

# Diane (Yian) Ding

Boston, MA 02215 | [ding.yia@northeastern.edu](mailto:ding.yia@northeastern.edu) | (413)210-0348 | [Website](#) | [LinkedIn](#) | [GitHub](#)

## Education

**M.A. in Data Science, Khoury College of Computer Science, MA** Sep. 2020  
**Northeastern University, GPA: 4.00** Expected Graduation: Sep. 2022  
Relevant Coursework: Supervised Machine Learning, Unsupervised Machine Learning, Database Management, Data Processing  
**B.A. in Data Science & Psychology, Mount Holyoke College, MA GPA: 3.82** Jan. 2016 - Jan. 2020  
Relevant Coursework: Natural Language Processing, Applied Regression Methods, Artificial Intelligence, Scientific Computing, Advanced Java, Data Structures, Computing Systems, Probability  
Honors: **Magna Cum Laude**, **Mary Lyon Scholar**, Sarah Williston Prize (top 3%), Academic Excellence Award in **Psychology**, Academic Achievement Award (top 5%)

## Skills Summary

Languages: **Python, R, SQL**, Java, MATLAB, C, JavaScript, HTML, CSS/SCSS  
Libraries: Pandas, Scikit-Learn, TensorFlow, Keras, Plotly, Matplotlib, Numpy, Pytorch, OpenCV  
Methods: Data Mining, Hypothesis testing, Classification, Clustering, Regression Analysis, A/B test, NLP, EDA, Time-series  
Technologies: RStudio, Postgre, Jupyter, Tableau, Power BI

## Experience

**Data Scientist Intern, Gap Inc., Shanghai, China** Jun. 2019 - Aug. 2019

- Performed data mining; utilized web-crawler to compile pricing data for competitive analysis using **Python** and **SQL**
- Conducted similarity test of products using **KNN** and **ANOVA** test; built a pricing model to dynamically adjust prices
- Created visualizations using **Tableau**; narrowed price gap by **21%** and increased quarterly revenue by **~13%**

**Research Assistant, Interactive Computing Research Lab, Mount Holyoke College, MA** Sep. 2018 - May 2019

- Tested, modified and productionized a programmable humanoid robot API to support children-robot interaction system
- Monitored and maintained the system for **4 months**; mentored and collaborated with 3 volunteers on using the system and completed **64 runs**; examined interaction data for early troubleshooting; improved overall interaction rate by **32%**

**Research Assistant, NLP Lab, Soochow University, China** Jun. 2018 - Aug. 2018

- Programmed scripts for face recognition and face tracking using Python (**OpenCV** and **TensorFlow**)
- Engineered and tested FaceSpy (Facial recognition Application) for community security to make sure FaceSpy works error-free; reduced classification error rate based on daily active test data (Accuracy **~90%**) on a daily basis

**Research Assistant, Cognition Attention Perception Speech Lab, Mount Holyoke College, MA** May 2017 - Jul. 2018

Project: The Cat in the Hat

- Programmed scripts for speech-text forced alignment using **Praat** and **Prosody-Aligner**; reached 98.6% accuracy
- Iteratively fitted control and metric models for word duration and inter-onset interval using **R**; the resulted model revealed unprecedented finding about hierarchical meter; related paper was **accepted** by **Cognition** at 2018

Project: Speech-to-sound illusion

- Assisted Professor in extracting and analyzing human pitches data using linear **regression** and **autocorrelation**
- Initiated and implemented a beat tracking model using **Java**, saving **~60%** of processing time

## Projects

**Mechanisms of Action (MOA) Prediction** Sep. 2020 - Nov. 2020

- Optimized and applied MLSTMOTE algorithm using **Python** to address class imbalance; built pipelines for consistent results
- Accomplished model Optimization using **PCA**, **Stacking** and **Blending**; Led a group of 4 and achieved top 1% submissions

**Android Development: Vocabulary Memorizer** Jan. 2020 - Mar. 2020

- Designed and developed Vocabulary Memorizer using **Android Studio**, utilized **Amazon DynamoDB** to store App data
- Implemented a spaced repetition algorithm (SM-2); improved user's quiz scores by **65%** over 3 months

**Natural Language Processing: Sentiments detection in Yelp reviews** Aug. 2019 - Dec. 2019

- Researched and conducted **Support Vector Machines** to classify reviews into categories
- Developed **RNN** and **CNN** using **Pytorch** to detect sentiments of reviews, the later achieved 95.46% accuracy

**Suicide Prediction and Prevention with Big Data: A Worldwide View** Aug. 2019 - Dec. 2019

- Identified and critiqued latent factors in suicide prediction through literature review and factor analysis
- Designed a **Random Forest** model to predict suicide rate with 92% accuracy, used K-means clustering to find country clustering and inspected the results to provide suggestions for national prevention methods
- Presented results at college-level seminar and to local communities; received department high honor

## Honors & Awards

Mathematical Contest in Modeling 2019, Honorable Mentions (10%) 2019  
Prize for Excellence in Mathematics and Computer Science, MHC Department of Mathematics 2017