

YI ZHU

Curriculum Vitae

zhu00331@umn.edu | 612-757-2450 | <https://carlsonschool.umn.edu/faculty/yizhu>

EDUCATIONS

- **Ph.D. in Information and Decision Sciences** Expected 2024
University of Minnesota at Twin Cities, Carlson School of Management, GPA: 3.72 / 4
 - **M.S. in Business Analytics** March 2019
Drexel University, LeBow College of Business, GPA: 3.97 / 4
 - **M.S. in Journalism** December 2015
University of Illinois at Urbana-Champaign, College of Media, GPA: 3.84 / 4
 - **B.A. in English** July 2014
Communication University of China, GPA: 3.38 / 4
-

RESEARCH INTERESTS

- **Areas**
Healthcare IT and Management, Financial Technology, IT Innovation, Digital Platform, Social Media, Personalization
 - **Methods**
Machine Learning, Deep Learning, Natural Language Processing, Big Data Analytics, Econometrics Modeling, Causal Inference
-

JOURNAL PUBLICATIONS

- Karaca-Mandic, P., Sen, S., Georgiou, A., **Zhu, Y.**, & Basu, A. (2020). Association of COVID-19-Related Hospital Use and Overall COVID-19 Mortality in the USA. *Journal of General Internal Medicine*.
 - Zhou, J., Everhart, A., Smith, L. B., Qiu, Y., & **Zhu, Y.** (2020). The Economic Burden of Extended-Release Pharmaceuticals in Minnesota. *Public Health Review*.
-

CONFERENCE AND WORKSHOP PAPERS

- **Zhu, Y.**, Chan, J., Bi, X., Guo, Y. & Wu, J. (2021). Investigating the Willingness to Pay for Enhanced Mobile Internet Services: Evidence from A Mobile Network Upgrade
 - 2021 Statistical Conference in E-Commerce Research (SCECR). Virtual.
 - 2021 American Marketing Association (AMA) Summer Conference. Virtual: American Marketing Association (AMA).
 - 2021 43rd Annual INFORMS Society for Marketing Science (ISMS) Conference. Virtual: Institute for Operations Research and the Management Sciences (INFORMS). Session Chair.
 - 2021 Industry Studies Association Annual Conference. Virtual: Industry Studies Association (ISA).

- **Zhu, Y.**, Everhart, A., Karaca-Mandic, P., & Sen, S. (2020). Using NLP to Extract Predicate History from Medical Device Approvals. International Conference on Information Systems (ICIS). Hyderabad, India: Association for Information Systems (AIS).
 - **Zhu, Y.**, Sen, S., Karaca-Mandic, P., Georgiou, A., & Vu, K. (2020). An Interactive Web-Based Dashboard for COVID-19 Hospitalizations in the U.S. Workshop on Information Technologies and Systems (WITS). Hyderabad, India.
 - Fang, J., Liu, C., & **Zhu, Y.** (2018). Sequence Clustering with Temporal Graph and Sequence Recovery. INFORMS Workshop on Data Science. Phoenix, AZ: Institute for Operations Research and the Management Sciences (INFORMS).
-

PAPERS UNDER REVIEW / WORKS-IN-PROGRESS

- Under review: Racial Disparities in COVID-19 Hospitalization and Inpatient Bed Occupation: Analysis Based on Healthcare Markets (Coauthored with Lu Li)
 - Under review: Rural versus Urban Trends in COVID-19 Hospitalizations: A County-Level Analysis of 6 U.S. States (Coauthored with Caitlin Carroll, Khoa Vu, Soumya Sen, Archelle Georgiou, Pinar Karaca-Mandic)
 - In progress: Investigating the Willingness to Pay for Enhanced Mobile Internet Services: Evidence from A Mobile Network Upgrade
 - In progress: Improving Sustained Activities Forecasting Accuracy: A Deep Learning Approach with Cross-Activity Relationship Awareness
 - In progress: Leveraging Data Science to Inform Regulation and Improve Cardiovascular Device Safety
-

