata, Git, Web Crawler, Text Mining