Huimin (Hannah) Han

huimin.han3@gmail.com | +1 301-454-9485

EDUCATION

Master of Science (MS), Business Analytics (STEM Program) **GPA:3.81/4**
University of Maryland, College Park, U.S. Expected 12/2019
Core Courses: Big Data and Artificial Intelligence for Business | Data Mining and Predictive Analytics | Data processing and Analysis in Python | Database Management Systems (SQL) | Data, Models and Decisions | Decision Analytics | Google Online Analytics Challenge | Price Optimization & Revenue Management

Bachelor of Management, Public Health Administration **GPA:3.4/4(Top 3)**
Capital Medical University, Beijing, China 09/2014-07/2018
Merit-based Scholarship
Core Courses: Advanced Mathematics | Linear Algebra | Probability& Statistics | Statistics in Health Management | Operational Research | Data Collection | Generality of Clinical Medicine | Health Service Administration | Health Care Economics | Computer System and System Software | Health Information Management

EXPERIENCE

Research Assistant College Park, MD, U.S.
University of Maryland, The Center for Health Information and Decision Systems (CHIDS) 08/2019- Present
Advisor: Gordan Gao

- Learn the evolution of the time series data forecasting in healthcare by historical literature reviews.
- Work with Inovalon, a provider of cloud-based platforms empowering data-driven healthcare, on predictive modeling projects
- Utilize historical data and applied ARIMA model on the data set to predict the pharmacy inventory
- Fit machine learning models after feature engineering to predict important outcome in healthcare field like pharmacy storage, patients cost, etc.
- Perform hyperparameter selection for models to improve model performance

Teaching Assistant (of Masters' Students) College Park, MD, U.S.
University of Maryland, R.H. Smith School of Business 03/2019- Present

- BUDT758P: Decision Analytics
  Main topics covered are linear, nonlinear optimization applications in industry segments, Monte-Carlo Simulation, and risk assessment
- BUSO758M: Decision Modeling
  Introduced how to use model to make better decision
- BUSI634: Operations Management
  Covered key concepts and tools of operations management, including process flowcharting, analysis of process flows and bottlenecks, capacity planning, batching and EOQ, waiting line management, newsvendor inventory model, quality management, six-sigma, and lean operations

Data Analyst Intern Beijing, China
Xuanwu Hospital 11/2017- 05/2018

- Worked with multiple sources of healthcare data (ex: financial and clinical data) to meet both internal and

external reporting requirements and support new initiatives
- Identified opportunities for improving clinical processes and outcomes by using sophisticated and creative approaches of analyzing outcomes and utilizing data
- As new sources of data became available, learned the data structures, wrote queries, linked, and accessed data for manipulation and analysis

PROJECTS

Capstone Project Collaborated with real estate investment company: Roofstock 09/2019- 12/2019(expected)
- Preprocessed and merged several datasets from a relational database of the real estate company Roofstock
- Extract listing fair market value using Zillow API
- Build model to predict 'time to sale' which means how long it will take from a listing once posted on the website to be sold

Predictive Modeling of Average Visitor Rating of Airbnb Listings 06/2019- 08/2019
- Preprocessed the datasets, removed redundant variables, imputed missing values, selected valuable variables, and applied target encoding to avoid high computational pressure
- Explore features and visualized data to uncover suspicious patterns in data
- Built several machine learning models in Python to predict whether a new listing will get high booking rate. Finally chose Random Forest and got an accuracy 10% higher than baseline

Machine Learning Readmission Modeling 03/2019- 05/2019
- Obtained data from JHU SoM, preprocessed datasets according to the principle of statistics in R, selected proper variables to deal with multicollinearity, and applied regularization to avoid overfitting
- Built models (regression models, tree models along with ensemble methods, KNN models, etc.) to predict if a discharged patient will return within 30 days.
- Chose F1 scores to evaluate models since the dataset is highly imbalanced.

Exploratory Data Analysis and Text Mining of Google Play Store Apps 03/2019- 05/2019
- Preprocessed Kaggle datasets, performed statistical and visual exploratory data analysis to find common characteristics of high-rating and high-downloads apps
- Extract features from the reviews text data. Prepared text data, selected features using bag-of-words and used topic modeling to discover any natural topics in the reviews
- Provided specific suggestions for the app-making business of different industries according to results of analysis

SKILLS

- **Machine Learning Algorithms:** Hands on experience of Regression (Linear, Logistic, Stepwise), Instance-based (KNN), Regularization (LASSO, Ridge), Decision Tree (CART), Clustering(K-Means), Dimensionality Reduction (PCA, LDA, QDA), Ensemble (Bagging, Boosting, Random Forest)
- **Programming languages & Visualization tools:** Python (pandas, numpy, sklearn, matplotlib, seaborn), R, SQL, Tableau