# Wanyue Xiao

xwanyue@syr.edu | +1-(315)278-0270 | www.linkedin.com/in/wanyue-xiao-60688196

## Education

**MS of Applied Data Science, Syracuse University,** School of Information Studies **(3.91/4.0)** **July 2019 - May 2021**

Coursework: Natural Language Processing and Deep Learning, Text Mining, Big Data Analytics, Data Visualization, Advanced Database Management, Business Analytics, Data analysis & Decision Making

**Bachelor of Management, Shanghai Ocean University,** AIEN Institute **(3.52/4.0)** **July 2015 - June 2019**

Coursework: Accounting and Financial Decision Making; Requirements Analysis and Modeling; System Acquisition and Implementation Management; Business and Information Analysis

**Bachelor of Information Systems, University of Tasmania,** School of Engineering **(6.89/7.0)** **July 2015 - June 2019**

Coursework: Data Management; Database Management Systems; Programming and Problem Solving; Dynamic Website Development; IS Project Management; ICT Project Management

## Technical Skills

| | |
|---|---|
| **Data Analysis:** | Python (scikit-learn, NumPy, pandas, gensim, nltk), R, Tableau, Spark, Probit & Logit, Excel (Lookup, Regression, Forecasting, Power Query, Pivot, Solver optimization), Google Analytics, ANOVA |
| **Database Development:** | MySQL, NoSQL, Oracle SQL, Hadoop Ecosystem (Hive, MapReduce, etc.), MongonDB |
| **Applications & System:** | MS Project, MS Visio, MS Access, Linux, Docker, Databricks |
| **Others:** | NLP (Topic Modeling, Word2Vec, Text Mining), Javascript, Java, HTML + CSS +PHP+Django |

## Professional Experience

**Database Developer | iConsult Collaborative at Syracuse University** (Syracuse, U.S.A.) **February 2020 - Now**

- Worked as part of a development team and established close working relationships with Eastside Soccer's stakeholders.
- Agilely worked with the website developers and management team to design, develop, document, test, and implement data models, database architecture, and DBMS solutions.

**Data Analyst Intern | Shouye Big Data Science & Technology Co., Ltd.** (Beijing, China) **July 2018 – August 2018**

- Implemented a spider that crawled 1.5 million rows of daily news from SINA, including title, date, content, click, tokenized comments. Used Python packages (urlib, RE, lxml) to process data and then store the data into MongoDB.
- Used word2vec to vectorize user comments to identify the popularity rate of each new topic.
- Analysis of One Belt One Road Project: Collected Transactions Data of 101 countries from 2013 to 2017. Cleaned around 20 Gb trade transaction data via Excel: VLOOKUP, Correlation, Linear & Multi-linear Regression, Pivot Table.

**Assistant Intern | Starbucks** (Shanghai, China) **July 2017 – August 2017**

- Promoted product and recommended activity to customers through maintaining WeChat Official Account.
- Collaborated with team members to produce weekly reports which recorded and summarized sales.
- Communicated with customers with different languages; understand customers' requirements.

## Academic Projects

**Flight Customer Satisfaction Enhancement** (Syracuse, U.S.A.) **November 2019 – December 2019**

- Manipulated JSON files and adopted Regression, Random Forest, SVM, and Association Rules etc. for building customer classification model, which achieved 87.27% accuracy rate.
- Used Logistic Regression to predict the possibility that if a customer will become a loyalty member
- Designed an A/B testing to evaluate customer classification model's performance.

**Web-based Shopping Forum Establishment** (Syracuse, U.S.A.) **August 2019 – September 2019**

- Constructed a database by using MySQL to store user information and via NoSQL to store users' reviews.
- Developed complex SQL scripts to query data across multiple databases and feed the website and APIs.

**Social Media Sentiment Analysis based on COVID-19** (Syracuse, U.S.A.) **November 2020 – December 2020**

- Established a transfer deep learning model using LSTM and pre-trained BERT model to quantify the sentiment and a LDA model to identify the COVID-19 related topics based on twitter posts published between March 29, 2020 to April 30, 2020.
- Used various algorithms, including Random Forest, XGBoost, etc. to verify the performance of the proposed classifier.

## Publication

Sang, X., Xiao, W., Zheng, H., Yang, Y. and Liu, T., (2020). HMMPred: Accurate Prediction of DNA-Binding Proteins Based on HMM Profiles and XGBoost Feature Selection. *Computational and Mathematical Methods in Medicine*, *2020*.

Wang, J., Zheng, H., Yang, Y., Xiao, W. and Liu, T., (2020). PredDBP-Stack: Prediction of DNA-Binding Proteins from HMM Profiles using a Stacked Ensemble Method. *BioMed Research International*, *2020*.

Yang, Y., Zheng, H., Wang, C., Xiao, W., & Liu, T. (2019). Predicting Apoptosis Protein Subcellular Locations based on the Protein Overlapping Property Matrix and Tri-Gram Encoding. *International journal of molecular sciences*, *20*(9), 2344.

## Leadership and Campus Involvement

**Team Leader** **April 2018 – May 2018**

- Lead 5-person Group, *2018 SAP ERP Simulation Game Competition* (The Recognition Award)

**Department Leader and Fund Raiser (**Voluntary Activity Department) **September 2015 – July 2016**

- Organizing "Campus Annual Blood Donation Voluntary Activity"; Raising Funds from local businesses

**eBoard Leader** (iBranch Student Association) **September 2019 – Now**