GU ZHAO

156-3281-0960
Hebei, China
guzhao1990@outlook.com

Website: guzhao.github.io
GitHub: github.com/GUZHAO
LinkedIn: linkedin.com/in/gu-zhao

PROFESSIONAL SUMMARY

Detail-oriented and highly adaptable professional with 4+ years' experience in data ETL, machine learning and data visualization to solve challenging business problems. Passionate about combining front end skills with data science.

SKILLS

Programming Languages: Python, R, SQL, SAS, JavaScript, HTML
Frameworks: Angular 6, Bootstrap

Database: Microsoft SQL Server, MongoDB
Tools: Google Cloud, Git, KNIME

EXPERIENCE

Dana Farber Cancer Institute (Harvard-affiliated) — Boston, MA
Senior Data Analyst — 08/2017 – 12/2018
- Executed Palliative Care improvement project using elastic net algorithm which resulted in 20% enhancement of patient sequencing and 35% higher patient satisfaction.
- Conducted Palliative Care clinical notes analysis with NLP bag-of-words model and Naive Bayes to predict missing notes in order to better meet patient needs.
- Transformed raw data into consolidated usable data bucket in Google Cloud with BigQuery

Brigham Health (Harvard-affiliated) — Boston, MA
Data Engineer — 01/2016 – 08/2017
- Rebuilt data platform based on Microsoft SQL for 10M+ patient data and provided effective data validation monitoring processes which resulted in organized data governance and 50% speed increase in data ETL processes.
- Collaborated with BI team for system support and data pipe maintenance.
- Led in building internal team request and issue tracking system with Atlassian Jira, internal team wiki and documentation system with Atlassian Confluence which enhanced team work flow by 50%.

Brigham Health (Harvard-affiliated) — Boston, MA
Data Analyst — 08/2014 – 12/2015
- Developed time series predictive model for surgery duration estimation to improve operating room block utilization which led to cost reduction.
- Designed Python Plotly dashboards for dynamic reporting on internal pharmacy daily OTC medication and non-OTC medication monitoring which increased drug restocking speed by 10%.
- Developed descriptive statistics dashboards using Tableau to enable provider data monitoring process.

PROJECT

User Churn Prediction Project — 01/2018 – 02/2018
- Supervised learning project with model training including Random Forest, KNN and logistic regression and the all three models' accuracy are above 87%.

IMDB Document Clustering and Topic Modeling Project — 08/2018 – 10/2018
- Unsupervised learning project with model training including tf-idf, K-means and Latent Dirichlet allocation.

EDUCATION

Udacity — Boston, MA
Front End Web Developer Nanodegree — October 2018

George Washington University — Washington, D.C.
Master of Science, Statistics. GPA: 3.82 — May 2014

Qiongzhou University — Sanya, China
Bachelor of Science, Mathematics. GPA: 3.37 — September 2012
- National College Mathematical Modeling Contest – Third Prize