

Yan Yin

DATA SCIENTIST

Cornell University, Ithaca, NY 14853

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

Education

Cornell University

Ithaca, NY, USA

MASTER IN APPLIED STATISTICS

Aug. 2018 - May. 2019

- GPA: 4.0/4.0
- Courses: Machine Learning, Data Science, Database Management, Computer Vision, Applied Statistical Computation with SAS, etc.

Renmin University of China

Beijing, China

B.S IN STATISTICS MINOR IN ECONOMIC STATISTICS

Sep. 2014 - June. 2018

- Cumulative GPA: 3.78 / 4.0 Ranking: 4 / 27 Minor GPA: 3.88 / 4.0
- Received Academic Excellence Scholarship in three consecutive years (top 5%)
- Courses: Statistical Computing, Time Series, Nonparametric Statistics, Probability Theory, Database Techniques, etc.

Research Experience

Design Algorithm for AB Test

Jan 2019- present

- Designing mechanism for A/B testing analysis to make e-commerce business decisions

Collaborative Deep Learning Model on Implicit Feedback Dataset(Python)

Feb 2018- Jun 2018

- Designed a recommender system for implicit dataset using matrix factorization
- Learned item latent representations using stacked denoising autoencoder

Textual Analysis Towards Spring Festival in Microblog(Python, SQL, ggplot2)

Jul 2017- May 2018

- Explored topics and relations applying latent Dirichlet allocation model and association rules; extracted and cleaned 50 million data; visualized data with ggplot2;
- Built the microblog transmission network; utilized the overall network structures to describe patterns of information flow; explored online information transmission rules by characterizing the behavior of follower and being followed respectively
- Designed a sentiment analysis model using Naive Bayes classifier based on an emoji training heuristic

Bike Demand prediction for Bike Sharing System(MATLAB,R)

Dec 2016- Jan 2017

- Predicted long-term and short-term occupancy of stations with Markov Chain system and Principle Component Regression respectively; located stations with maximal covering location model
- Designed a k-means clustering algorithm to find unbalanced stations and damaged bikes

Mining Relationship between the Value of Children and Desired Fertility(R)

May 2016- Apr 2018

- Explored latent variables using factor analysis; found significant variables applying hypothesis testing; explained fertility desire with logistic regression
- Received ¥20,000 research fund in 2016 Undergraduate Innovative Program (top 5%)
- Designed a questionnaire and surveyed more than 500 people with students in Psychology department

Internship

Drilling Dynamics and Computation Engine Team, Schlumberger Ltd.

Beijing, China

DATA SCIENTIST

Oct. 2017 - Mar. 2018

- Designed a recommendation engine using Bayesian method for drilling software
- 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
- Worked with product managers to define scopes and demand, and prioritized the project

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 about drowsiness detection devices

Campus Activities

Statistical Research Association, Renmin University of China

Beijing, China

MINISTER OF HUMAN RESOURCE DEPARTMENT IN THE ASSOCIATION

Sep. 2014 - Sep. 2016

- Organized research projects such as annual Campus Satisfaction Survey
- Responsible for staffing in activities held by the association

Skills

- Machine Learning, Statistics, Data Analysis, Data Visualization, Natural Language Processing, Recommender System, Deep Learning
- Python, SQL, R, Tensorflow, Pytorch, ggplot2, MATLAB, C++, SPSS, LaTeX, SAS, Photoshop