ZHAOYU QIAO

zhaoyuq.career@gmail.com | +16463317832 | www.linkedin.com/in/zhaoyuqiao | https://github.com/piapiaQ

EDUCATION

Carnegie Mellon University (CMU) *Pittsburgh, PA*
M.S. in Healthcare Analytics & IT | 30% Merit-Based Scholarship *Aug 2019 - May 2021*

- **Analytics Coursework**: Big Data & Large Scale Computing, Data Science and Big Data, Machine Learning, Applied Analytics: ML Pipelines, Unstructured Data Analytics, Data Mining, Database Management, Operation Research, Econometrics
- **Healthcare Coursework**: Healthcare Information Systems, Health Economics, Policy & Finance

Shandong University *Shandong, China*
B.A. in Risk Management and Insurance | Second Class Honorable Scholarship *Sep 2015 - Jun 2019*

- **Economics Coursework:** Actuarial Studies, Theory of Interest, Statistics & Probability, Macroeconomics& Microeconomics
- **Insurance Coursework:** Actuarial Studies, Theory of Interest, Risk Management, Life and Health Insurance

TECHNICAL SKILLS

Programming Language: Python (PySpark, PyTorch, scikit-learn, TensorFlow, Gurobi), R, R Shiny, SQL
Tools: Tableau, AWS, Git, ArcGIS, SAS Viya, Visio, Stata, Oracle EBS, Excel
Algorithms: Linear/Logistic Regression, KNN, Random Forest, XGBoost, MLP, CNN, RNN, PCA, T-SNE, Linear Programming

WORK EXPERIENCE

Great Lakes Behavioral Research Institute *Pittsburgh, PA*
Data Analyst, Behavioral Health Analytics Team *Aug 2021 - Present*

- Evaluated Hepatitis C testing/diagnosis/prescription prevalence and gap on 220k Allegheny County Medicaid members, mapped low testing rate region and providers for Health Department epidemiologists.
- Identified bereaved children and summarized their service utilization pattern (housing, education, mental health, jail).
- Maintained Mental Health Public Dashboard and fixed a data flow issue impacting 4000+ people with data warehouse team.

CMU, Highmark Health Funded Research *Pittsburgh, PA*
Research Assistant, Chronic Kidney Disease (CKD) Patient Risk Stratification and Profiling *Mar 2020 - Mar 2021*

- Streamlined claim data preprocessing of 10,000+ patients with 16 million records and conducted data Quality Control.
- Stratified patient cost and disease progression risks using longitudinal Group-Based Trajectory clustering model.
- Identified distinctive patient features with a clinical-claim cluster crosswalk analysis by profiling the clustering results.
- Pinpointed the high-risk group using 5 quarters of initial visit data with 80% accuracy and advised intervention strategies.

Highmark Health *Pittsburgh, PA*
Data Scientist Summer Intern *May 2020 - Jun 2020*

- Collected Highmark-COVID-19 related articles and classified them into 4 topics: testing, volunteer work, technology, and coverage
- Evaluated the social sentiment towards the 4 topics and suggested intervention strategy from identified operation gaps.

GE Healthcare *Beijing, China*
Supply Chain IT Intern *Jul 2018 - Jan 2019*

- Maintained Oracle EBS sales and distribution modules in 8 countries and troubleshoot over 200 system cases regarding global parts shipping, inventory management, part returns, repairs, and finance.
- Standardized data extraction process for daily case report using Python. Reduced case response time by 30%.

ACADEMIC PROJECTS

COVID-19 Vaccine Allocation Recommendation for Nursing Homes, Sponsored by SAS, Inc. *Sep 2020 - Jan 2021*

- Mined and engineered over 300 features from 8 data sources based on CMS expertise, NLP and RF feature selection models.
- Predicted infection and mortality risk in 13k US nursing homes by customizing RF and Light GBM classification algorithms.
- Optimized vaccine dispensing locations to minimize total supply chain traveling distance and built SAS dynamic web dashboard for CMS health officials to figure out the next location as well as strategy for vaccine allocation.

Song Hotness Classification using Million Song Dataset with PySpark *March 2021 - April 2021*

- Converted Million Song Dataset of 580k records to csv format using AWS EMR and stored resulting files on S3 bucket.
- Evaluated AUC performance improvement before and after adding TF-IDF converted features in LR and RF models.

Alzheimer's Disease (AD) Transition Risk Prediction Interactive Dashboard in R-Shiny *Feb 2021 - May 2021*

- Predicted risk of transition to AD within 6,12 and 24 months after initial visit with ADNI data.
- Embeded data summary, risk stratification over time and ad-hoc new patient import and prediction in R-shiny Dashboard.

Electricity Consumption Forecasting, Sponsored by Tata Consultancy Service, Inc. *Jun 2020 – Aug 2020*

- Managed remote team collaboration, communicated progress and roadblocks to clients and met 100% deadlines and goals.
- Forecasted monthly electricity consumption and achieved 4.3% mean absolute percentage error with Neural Network Model.