

# Yan Yin

DATA SCIENCE · DATA ANALYSIS

Cornell University, Ithaca, NY 14853

☎ (607) 319-9282 | ✉ yy855@cornell.edu | 🏠 yinyan5988.github.io/ | 📱 yan-yin-588461168

## Summary

- An enthusiastic learner with experiences in Data Science, Data Analytics, Recommender System, Natural Language Processing and Machine Learning. Skilled in data analysis, machine learning, mathematics, statistics, programming and creative problem solving. Having zeal to explore new areas and take challenges. Good at communicate with different kind of people.
- **skills:** R, Python, Tensorflow, Pytorch, MATLAB, C++, SPSS, LaTeX, SQL, SAS, Photoshop

## Education

### Cornell University

MPS IN APPLIED STATISTICS

Ithaca, NY

Aug. 2018 - May. 2019

- Courses: Machine Learning, Database Management, Applied Statistical Computation with SAS, etc.

### Renmin University of China

B.S IN STATISTICS MINOR IN ECONOMIC STATISTICS

Beijing, China

Sep. 2014 - June. 2018

- Cumulative GPA: 3.78 / 4.0 Ranking: 4 / 27 GRE: 157+168+3.5 TOEFL: 105
- Received Academic Excellence Scholarship in three consecutive years (top 5%)
- Courses: Statistical Computing, Time Series, Nonparametric Statistics, Probability Theory, Database Techniques, C Programming, etc.

## Research Experience

### Deep Learning Collaborative Model on Implicit Feedback Datasets

Feb 2018- Jun 2018

- Combined matrix factorization and stacked denoising autoencoder algorithm on implicit feedback dataset aiming at cool start and sparse matrix and compared the model with PMF model

### Split Questionnaire Design for Long Questionnaires

Jan 2018- Jun 2018

- Proposed a design dividing a lengthy questionnaire into subsets of survey items and then administered each subset to distinct subsamples to solve low response rate and poor quality of survey data
- Used predictive mean matching multiple imputation dealing with empty items

### Textual Analysis Towards Spring Festival in Microblog

Jul 2017- May 2018

- Extracted 60GB data with SQL and cleaned it; did Exploratory Data Analysis with ggplot2 to maximize insight into our data; used light-LDA and association rules to explore topics and relations in shows and customers
- Explored online information transmission rules by characterizing the behavior of follower and being followed respectively, and used a Bayesian approach to capture repost dynamics

### Bike Demand prediction for Bike Sharing System

Dec 2016- Jan 2017

- Predicted occupancy of stations in long term and short term using stochastic network model and Principle Component Regression with R and Matlab; used the maximal covering location model to locate stations.

### Investigation of Land Rights Reform Process

Oct 2016- Feb 2017

- Conducted a national survey in 152 cities; found factors influencing satisfaction towards land rights reform using A/B test
- Conducted cross-sectional and longitudinal studies to analyze the status of women in rural areas with regression model, data visualization techniques and hypothesis testing

### Relationship between the Value of Children(VOC) and Desired Fertility

May 2016- Apr 2018

- Received 20000 research fund in 2016 Undergraduate Innovative Test Program (top 5%)
- Designed a questionnaire and surveyed more than 400 people under the two-child policy background
- Used factor analysis, lasso and hypothesis testing to explore the influence of VOC on desired fertility in three dimensions and established a hybrid model to explain fertility desire

## Internship

### Drilling Dynamics and Computation Engine Team, Schlumberger Ltd.

Beijing, China

DATA SCIENTIST

Oct. 2017 - Mar. 2018

- Predicted users' preference in trajectory design with modified Random Forest
- Communicated with software engineers and product managers about demand and process
- Wrote reproducible python scripts to extract, clean and convert 3.5 million data from database
- Characterized trajectory structures using feature engineering and partition-and-group clustering

### Brand Marketing Department, Baidu Inc.

Beijing, China

AI RESEARCHER

Jul. 2017 - Oct. 2017

- Collected and tracked daily international Artificial Intelligence reports by building a web crawling
- Designed a market research towards truck drivers before drowsiness detection devices released