

# Xiaohan (Aria) Wang

xiaohanwang2020@u.northwestern.edu | (424) 355-1425 | LinkedIn: linkedin.com/in/aria-wang | Website: ariawangxh.github.io/

## SUMMARY

---

A master's student in Analytics with a solid background in Math and Statistics. Possessing hands-on data science experience in multiple research and intern projects. A fast learner and a dedicated problem solver with effective communication skills.

## EDUCATION

---

**Northwestern University** Evanston, IL

**September 2019 – December 2020**

*Master of Science in Analytics*

- **Cumulative GPA: 3.95/4.00**

**University of California, Los Angeles (UCLA)** Los Angeles, CA

**September 2015 – June 2019**

*B.S. in Statistics and B.S. in Mathematics/Economics*

- Cumulative GPA: 3.86/4.00; **Major GPA: 3.92/4.00; Dean's Honors List of UCLA (8 out of 12 quarters)**

## SKILLS

---

- **Software:** Python, R, MySQL, PySpark, MapReduce, Git, AWS, Java, Tableau, D3.js, ArcGIS, SAS
- **Data Science:** XGBoost, Random Forest, SVM, Lasso, Linear and Logistic Regression, Data Mining, Deep Learning, Natural Language Processing (NLP), Time Series, A/B Testing, Data Visualization
- **Language:** Native Speaker of Chinese (Mandarin), Fluent English

## RESEARCH & INTERNSHIP EXPERIENCES

---

**Spiegel Research Center, Northwestern University** Evanston, IL

**June 2020 – August 2020**

*Research Assistant*

- Built end-to-end machine learning solutions for predicting customer churn and win-back for WEHCO Media. Developed the data cleaning and feature engineering pipeline in PySpark. Achieved high prediction accuracy (AUC 0.80) using logistic regression models in R
- Extracted patterns in the reading engagement behaviors of WEHCO customers, and provided actionable insights for supporting WEHCO to shift from advertising-based strategy to reader-based revenue models
- Provided thorough evaluation on the effectiveness of newsletters and pricing strategies to improve customer retention and lifetime value. Supported WEHCO's efforts in increasing customer stickiness by investigating the reliability of current engagement metrics

**Acumen, LLC** Burlingame, CA

**June 2018 – September 2018**

*Statistical Programmer Intern*

- Improved the definitions of control and risk windows for FDA's real-time influenza surveillance on a rare syndrome by analyzing its presence in diagnoses from historical Medicare claims for the past 2 decades using SAS
- Optimized the prediction of beneficiaries' choices of pharmacy chains for prescription refills by identifying key demographic factors influencing beneficiaries' decisions. Improved the classification accuracy by 17% by implementing Random Forest models

## SELECTED PROJECTS

---

**Everybody Eats**, the City of Evanston

**June 2020 – September 2020**

- Identified influential factors in local food insecurity rates in Cook County by developing regression and decision tree models on data collected from US Census, Feeding America and Food Access Research Atlas
- Implemented kepler.gl maps to visualize the vulnerable community and the active charitable food assets in Evanston
- Communicated the insights with the Evanston food insecurity task force to facilitate transparent and informed cooperation amongst partner agencies, and to balance the food supply and demand for the food insecure community

**Mask On – Face Mask Detection**, Northwestern University

**April 2020 – June 2020**

- Built a face mask detection model with Convolutional Neural Network (VGG16). Reached 82% accuracy in recognizing whether the image subjects wear face masks and the type of masks they wear
- Developed a real-time face mask detection tool with Labelling that outputs bounding boxes around masks on image input. Facilitated close monitoring of the local mask rate for public alert during the COVID-19 pandemic

**Open-Source Marketing Intelligence Repository**, HSBC Bank

**October 2019 – June 2020**

- Created an open-source SQL data repository that can be leveraged for existing revenue analysis within HSBC
- Detected market dynamics from geospatial data with Lasso and Random Forest models. Identified groups of branches with consolidation/closure potentials using K-means clustering. Created ArcGIS visualization dashboard

**Predicting "Match" in a Speed Dating Experiment**, Northwestern University

**January 2020 – March 2020**

- Built multiple machine learning models (Support Vector Machine, Neural Network and XGBoost) to predict the matching results with 74% sensitivity rate for a speed dating experiment conducted at Columbia University in 2004
- Supported future studies about the disparity between people's stated interests and actual preferences of their partners in dating. Provided online dating apps with suggestions on increasing matching rates for different groups of users