Tonia (Tong) Chu

Eagan, MN 55122 Cell: 510-693-7608 E-Mail: tonia.chu@gmx.com
https://www.linkedin.com/in/tonia-chu-6021686a | https://github.com/anxin16

Skills:

- **Machine Learning:** classification, regression, clustering, feature engineering
- **Software and Programming Languages:** Python (scikit-learn, NumPy, SciPy, pandas, nltk, gensim), R, Rstudio, SQL, Microsoft Excel
- **Data visualization:** ggplot2, matplotlib, pandas, seaborn, Bokeh
- **Language:** Mandarin Chinese, English

Education:

Data Science Career Track, Springboard **2017**

- Use Python, R, SQL, Spark to do hands on mini-projects and capstone projects.
- Data Wrangling, Data Visualization, Statistics, Data Storytelling, machine learning, and NLP
- Machine learning methods include Linear and Logistic Regression, SVM, Decision Trees and Random Forests, Bayesian Methods, KNN, K-Means clustering.
- Advanced topics on machine learning include Recommendation Systems, Time Series Analysis, Anomaly Detection, Modern NLP, Neural Networks, Density-based clustering methods.
- Data Science at Scale include Spark with Python, Machine Learning on Apache Spark MLlib

Tsinghua University, Beijing, China

- Master of Science in Microelectronics **1996**
- Bachelor of Electronic Engineering **1994**

Projects:

Engagement and satisfaction of employees in the Federal workforce (coded in R)

- Analyzed and identified the working status with data of Federal Employee Viewpoint Survey.
- Built logistic regression model to predict employees' satisfaction and engagement.
- Gave results on how to improve the feelings of federal employees.

Factors associated with differences in Life Expectancy across the United States (coded in Python)

- Analyzed the differences of life expectancy by income, over time, and across areas.
- Built models of Linear Regression, SVR, and Random Forest Regressor to predict the life expectancy of individuals by their age, income, living area and other aspects.
- Used Feature Selection methods of PCA, Regularization, and Random Forests to find out the factors that associated with Life Expectancy most.
- Used Scikit-learn, Pandas, NumPy, Matplotlib, and Seaborn.

How many stars will I give? Predicting ratings of Amazon reviews (coded in Python)

- Used techniques of Tokenizing, Removing Special Characters, Expanding Contractions, Case Conversions, Removing Stop words, Correcting Words, and Lemmatization for text normalization of Amazon reviews.
- Developed models of Naive Bayes, SVM, and Random Forest to predict rating from text review.
- Used nltk, gensim, scikit-learn, re, Pandas, NumPy, Matplotlib, and Seaborn.

Working Experience:

SQL Server DBA & Data Analyst **01/2014 – 01/2017**
ACE Data Services LLC **Chicago, IL**

- Analyzed data and developed recommendations to improve marketing efforts.
- Improved performance and functions by resolving database related incidents and requests.
- Worked with software programmers to manage SQL Server tables and Stored Procedures.

Product Instructor **03/2012 – 05/2013**
Raqsoft Co. Ltd. **Beijing, China**

- Created SQL self-study materials and video media for university student candidates.
- Composed Raqsoft product manual which was used by employees and public users.
- Evaluated potential employees and made hiring recommendations to manager.

Work Authorization: Fully Work Authorized. Permanent resident.