OTHER RESEARCH CONTRIBUTIONS

- **COVID-19 Hospitalization Tracking Project**
 - This project tracks COVID-19 hospitalizations data to estimate the COVID-19's current impact on local hospital systems and model future hospital utilization needs. Data are collected from the department of health website of each U.S. state
 - This project has won numerous media exposures and academic/business awards. Projects details can be found at <https://carlsonschool.umn.edu/mili-misrc-covid19-tracking-project>
 - Extensively used Tableau to build and launch the COVID-19 hospitalization tracking online dashboard for data visualization.
 - Contributed to the following publications by building and launching the online dashboard and conducting data analysis:
 1. Karaca-Mandic, P., Georgiou, A., & Sen, S. (2020). Assessment of COVID-19 Hospitalizations by Race/Ethnicity in 12 States. JAMA Internal Medicine.
 2. Levin, Z., Choyke, K., Geogiou, A., Sen, S., & Karaca-Mandic, P. (2021). Trends in Pediatric Hospitalizations for Coronavirus Disease 2019. JAMA Pediatrics.
 3. Sen, S., Karaca-Mandic, P., & Georgiou, A. (2020). Association of Stay-at-Home Orders With COVID-19 Hospitalizations in 4 States. JAMA, 323(24):2522-2524.
-

REVIEWER

- 2021 International Conference on Information Systems (ICIS)
 - 2021 American Marketing Association (AMA) Summer Conference
-

TEACHING ASSISTANCE

- MSBA 6130: Introduction to Business Analytics in R (Fall 2020)
 - MSBA 6440: Causal Inference via Econometrics and Experiments (Spring 2020)
 - IDSC 6050: Information Technologies and Solutions (Fall 2019)
-

AWARDS AND GRANDS

- Carlson School of Management Ph.D. Student Conference/Travel Fellowship (\$355) April 2021
Carlson School of Management, University of Minnesota
 - Champion of the first UMN Interdisciplinary Health Data Competition (\$4,000) April 2020
Carlson School of Management, University of Minnesota
 - BOLD IDEAS Grant (\$30,000) February 2020
Office of Academic Clinical Affairs, University of Minnesota
 - Ph.D. Fellowship 2019 – 2024
Carlson School of Management, University of Minnesota
 - Teck-Kah Lim Graduate Student Domestic Travel Subsidy Award (\$300) October 2018
Graduate College, Drexel University
 - Donald T. Sheehan International Funding September 2017
The Wharton School, University of Pennsylvania
 - LeBow Alumni Merit Scholarship 2017 – 2019
LeBow College of Business, Drexel University
 - Marilyn Miller Kaytor Scholarship 2014 – 2015
College of Media, University of Illinois at Urbana-Champaign
-

EMPLOYMENTS

- **University of Minnesota at Twin Cities** Minneapolis, MN
Research Assistant September 2019 – Present
Selected Analytical Experiences:
 - Applied NLP techniques (bag of words, n-gram, rule-based search) to extract the predicate history of medical devices from more than 50,000 new device approval filings.
 - Applied NLP techniques (word embeddings, sentiment analysis) to classify and compute sentiment for more than 600,000 customer complaints on mobile network services.
 - Applied deep learning (CNN, LSTM) to extract features from physical activity sensor data and build up model to forecast user adherence to running exercises.
 - Applied econometrics modeling and causal inference techniques to evaluate the impact of mobile network upgrade on consumer reactions and estimate the association of COVID-19-related hospital use and overall COVID-19 mortality in the U.S.
- **Drexel University** Philadelphia, PA

Research Assistant

March 2018 – March 2019

Selected Analytical Experiences:

- Conducted topic analysis with LDA model on more than 3000 company location selection criteria articles to identify key elements in considering startup locations.
- Analyzed more than 200,000 music critics reviews' evolvement across 18 years with econometric modeling to evaluate relationships between album sales and insights of the reviews.
- Computed Temporal Graph based on uni-gram and PrefixSpan on 540 thousand rows of e-commerce data to perform sequence clustering for better customer segmentation.
- Built customer segmentation methodology based on customers' gas usages and their usage responses to temperature change. Validated the cutting thresholds with K-shape method.
- Forecasted potential excess funds in more than 60 thousand checking accounts across multiple time windows with Holt-Winters Exponential Smoothing model.

