

# Yumiao (Jenny) Hui

3224 Dunwoodie Rd, Ann Arbor, MI, 48105 | 734-881-8445 | huiym@umich.edu

## EDUCATIONAL BACKGROUND

|  |                 |
|--|-----------------|
| <b>University of Michigan</b>  | 08.2021-05.2023 |
| Applied Statistics, Master of Science  |                 |
| <b>GPA: 3.76/4.0</b>   |                 |
| <b>Sun Yat-sen University, China</b>   | 09.2016-07.2020 |
| Information and Computing Science, Bachelor of Science   |                 |
| <b>GPA: 3.7/4.0</b>  |                 |
| <b>Major Courses:</b> Computational Methods and Tools in Statistics, Machine Learning and Data Mining, Computer Network, Data Structures and Algorithms, Operating System, Statistical Inference, Database Application Design, Statistical Learning II: Multivariate Analysis, Cloud Computing |                 |
| <b>Relevant Skills:</b> C++, Python (PyTorch, TensorFlow, Keras, scikit-learn, ggplot2), MATLAB, R, SPSS, Java, MySQL, Tableau, Hadoop, Jupyter Notebook, AWS tools, Django, Machine learning  |                 |
| <b>Student Summer Program: Oxford Prospects Program</b> , at University of Oxford, UK  | 07.2018-08.2018 |

## INTERNSHIP EXPERIENCES

|   |                 |
|---|-----------------|
| Data Internship, <b>Edugion</b> , U. S  | 12.2021-Current |
| Summarized and updated worldwide stock exchange information in website database collaborating with colleagues                 |                 |
| Visualized trade associations with industry information on a world map in Tableau and Power Map                               |                 |
| Conducted data cleaning on 50000 companies' information on nice formatting and missing values                                 |                 |
| Web & IT Services Internship, <b>Capgemini</b> , China  | 04.2021-07.2021 |
| Created and executed several AWS ETL jobs with AWS Glue and applied Kinesis and SQL to analyze stream 5GB data                |                 |
| Assisted with database management and provided comprehensive IT Desktop operation & maintenance services                      |                 |
| Part-Time Assistant, <b>Boston Consulting Group</b> , Remote, China   | 07.2019-09.2019 |
| Attended industry research on digital asset custody: using python crawler to collect keywords from over 1000 webpages         |                 |
| Estimated the proportion of a specific future product with linear regression model based on the processed data                |                 |
| Implemented a field survey included cold calls, gathering fresh first-hand information over 80 beer products                  |                 |
| Made basic analysis and created slides to illustrate the insights from market proportions of beer brands proportions in local |                 |
| Data analyst Internship, <b>Ernst &amp; Young Global Limited</b> , China  | 01.2019-03.2019 |
| Built Journal Entry and Trial Balance forms to check the accuracy of structured data from clients                             |                 |
| Used data visualization tools like Tableau to identify key issues and opportunities in due diligence & valuation              |                 |
| Analyzed reports of target international companies in Japanese and summarized key figures and metrics for team                |                 |

## RESEARCH EXPERIENCES

|  |                 |
|--|-----------------|
| Membership, <b>Query-Based Recommendation by HIN Embedding with PRE-LSTM</b>   | 09.2019-11.2020 |
| <i>Dissertation, accepted by ADMA2020, published in Lecture Notes in Artificial Intelligence (LNAI) by Springer</i>  |                 |
| Used pre-trained word embeddings and LSTM to train network embeddings and learned the relationship among the tables, users, and keywords                                       |                 |
| Transferred keywords, tables, and users into Heterogeneous Information Network and adopted Random Walk method to generate meta-paths to characterize semantic meanings for HIN |                 |
| Made the query-based recommendation for keywords from users and showed user-specific recommendation tables sorted in a reasonable order  |                 |
| Membership, <b>Combined Density Peak and K-means Clustering on Image Segmentation</b>  | 03.2018-07.2018 |
| <i>Working Paper</i>   |                 |
| Practiced various kinds of clustering methods like K-means and Density Peak to conduct image segmentation  |                 |
| Combined K-means and Density Peak Clustering together and made improvements on running time  |                 |

## PROJECT EXPERIENCES

|  |                 |
|--|-----------------|
| Group Project, <b>Hotel Booking Demand Analysis from Kaggle</b>  | 11.2021-12.2021 |
| Conducted data manipulation and exploratory data analysis on dataset after cleaning from the raw data with R   |                 |
| Applied statistical tools like Bootstrap/Permutation Test and several machine learning algorithms such as KNN, SVM, Random Forest, Logistic Regression, Naïve Bayes Classifier on prediction of hotel booking cancellation |                 |
| Independent Project, <b>Database Application Design</b>  | 01.2022-04.2022 |
| Built several Create, Retrieve, Update and Delete Applications with Log-in page on Website with Django framework   |                 |
| Designed a classified ad website that allowed GitHub accounts with creating, presenting, editing, and searching ads  |                 |
| Independent Project, <b>Biostatistical Analysis on MIMIC Dataset</b>   | 05.2018-07.2018 |
| Performed various SQL tasks like creating new tables, making joins, and conducting data manipulation on the dataset  |                 |
| Made statistical inferences and exploratory data analysis on several related biostatistics' metrics with Python & Tableau  |                 |