

# YIFEI WANG

Master student in Interdisciplinary Data Science at Duke University

@ yifei.wang828@duke.edu

☎ 984-245-5566

in linkedin.com/in/yifei-wang-0249b5170/

📄 github.com/ywang512

## EDUCATION

**Duke University** Durham, NC

*Master of Interdisciplinary Data Science*

📅 Aug 2018 – May 2020 🎓 3.9/4.0

**University of California, Berkeley** Berkeley, CA

*Exchange Student*

📅 Aug 2016 - Dec 2016

**Sun Yat-sen University** Guangzhou, China

*Bachelor of Science in Mathematics*

📅 Aug 2014 – June 2018 🎓 3.7/4.0

## EXPERIENCE

**Hearful Technologies, Inc.**

*Data Scientist Intern*

📅 May 2019 – Aug 2019 📍 Chapel Hill, N.C.

- Researched extensively on Aspect-based sentiment analysis for online reviews.
- Deployed a coherence topic model for the simultaneously discovery of latent aspects and associated sentiment polarity in python from scratch.
- Implemented tools to thoroughly analyze model outcomes and provided instant and deep insights into a new domain.
- Achieved automatic aspect extraction and sentiment polarity assignment with 0.82 F1-score on hand-labeling dataset.

**National Supercomputer Center in Guangzhou**

*Research Assistant*

📅 Oct 2017 – Jan 2018 📍 Guangzhou, China

- Contributed to “Translingual Literature Recommendation System” under Prof. Lu, Yao.
- Implemented previous algorithm, including RNN, node2vec and metapath2vec, using python and TensorFlow.
- Undertook literature text mining to extract key characteristics and to store them using MongoDB.
- Generated a large heterogeneous network of multilingual papers, authors and venues using extracted data.
- Operated data and ran scripts on Tianhe-2, a supercomputer located in National Supercomputer Center in Guangzhou.

## HONORS & AWARDS

- Winner of HL7 FHIR DevDays Student Track at Seattle for an Asthma Management App on IOS *June 2019*
- “Best Use of Outside Data” Prize at Data Fest at Duke (one of best 4 teams among 82 teams) *Apr 2019*
- 7th prize of Duke Datathon among 200+ students *Nov 2018*

## SKILLS

- **Programming Languages:**  
Python, C/C++, R, Shell script, MATLAB
- **Database:**  
SQL, MongoDB, Hadoop, Apache Spark
- **Deep Learning:**  
TensorFlow, keras, PyTorch
- **Cloud Computing:**  
Continuous Integration, AWS Cloud9, AWS Lambda
- **Visuals:**  
Tableau, Markdown, HTML

## PROJECTS

**Capstone Project: Image Features Extraction for Auction Price Prediction**

📅 Aug 2019 - Now

- Coordinated regular meetings with stakeholders following Agile project management.
- Extracted three interpretable and meaningful features from the auction images.
- Implemented an ETL pipeline for image feature extraction and a price prediction pipeline with both auction data and extracted features.

**Solar Panels Recognition from Satellite Images**

- Applied several image preprocessing techniques, including RGB rescaling, image augmentation, image gradients and relative luminance.
- Implemented several statistical learning models and a Convolution Neural Network model with accuracy 98.54%.

**Real-time Nascar Racing Data Analysis and Visualization**

- Implemented comprehensive and interactive visualizations of real-time refreshing racing data using R Shiny.
- Incorporated heterogeneous historical racing data to help engineers make optimal decision on pit stop strategies.

**New York Housing Price and Transportation Analysis**

- Quantitatively analyzed the effect of different transportation accessibility (Taxi, Uber, subway, bike, etc.) on housing price in New York.
- Conducted prediction on New York housing price data using various machine learning models and significantly raised prediction accuracy after incorporating feature engineered transportation data. ( $0.61 R^2$ )