• **Tencent (HK: 0700) Tencent News Securities Channel** Beijing, China
Financial Analyst and Writer (freelance after September 2017) March 2017 – August 2019

- Wrote analytical articles on U.S. financial markets to provide investors with investment guidance. Major focuses were on financial statements, asset price movement, economic policy, etc.
- Monitored mainstream U.S. financial media reports on Wall Street analysts' comments, U.S. public companies, Fed's monetary policy, etc. Translated valuable content into Chinese for reposting.
- Investigated market sentiment and generated reports every week with online questionnaires. Finished a featured report on the 2017 Berkshire Hathaway Annual Shareholders Meeting.

• **Phoenix New Media (NYSE: FENG) Finance Channel** New York, NY
Financial Correspondent May 2016 – December 2016

- Hunted for exclusive stories on Wall Street. Wrote analytical articles and instant reports on market movement, economic data, and policy by monitoring and analyzing market data.
- Conducted face-to-face interviews with experts in the financial industry to develop analytical stories. Attended industry events and meetings to expand connections and story sources.
- Covered remotely or on the spot the major headlines in 2016, including RMB depreciation, Fed rate hike, Brexit, IMF, and WB annual meeting, U.S. stock market rally, U.S. presidential election, etc.

• **Bloomberg** Beijing, China
Financial Reporter Intern (Global Summer Internship Program) July 2015 – September 2015

- Wrote and published more than 20 financial news stories, covering Chinese economic data and policy, market movements, financial statements, mergers and acquisitions, box office, etc.
- Covered 2015 China's stock market meltdown by interviewing industry experts on market movements and analyzing market data variations on Bloomberg Terminal independently.
- Collected market facts via Bloomberg Terminal and got experts' quotes by doing phone interviews to provide and verify the information for news stories written by senior reporters.

MEMBERSHIPS

- The Association for Information Systems (AIS)
 - The International Business Honor Society of Beta Gamma Sigma
 - The Honor Society of Phi Kappa Phi
-

SKILLS

- Substantial experience in Python and its deep learning tools (e.g., Keras, TensorFlow, PyTorch, etc.)
 - Fluent in R, Excel, Tableau, Bloomberg Terminal
 - Some knowledge in SQL, SAS, Hadoop, Spark, MATLAB, Maple, SWIFT
-

OTHER ATTENDED EDUCATION PROGRAMS

- **Certificate for Wharton Seminars for Business Journalists** October 2017
University of Pennsylvania, The Wharton School
 - **Courses in Mathematics** January 2017
The City University of New York, Borough of Manhattan Community College
 - **Certificate in Communications** September 2012 – February 2013
Hanze University of Applied Sciences
-

OTHER RESEARCH AND ANALYTICS EXPERIENCES (PRE-Ph.D.)

