Huidi (Scarlett) Wang

Phone: (626)888-0650 | Email: huidiwlett@gmail.com

EDUCATION

University of California, Berkeley Expected: 05/2020
Master of Engineering - Industrial Engineering and Operations Research GPA: 3.8/4.0
University of California, Berkeley 12/2016
Bachelor of Art in *Statistics and Bachelor of Art* in *Economics* GPA: 3.6/4.0

SKILLS

Python, R, SQL, Tableau, Excel, Stata, AMPL, Git, JIRA, Experience with AWS

WORK EXPERIENCE

Beam Solutions, San Francisco
Data Scientist Intern 09/2019 – Now

- Lead the project to identify potential fraudulent or malicious financial activities, by building machine learning models on client text data and contextual feature in Python Numpy and Pandas
- Transform corpus of customer feedback and project summary text into model-ready summaries by using NLTK
- Train word and sentence embeddings from Facebook FastText and Google's Word2Vec
- Build and validate neural network models, e.g. framework Keras, Recurrent neural network (RNN), sequence models, Gated Recurrent Unit (GRU), and Long-short term memory (LSTM), for multi-label classification cases

Squaretrade, San Francisco
Customer Operations Analyst 04//2018 – 07/2019

- Collaborate with data analytics team to analyze Amazon reviews, identify customer preference, and develop a more efficient file-a-claim process
- Lead manufacturer referral project in analyzing customer complaints, optimizing systems and improving customer service
- Define metrics to measure new features, conduct A/B testing for significant impact in further make launching decisions

Arrowhead Credit Union, Rancho Cucamonga
Business Intelligence Analyst 08/2017 – 04/2018

- Work on large dataset of real time financial transactions, design SQL query, and use business intelligence tool, e.g. SQL Server Reporting Services, to develop insights in dashboard
- Collaborate with Loss Prevention team to automate merchant credit detection for card disputes by creating a platform to monitor accounts and search on a variety of criteria. Saved projected $96,931.26 annually for organization
- Use statistical predictive modeling to evaluate loan promotion scenarios and make predictions on future outcomes

PROJECTS

Airline Sentiment Analysis

- Build supervised learning model in R that efficiently predicts multiple US airlines passengers' sentiment (Negative, Positive/Neutral) out of tweets
- Use tokenization with stemming and lemmatization to convert user reviews to vector space of word frequency (tf-idf)
- Generate supervised sentiment prediction model by comparing Logistic Regression, LDA, CART, Random Forest, Boosting, etc. Apply model blending method to improve accuracy
- Validate model performance through ROC curve, classification rate and k-fold cross validation

Kaggle Housing Price Prediction

- Build a multivariate linear regression model on Housing Price prediction dataset in R
- Impute missing data using multiple ways like mean, KNN, multivariate imputation by chained equation (MICE) algorithm, etc. and performed feature selection through exploratory analysis
- Fit linear regression model with regularization to control for multicollinearity by Lasso and Ridge
- Build and compare decision tree, random forest, and boosting models to predict housing price

Collaborative Filtering for Music Listeners

- Infer each user's ratings of all songs based on other users' listening behavior. Recommend specific songs to specific users
- Train and evaluate Collaborative Filtering, Random Forest and Blending models with over 24K observations