YANNAN PAN

panxx396@umn.edu 612-814-5133 Minneapolis, Minnesota

EXPERIENCE

Data Scientist Intern

Seagate Technology

May 2017 – Present Edina, MN

- Work with engineers to manipulate, visualize and analyze large data sets using SQL, Python and R.
- Use NLP techniques such as Tf-Idf, N-grams to measure the text similarity for both structured and unstructured data.
- Performed anomaly detection in images using quantile thresholding and One-class SVM.
- Developed a scoring system to predict product scrap using various classifiers, such as GLM, XGBoost, Random Forest, KNN and performed unsupervised clustering to detect anomaly.
- Handled extremely imbalanced datasets by sampling with SMOTE, training cost-sensitive classifiers, etc.

Teaching Assistant

University of Minnesota Twin Cities

Jan 2016 – Dec 2016 Minneapolis, MN

- Lead discussion sections and demonstrate how to solve statistical problems in regression using R.
- Meet with instructors to discuss lab topics and hold regular office hours to meet with students.

PROJECTS

Toxic Comments Classification

Kaggle Project

Jan 2018 – Present

- Preprocessed large text datasets using regular expression functions, tokenization and lemmatization tools
- Extracted features by Bag-of-Words, Tf-Idf representations and used machine learning alorithms to perform classification
- Trained RNNs with pre-trained GloVe embeddings
- Tools: Python (Sklearn, NLTK, Keras)

Predictive Modeling on Auto Claim Data

Coursework Project

Oct 2015 - Dec 2015

- Built predictive models to predict claim costs based on historical data
- Performed model selection and variable selection to choose from generalized linear models
- Trained a model with high Gini coefficient and provided a good segmentation of the policyholders to help reduce claim costs
- Tools: R

EDUCATION

Master of Science, Statistics

University of Minnesota Twin Cities

2015 - 2018 Minneapolis

GPA: 3.83/4

Bachelor of Economics, Mathematical and Financial Statistics

Zhongnan University of Economics and Law

2011 - 2015 Wuhan

GPA: 3.72/4

COURSEWORK

- Advanced Regression Techniques
- Engineering Optimization
- Machine Learning
- Design of Experiments
- Statistical Computing
- Statistical Consulting

AWARDS

First Prize
in the 4th Chinese National College Student Market Research and Analysis Final Competition, Aug 2014

Honorable Mention
in American College Mathematical Contest in Modeling, Jan 2014

STRENGTHS

R | Python | SQL | SAS | LaTex | Microsoft Office | Matlab

CERTIFICATIONS

- SAS Certified Base Programmer for SAS 9
- Chinese National Market Research and Analysis Specialized Skills Certificate