- **Determining the Right Location for Starting a New Company**
 - Conducted deep text cleaning on 3000 company location selection criteria articles in Python. Trained Phrases model to detect common phrases in the texts, so phrases can be treated as individual tokens. Applied TF-IDF model to extract highly-weighted words/phrases in each article.
 - Applied word2vector model to produce word embeddings, based on which a PCA was conducted. Clustered texts with K-means based on PCA and plotted inter-word distance map to visualize words connections. Captured word usage difference across multiple articles with a comparison word cloud.
 - Conducted topic analysis across the articles with LDA model to identify key elements in considering company locations. Computed word-topic probabilities and document-topic probabilities.
- **Analyzing Music Reviews and Their Influence on Customer Behavior**
 - Conducted broad literature review on how critics reviews affect business. Investigated how this problem had been studied and proposed new research directions in the music industry.
 - Analyzed 200 thousand music critics reviews' evolvement across 18 years in R. Built relations between album sales and insights of the reviews. Identified influential reviewers who drive the trend.
 - Grouped the reviews by year and applied word2vector model to locate top words related to certain sentiments, music genres, playing platforms. Plot their sum to analyze the changes across years.
- **Sequence Clustering with Temporal Graph and Sequence Clustering**
 - Co-proposed a novel sequential data clustering framework based on temporal graph. Co-developed a sequence recovering algorithm to recover clusters' centroids for interpreting the clusters. Collected real-world sequential datasets to test the performance of the framework and algorithm.
 - Computed Temporal Graph, Uni-gram, and PrefixSpan on 540 thousand rows of e-commerce data in Python. Experimented sequence clustering based on these three transformations with K-means and Hierarchical methods and compared the results by computing confusion matrices.
 - Co-authored the paper "Sequence Clustering with Temporal Graph and Sequence Recovery", which recorded the entire research of the proposed framework and algorithm. The paper was accepted by the 2018 INFORMS Workshop on Data Science, Phoenix AZ, November 2018.
- **Customer Segmentation and Gas Consumption Forecasting**

- Built customer segmentation metric for local energy company based on gas usage magnitude and response to temperature. Validated with identified cutting thresholds and K-shape method in Python.
 - Identified customers in each new segment and mapped them with the company's old segments clustered without considering weather influence to detect errors in the old segments.
 - Identified influential variables and applied linear regression to each old and new segment to forecast gas usage across multiple time windows. Compared forecasting errors with RMSE and MAE.
- **New York City Taxi Fare Prediction**
 - Conducted EDA and data preprocessing on 55 million rows of data in Python. Implemented feature engineering based on official taxi fare charging principles and Great Circle Distance metric.
 - Applied LGBM model for prediction and repeatedly tested the parameters for the best performance. Trained the model with 85% of data and tested the prediction errors with RMSE.
 - Final ranking on the leaderboard was 35th place out of 1488 teams, which was among the top 3%. Posted a Kernel of this project on Kaggle and received considerable views and likes.
 - **Excess Funds Detection and Forecasting**
 - Forecasted excess funds in checking accounts across multiple time windows in R to help a local bank identify customers who are likely to transfer their money from checking to saving accounts.
 - Processed roughly 38 million rows of data and 60 thousand accounts. Employed Holt-Winters Exponential Smoothing model to perform time series analysis for forecasting excess funds.
 - Forecasted minimum balance of target accounts across weekly, bi-weekly, and monthly time windows. Used Actual Prediction Error to test errors and determine the best-fit model.
 - **Chatbot Development for Company Recruitment**
 - Developed a chatbot that can output appropriate answers based on different input questions in R to help the client company answer frequently asked questions during recruitments.
 - Used Support Vector Machine, Uni-gram, Text Clustering and other machine learning tools to develop the chatbot. The development was based on given Q&A data and self-enriched data.
 - The chatbot was developed successfully in full function within 12 hours and received the merit award for the 2018 Drexel University Datathon competition.
 - **Crime Data Analysis and Safety Rating**
 - Designed, developed, and launched the IOS app "Safety Fellow" in Swift to provide safety information for the local community based on the mobility features of cell phones.
 - Proposed a novel safety rating algorithm to compute the "Safety Index" by blocks, based on crime frequency, type, occurring time, block area, and population size.
 - Processed and analyzed 10 million rows of crime data. Managed to tell a reasonable data story based on the analysis. Closely used Xcode IDE, Excel, Photoshop, ArcGIS for the development.