Xu Dou

Ontario CA 91761 dowxdou@gmail.com
405-414-9483

Summary

- ❏ Data analysis and visualization, machine learning methods; quantitative and research skills; more than five years' experiences in theoretical modeling and numerical research of condensed matter physics; experience with data ETL processes (CSV, JSON).
- ❏ Focus on identifying and solving problems.
- ❏ Courses: Machine Learning (Coursera)

Skills

- ❏ Programming: Python, SQL
- ❏ Software: Mathematica, Matlab
- ❏ Technical tools: Spark, LaTeX, Git, scikit-learn, Pandas, Numpy, matplotlib

Education

University of Oklahoma，US - *Ph.D. in theoretical physics; focus on modeling*

August 2011 - August 2018

Nankai University，China - *BS in physics*

September 2005 - June 2009

Experience and projects

University of Oklahoma, 06/2012 - 11/2018

Research assistant

- ❏ Quantitative model building and numerical calculations (Python, Matlab, Mathematica); made experimental predictions and data analysis; gave talks in local seminars and the American Physical Society meetings.
- ❏ Prepared paper drafts; published papers in peer-reviewed journals as the first author.

Nankai University, 06/2008 - 01/2011

Student research assistant

- ❏ Build phenomenological cosmology models and analyzed cosmological data; determined cosmological model parameters by using regression analysis; published two papers in a peer-reviewed journal.

Data Projects

Yelp Business data analysis

- ❏ Used python to clean, preprocess and analyze large data sets of customers' reviews. Used unsupervised learning techniques (PCA, K-Means clustering) to categorize users' short reviews into groups; Used machine learning methods (Logistic Regression, Random Forest) to understand business performance information.

Bank Marketing Conversion Prediction

- ❏ Extracted categorical features from a bank product conversion rate data set. Performed features selection through exploratory analysis; performed systematic feature engineering to reach significant prediction results.
- ❏ Built different machine learning models with regularization to predict conversion rate, including Logistic Regression, Gradient Boosting, and Random Forest; obtained high ROC_AUC score by grid searching.