

# Po-Yi (Bryce) Du

3750 E., Via Palomita., Apt. 1203, Tucson, AZ, 85718

+1 520 302 0069 · [pydu@email.arizona.edu](mailto:pydu@email.arizona.edu) · <https://www.linkedin.com/in/pydu-brys/> · <http://pydu-brys.github.io/>

Self-motivated data enthusiastic with 4-year data analytic/processing experience specializing Python/R/SQL/other Big Data tools, seeking full-time opportunity as data scientist.

## Education

**The University of Arizona, Eller College of Management; GPA: 3.77/4.0** Aug. 2016 – May, 2018

- Master of Science in Management Information Systems, Research Track

**National Cheng Kung University**

Sept. 2011 – Jun. 2015

- Bachelor of Business Administration in Statistics

## Professional Experience

**Research Engineer** | Research & Development | Metropia, Inc.

May 2018 – Present

### Incentive Engine Development

- Designed and implemented personalized incentive algorithm by quantifying BJ Fogg's Behavior Model
- Collaborated with software team to established data transferring pipeline and system designs
- Set up central databases and schemas/tables with AWS RDS and DynamoDB
- Designed statistical experiments to verifying the algorithm and calibrating parameters

### Bay Area Rapid Transit (BART) Perks Program; Prediction, Assignment, and Incentive Calculation

- Developed multi-source data collection and ETL procedures to acquire various features for modeling
- Utilized machine learning to predict crowdedness for links in BART with 95% accuracy
- Designed and implemented a Randomized Incremental Assignment algorithm with incentive calculation mechanism to recommend passengers different departure times, reducing severe crowding situation by 20%
- Collaborated with BART operation team to establish automated data transfer, preprocess, and update pipeline with AWS EC2 and S3 to ensure system robustness and sustainability
- Created and maintained program dashboard to monitor system performance using Periscope and Tableau

**Research Assistant** | Artificial Intelligence Laboratory | The University of Arizona

Aug. 2016 – May 2018

### DarkNet Marketplace (DNM) Analytics: High-impact Opioid Product Prediction

- Applied semi-supervised SVM classifier on identifying unlabeled opioid listings from collected DNM dataset
- Performed textual listing features engineering through TF-IDF and LSA for further modeling
- Applied GBM, SVM, Neural Network to predict high-impact opioid product, reaching an accuracy of 95%

### Hacker Web Data Collection Project (First-authored paper published at IEEE ISI 2018)

- Identified and crawled two major DNM's product listings and seller information on Tor using Python
- Extracted, transformed, and loaded 300k+ DNM html files into MySQL database

## Projects and Research Experience

**Social Media Opinion Mining for Nanotechnology Development (Supervised by Dr. Hsinchun Chen)**

- Collected and preprocessed 100k+ tweets through APIs, using Python and MySQL
- Clustered Nanotechnology articles' content into 13 subtopics through Latent Dirichlet Allocation
- Utilized term-frequency and polarity analysis to explored public opinions on Twitter toward the 13 subtopics

**Pokémon Respawn Prediction: PokémonGo Data Analysis**

- Generated Pokémon co-spawning features through network analysis, realized with MapReduce on AWS EMR
- Conducted exploratory data analysis through data visualization in R to determine important features
- Established Decision Tree model to predict next-spawning Pokémons with 90% accuracy

**Dialysis: Long-term Pattern Analysis**

- Manipulated and preprocessed Taiwan National Health Insurance Database (30Gb) using SAS
- Applied exploratory data analysis with patterns visualization on revealing research direction
- Identified high-risk individuals of renal failure with 94% accuracy through Lasso Logistic Regression

## Professions & Skills

Python	Expert in ML libraries, including pandas, numpy, sklearn, requests and BeautifulSoup
R	Proficient in statistical built-in functions and visualization libraries, e.g. ggplot2 and ggmap
Java	Fluent object-oriented system design and programming skills
Relational DBMS	Experienced in MySQL, Oracle SQL Developer, PostgreSQL, and MS SQL Server
Analytic Tools	SAS, Tableau, Periscope, SPSS, AWS (EC2, S3, EMR, RDS, DynamoDB), Gephi, Git, Linux, Kafka