

# Yidan Wang

richardyidan@gmail.com | (310) 210-5614  
870 Hilgard Ave, Los Angeles, CA 90024

## Summary

---

Future data scientist. Current M.S. Analytics graduate student at UCLA Anderson (expected Dec 2018). Former Data Engineering & Algorithm Intern at TalkingData, China's top 1 mobile data firm. Experience with mining 200 TB of data using Python, Scala, Java and Apache Spark. Proficient in data mining, feature engineering, machine learning, and large-scale parallel data processing.

**Languages:** Python, R, MySQL, SAS, Tableau | Scala, Java, Apache Spark, OCaml | MATLAB, VBA | LaTeX

**Packages/APIs:** Scikit-Learn, Matplotlib, Pandas, NumPy, SciPy | Scala/Java collections, Spark Core/SQL | ggplot2, tidyr

**Math Awards:** Silver Medal, China Mathematical Olympiad; Meritorious Winner, the Mathematical Contest in Modeling

## Work Experience

---

### TalkingData

Beijing, China

*Data Engineering & Algorithm Intern, Python/Scala/Java/Apache Spark*

*Aug 2017 - Dec 2017*

- **Apache Spark:** Mined 200 TB mobile device location data (1000 executors, 1 TB memory) using Spark Core, Spark SQL and Scala Collections to detect patterns in population's geographical distribution and mobility.
- **Machine Learning:** Constructed XGBoost and Random Forest Classifiers in python to help categorize GeoHash block by its functions. Engineered hourly human traffic features by encoding coordinates using a GeoHash Class written in Scala.
- **Bitmap:** Studied billion-scale population migration characteristics. Created multiple Compressed EWAH Bitmaps to store presence/absence info across different time frames and regions, which drastically improved the batch query processing time.
- **Data Engineering:** Parsed ArcGIS Geofence raw data using Pandas (Python) and Scala Collections to converted it into a standardized format for creating a Scala Map of R-trees.
- **Brand Association Mining:** Measured brand connections, mined association rules and brand aggregation patterns using the Apriori algorithm to help improve store operations and optimize brand synergy effects.
- **R-tree Indexing:** Sped up geographical area indexing of GPS coordinates by replacing traditional polygon methods with R-tree indexing (implemented in Scala), which uses tree data structure for spatial access methods.

### McKinsey & Company

Beijing, China

*Strategic Consulting & Data Visualization Intern, Tableau/think-cell*

*Nov 2016 - Jan 2017*

- **Data Visualization:** Developed corporate strategy for a cosmetics company to improve its offline sales performance. Visualized key business indicators using think-cell and Tableau to perform brand assessment and customer segmentation analysis.
- **Strategy Diagnosis:** Conducted marketing strategy analysis for peer companies in terms of total spendings, investment modes and promotion channels through literature research and expert interviews.
- **Improvement Matrix:** Created operations improvement matrix in terms of BC consultation and skin test services.

### Bain & Company

Beijing, China

*Data Analytics Intern, VBA/think-cell*

*Apr 2016 - Aug 2016*

- **VBA Automation:** Created a VBA mini project (100+ lines) to automate formerly manual process of merging 2,000 sales data worksheets according to predefined rules. Increased data loading and processing efficiency by 2500%.
- **Predictive Models:** Assessed the potential market value of a state-owned agricultural company by conducting data-mining, comparable company analysis and building predictive models on net profits of various product lines.
- **Workflow Standardization:** Created a generator that automatically imports raw data form Wind Database and outputs all key charts/figures (earnings, margins, cost split, etc.) given only the ticker of the firm.

## Education

---

### University of California, Los Angeles

Los Angeles, CA

*M.S. Analytics; GPA: 4.0/4.0*

*Expected Dec 2018*

- Courses: R, MySQL, Data Management, Optimization, Prescriptive Models, Types & Programming Languages

### National University of Singapore

Singapore

*Exchange Student, Statistics & Applied Probability; GPA: 5.0/5.0*

*Jul 2015 - Dec 2015*

- Courses: Regression Analysis (SAS, R), Applied Time Series Analysis (R), Stochastic Processes (R)
- Topics: ARIMA/SARIMA, Holt-Winters, Regression & Hypo Test, MCMC, M-H Algorithm, Gibbs Sampler, Filtering

### Central University of Finance and Economics

Beijing, China

*B.A. Management Science & Engineering (Real Estate); GPA: 92.6/100*

*Sept 2013 - June 2017*

- Highest Distinction (Rank: 2/185), 4 Full Marks on Math/Stats Courses

## Additional Information

---

Avid Fan of Functional Programming, Badminton Champions for Mixed Doubles, Aficionado of Chili & Spicy Food