Chang SU
Email:suchang97@outlook.com

EDUCATION

Sun Yat-Sen University (SYSU) Guangzhou, China
Bachelor of Science in **Statistics** 09/2014-06/2018 (Expected)

- Overall GPA:3.9/4.0

University of California - Berkeley (Exchange Student) Berkeley, California
Department of Mathematics & Department of Statistics 08/2016-12/2016

- Overall GPA:4.0/4.0
- Introduction to Abstract Algebra(A+); Concepts of Statistics(A); Concepts in Computing with Data(A)

GRE: V159 + Q169 + 4.0; **TOEFL**: 111

PROJECTS

- **2016 US Presidential Election Debrief** (**Concepts in Computing with Data** course project) Berkeley, US
- Advisor: Prof. Deborah Nolan 11/2016-12/2016
  - *Abstract:* Carried out EDA (exploratory data analysis) on the election and census data and built prediction models to identify factors vital to the victory of Republican
  - Visualized the comparison of election results in 2012 and 2016 on a map; Visualized relations between input variables via correlation matrixes of Pearson correlation and Kendall rank correlation
  - Identified factors important to the victory of Republican via the variable importance by Random Forest and the step-wise feature selection via k-NN (k-Nearest Neighbor) algorithm

- **Job Salary Prediction** (**Data Mining** course project); SYSU, China
- Advisor: Prof. Xueqin Wang & Yanbo Shen 03/2017-06/2017
  - *Abstract:* Predicted the job salary based on the recruitment data with 240,000 pieces of ads
  - Visualized the mean salaries at different locations (clustered by the k-means algorithm) in UK on a map. This relation between salary and location is similar to that between GDP and location as revealed by a map from Eurostat. (Link for the graph)
  - Applied k-means clustering on the result of Word2Vec to categorize the words similar in meaning. The job title is then modelled by one-hot encoding of these synonym groups. The power of these generated features is justified by the variable importance of XGBoost, our prediction model.

- **Prediction of Breast Cancer Data with Lasso Cox Model** (**Survival Analysis** course project); SYSU, China
- Advisor: Dr. Xiaobo Guo 06/2017-07/2017
  - Adopted LASSO and assumed linearity in predictor variables to avoid overfitting
  - Quantified the predictive ability of the model using overall C-index, AUC and calibration curve
  - Plotted nomogram to visualize the model

- **Classification of Glasses** (**Nonparametric statistics** course project); SYSU, China
- Advisor: Prof. Xueqin Wang 07/2017
  - *Abstract:* Experimented with nonparametric statistics methods on the glass identification data set
  - Improved k-NN algorithm with kernel smoothing (adopted the so-called weighted k-NN) to reweight the nearest neighbors by respective distances
  - Improved the performance of tree-based models on an imbalanced dataset with SMOTE

Chang SU
Email:suchang97@outlook.com

- Compared k-NN-based models with tree-based models (Decision Tree, Random Forest, Adaboost) on their performances on an imbalanced dataset; found that tree-based models performed poorly on minority class, while k-NN was more robust to the class imbalance problem; inferred that the distribution of samples in a multidimensional feature space was more crucial than the balance among classes for k-NN-based models.
- Rigorous experiments on more datasets are needed to generalize the inference above.

WORK EXPERIENCE

Institute of Advanced Computing and Digital Engineering (Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences) 07/2017-Present

Working Project: **Mobility Prediction Algorithm for BMW Premium Carpooling**

- *Abstract:* Predicted the departure time, destination and mode of a user with driving recorder data
- Calculated the mutual information between destinations and predictor variables to quantify how predictable a user is
- Introduced kernel density estimation into Naïve Bayesian classifier, resulting in an increase of 15.2% in coverage. (The goal of this algorithm is to achieve the accuracy of 0.7 covering 70% of all users)
- Tried bagging to ameliorate over-fitting; verified the conclusion by a paper that as a stable model, Naïve Bayesian classifier could hardly be improved by bagging.
- Implemented k-means clustering and the elbow method for optimal k selection in Scala
- *Achievement:* The destination prediction algorithm achieved an accuracy of 0.7 in 70% of the users, while that of DIDI (China's Uber) is 0.9 in 30%. Results of all algorithms exceeded the criteria by BMW.

SKILLS

Proficient in R; Familiar with C, C++; Basic in Scala and Spark, Python, LaTeX

ACTIVITIES

- **Main Debater** in **the Debate Team** 09/2014-12/2015
  - Participated in debate contests on a wide range of topics
- **Head** of **the Entertainment Department** 06/2015-06/2016
  - Organized two inter-school activities and one inter-university activity
- **Volunteer** in **UCB Circle K** 08/2016-12/2016
  - Provided service for the underrepresented and the community

HONORS AND AWARDS

National Academic Scholarship (top 2%) 2015-2016
First Prize Merit-based Scholarship, SYSU (top 5%) 2015-2016
Honorable Mention in Mathematical Contest in Modeling (top 25%) 2015, 2016
Excellent Student Leader of School of Mathematics 2015-2016
Team Championship in the SYSU Inter-School Debate Competition 2015
Second Prize Merit-based Scholarship, SYSU (top 15%) 2014-2015, 2016-2017