Liyang Xue

PhD candidate in Information Science liyang.xue@rutgers.edu
+12065198336 https://www.linkedin.com/in/liyangxue/

EDUCATION

Rutgers University, School of Communication and Information, New Brunswick, NJ **September 2021 – Present**
Ph.D. Candidate in Library and Information Science
Syracuse University, Martin J. Whitman School of Management, Syracuse, NY **August 2019 – May 2021**
M.S. Business Analytics | GPA 3.9/4.0
University of Washington, College of Art & Science, Seattle, WA **September 2014 - June 2019**
B.S. Physics, Minor: Mathematics | GPA 3.1/4.0

RESEARCH INTERESTS

- Current Interests: Algorithmic Fairness and Health Informatics
- General Interests: Machine Learning, Health Data Science, Responsible AI, LLMs

SKILLS

- Programming: Python, R, Java, SQL, MATLAB.
- Tools: EXCEL, ACCESS, Power BI, Tableau, Illustrator, Google Analytics, VISIO, SQL Server, Mini Tab, Visual Studio, SolidWorks
- Industry Knowledge: Machine Learning, Data Analysis, Database Management, Lean Six Sigma, Data Warehouse, LLMs

PUBLICATIONS

- Liyang Xue, A M Muntasir Rahman, Charles R Senteio, Vivek K Singh, Automated detection of stigmatizing language in Electronic Health Records (EHRs) using a multi-stage transfer learning approach, Journal of the American Medical Informatics Association, 2025;, ocaf193, https://doi.org/10.1093/jamia/ocaf193
- Ahmed, E., Xue, L., Sankalp, A., Kong, H., Matos, A., Silenzio, V., & Singh, V. K. (2023a). Predicting loneliness through digital footprints on Google and YouTube. Electronics, 12(23), 4821. https://doi.org/10.3390/electronics12234821
- WIP: Documenting Equity: Creating and Modeling a Dataset of Stigmatizing Language Toward Gender-Expansive Patients in EHRs

EXPERIENCE

Research Assistant - With Dr. Vivek Singh, Behavioral Informatics Lab, Rutgers University **September 2021 – Present**

- Investigating stigmatizing language in electronic health records (EHRs), with emphasis on linguistic harms affecting gender-expansive patients. Fine-tuned LLMs (Longformer, BERT, ClinicalBERT) using a multi-stage transfer learning framework to detect stigmatizing language in long clinical narratives (89% accuracy) and systematically compared performance and fairness outcomes against GPT-4o using zero-shot and few-shot prompting strategies.
- Led the development of annotation guidelines to identify stigma subtypes, including misgendering, ensuring reliable labeling for gender-expansive populations and supporting fairness-aware evaluation across demographic subgroups.
- Contributing to ongoing research on digital well-being and loneliness prediction by building machine learning models, including random forest, XGBoost, Logistic Regression, and SVM that integrate digital traces and demographic information.
- Assisted in writing a grant proposal on social work regulation in the U.S. by collecting and comparing social worker data from each state.
- Analyzed people's reactions on Facebook to COVID-19 news using Python, LDA, and data visualization techniques to determine when people focus more on COVID-19-related news.

Data Analyst Intern - Syracuse University, Whitman School of Management **August 2020 – December 2020**

- Collected thousands of data points from past alums to predict their wealth by developing an automated web scraping program using Beautiful Soup and Selenium in Python.
- Monitored and maintained the web scrapers to work reliably.
- Delivered the datasets and presented the key insights to the Business School dean's office.

Lean Six Sigma Consultant, Indium Corporation **August 2020 – January 2021**

- Initiated a lean project within the facilities to increase the RTP inventory accuracy from 77% to 98%.
- Analyzed the data and provided insightful recommendations for the RTP inventory control process.
- Conducted a pilot test on the proposed solutions and fixed some problems during the pilot test.
- Obtained Lean Six Sigma green belt after this project.

TEACHING

Teaching Assistant - ITI 220, Rutgers University **September 2024 – December 2024**
Course Title: Data in Context

Teaching Assistant - MI 557, Rutgers University **September 2024 – December 2024**
Course Title: Database Design & Management
Lecturer - MI 561, Rutgers University **January 2024 – June 2024**
Course Title: Data Analytics
Teaching Assistant - MI 562, Rutgers University **September 2023 – May 2024**
Course Title: Problem Solving with Data
Teaching Assistant - ITI 202, Rutgers University **September 2021 – May 2024**
Course Title: Object-Oriented Programming
Teaching Assistant - MAR 301, Syracuse University **August 2020 – December 2021**
Course Title: Essentials of Marketing

PROJECTS

Data Analyst, Amazon Review Helpfulness Classification **January 2021 - May 2021**
• Developed a topic modeling system with the LDA algorithm to understand the contents of reviews.
• Combined LDA scores and random forest to determine if the review is helpful to other users.
• Optimized model performance by grid searching for optimal topic numbers and evaluated them using a confusion matrix.

Data Analyst, Breast Cancer Diagnosis Classification **August 2020 - December 2020**
• Developed an SVM and K-means clustering model to identify the diagnosis result with over 90% accuracy.
• Compared the performance of the decision tree, XGB, Random Forest, and MLP models based on recall evaluation.

Data Analyst, Football Player Data Analysis **January 2020 - May 2020**
- Developed and tuned predictive models for players' overall scores using advanced machine learning techniques, such as random forest, linear, GBT, and TensorFlow Keras sequential model in Python.
- Designed and implemented a robust recommender system that utilized principal component analysis and K-means clustering to provide tailored player recommendations to club managers.
- Utilized various neural network models to accurately predict players' positions based on stats with high precision, achieving over 85% accuracy.

Data Analyst, Marketing Case Study on Orange Juice Brands **March 2020 - May 2020**
- Applied principal component analysis with factor rotation and k-mean clustering to study the relationship between customers and orange juice brands.
- Developed a linear model, identified critical factors that impact the demand for each brand, and conducted hypothesis testing to interpret the marginal effect of price on demand.

Data Analyst, Customer Churn Analysis in the Airline Industry **September 2019 - December 2019**
- Improved predictive model accuracy by performing data quality checks and resolving data quality issues.
- Discovered crucial factors that contribute to customer churn by mining association rules.
- Classified customers into detractors, passive, and promoters using the Support Vector Machines (SVM) algorithm.

Database Engineer, Database Design for Global Consolidate Shipping Company **September 2019 - December 2019**
- Designed a database to support customers and the company and implemented it using MS SQL Server. Designed the user interface using ACCESS.
- Wrote queries to solve business problems and created views for easy access.

HONORS

- Mlis And Phd Student Support Fund at Rutgers University
- SC&I Scholarship Fund - LIS Student Support Fund Award at Rutgers University
- Tefko Endowed Fund for Doctoral Students Award at Rutgers University
- Merit-based Scholarship at Syracuse University
- Multiple Dean’s List recognitions at the University of Washington