

# Yan Yin

4 YEARS OF EXPERIENCE IN DATA ANALYSIS AND MACHINE LEARNING

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 GPA: 4.0/4.0

Aug. 2018 - May. 2019

- Courses: Machine Learning, Data Science, Database Management, A/B testing by Google (Udacity)

### Renmin University of China

Beijing, China

#### B.S in Statistics Minor in Economic Statistics

Sep. 2014 - June. 2018

- Courses: Statistical Computing, Time Series, Statistical Software, Probability Theory, Econometrics

## Skills

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

## Internship & Research Experience

### Design Algorithm for AB Test, Data Analyst, Nation Info Cop

Jan 2019- present

Adivisor: Prof. Yang Ning, Associate Professor of Statistics and Data Science Department

- Designed an A/B test to improve conversion of the home-rental website
- Built a dashboard to analyze and present A/B test results
- Validated independence assumptions using mutual information and bootstrapping

### Textual Analysis Towards Spring Festival in Microblog, Research Analyst, CTR China

Jul 2017- May 2018

Adivisor: Prof. Yang Li, Professor of Statistics Department

- Explored topics and relations applying latent Dirichlet allocation model and association rules; extracted and cleaned 50 million data from SQL; 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

### Drilling Computation Engine Team, Data Scientist, Schlumberger Ltd. Beijing, China

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
- Integrated with product managers and peers to define scopes and demand, and prioritized the project

### Brand Marketing Department, AI researcher, Baidu Inc. Beijing, China

Jul. 2017 - Oct. 2017

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

### Collaborative Deep Learning Model on Implicit Feedback Dataset(Python)

Feb 2018- Jun 2018

Adivisor: Prof. Yanyun Zhao, Dean of Statistics Department

- Designed a recommender system for implicit dataset using matrix factorization
- Learned item latent representations using stacked denoising autoencoder
- Put forward ideas of designing hierarchical neural network in future work

## Project

### Text Analysis of Trump's tweets

Dec 2018- Jan 2019

- Cleaned and visualized Trump's tweets on Android and iPhone. Concluded that the Android and iPhone tweets were from different people, posting during different times of day and using hashtags, links, and retweets in distinct ways.
- Predicted the devices Trump's tweets came from using sentiment analysis, decision tree and Xgboost

### 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 results 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