TONIA CHU

DATA SCIENTIST
DATA ANALYST

I love Data Science! I think Data Science is magic, it can get very important and valuable information from big messy data. As a data scientist, I will help change the world and make people live better.

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
- Highly developed problem solving skills

CONTACT DETAILS

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

PROJECTS

1. Engagement and satisfaction of employees in the Federal workforce
+ Analyzed and identified the working status with data of Federal Employee Viewpoint Survey.
+ Built logistic regression model to predict employees' satisfaction and engagement.
+ Presented results on how to improve the feelings of federal employees.
2. Factors associated with differences in Life Expectancy across the United States
+ 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.
3. How many stars will I give? Predicting ratings of Amazon reviews
+ Used techniques of Tokenization, Removing Special Characters, + Expanding Contractions, Removing Stop words, Correcting Words, and Lemmatization for text normalization of Amazon reviews.
+ Developed models of Logistic Regression, Multinomial Naive Bayes, Linear SVC, SGD and Random Forest to predict rating from text review. Optimized models with Hyperparameter Tuning.

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.

Tsinghua University, Beijing, China
Master of Science in Microelectronics 1996
Bachelor of Electronic Engineering 1994

WORKING EXPERIENCE

Product Instructor 03/2012 - 05/2013
Raqsoft Co. Ltd. Beijing, China
+ Composed Raqsoft product manual which was used by employees and public users.
+ Created SQL self-study materials and video media for university student candidates.
+ Evaluated potential employees and made hiring recommendations to manager.

Work Authorization: Fully Work Authorized. Permanent resident.