

Huidi (Scarlett) Wang

Phone: (626)888-0650 | Email: scarlett128whd@gmail.com

SKILLS

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

WORK EXPERIENCE

Beam Solutions, San Francisco

Data Scientist – Capstone Team

09/2019 – 05/2020

- 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
- Apply pre-trained word and sentence embeddings from FastText and Word2Vec to OneVsRest regressions and classification models like Decision Tree and Random Forest. Receive 80.28% accuracy
- Build and validate neural network models using Tensorflow. Train sequence models – RNN, GRU, and Long-short term memory (LSTM), for multi-label classification forecast. Improve performance to 89.38% to detect suspicious accounts

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. Use SQL in a daily basis
- 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 visualization tools to show data 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, ensemble learning like Random Forest and Boosting. Apply model blending method to improve accuracy
- Validate model performance through ROC curve, precision-recall score, confusion matrix, etc.

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

EDUCATION

University of California, Berkeley

Master of Engineering - Industrial Engineering and Operations Research

05/2020

GPA: 3.8/4.0

University of California, Berkeley

Bachelor of Art in Statistics and Bachelor of Art in Economics

12/2016

GPA: 3.6/4.0