Xiao (Claire) Zhang

xzhang7@scu.edu, 408-930-8189

Objective: Applied Machine Learning Engineer/Data Scientist (Full-time)

WORK EXPERIENCE

KLA-Tencor | Global Service Department Milpitas, CA
Data Scientist Summer Intern July 2018 – Nov 2018

- Worked in a 10-people team across 2 divisions and owned 3 projects.
- Cleaned, preprocessed data and applied Cpk and Specification analysis using JMP and python on 10 tools' data with around 200 features; identified latent tool problems based on analysis results and increased original team efficiency and accuracy by 75%.
- Conducted correlation analysis to accelerate feature engineering; came up with models using XGBoost, Random Forest, Logistic Regression to forecast future tool service action in POC process; visualized results in Tableau.
- Created data parsers using python to extract specific data from html and txt log files to replace original manual data collection and boost efficiency to next level.

EDUCATION

Santa Clara University (GPA: 3.5/4.0) Santa Clara, CA
Master of Science in Information Systems December 2018
Relevant Course: Machine Learning, Deep Learning, Business Intelligence and Data Warehouse, Data Science Analysis with Python, Dashboard and Data Visualization, Big Data Modeling & Analytics, Object-Oriented Analysis & Programming, Database Management System.

Hangzhou Dianzi University (GPA: 4.02/5.0) Hangzhou, China
Bachelor of Engineering in Electrical Engineering of Automation June 2015

ACADEMIC PROJECTS

Black Friday Purchase Prediction (Boosting & Bagging): https://github.com/clairezhang2018/Machine-Learning

- Data exploration on 550068 samples; cross checked missing data ratio and correlation heatmap and dropped columns with high null value ratio; created 3 new features and proposed models with regard to EDA results.
- Evaluated XGBoost Regressor and Random Forest models using cross validation; Increased model accuracy by around 3% after introducing 3 new features and 2 out of 3 become top 5 important features.

Ames House Price Prediction (Stacking): https://github.com/clairezhang2018/Machine-Learning

- Data exploration on 2920 samples with 80 features in each sample; applied bivariate and multivariate analysis to remove outliers; imputed missing data and drop columns based on null value ratio and correlation heatmap; used Box-Cox Transformations on skewed numerical data.
- Evaluated Lasso, Elastic Net, Gradient Boosting Regressor and XGBoost Regressor models using cross validation with around 0.12 RMSE score, increased model accuracy by 6% via building up an averaged stack model with four models together.

Cat Image Recognition (Neural Network with Classification): https://github.com/clairezhang2018/Deep-Learning

- Preprocessed data including reshaping image data into vectors and data normalization.
- Built a 2-layer shallow neural network. After initializing parameters, defined the forward and backward propagation to learn parameters; identified cost function and computed derivatives to optimize model and achieved 80% accuracy.

TECHNICAL SKILLS

- **Skills:** Python, Java, Pyspark, MySQL, Octave, Matlab, C, C++, Tableau, Pentaho
- **Libraries:** SciKit-Learn, Keras, TensorFlow, Statsmodels, Scipy, Numpy, Pandas, Matplotlib, Seaborn