

# Yashu Kang

✉ yashu.kang@outlook.com ☎ 402-429-5459 in [www.linkedin.com/in/yashukang](http://www.linkedin.com/in/yashukang)

## CAREER OBJECTIVE

I am interested in executing data-driven solutions to increase efficiency, accuracy, and utility of data processing. Experienced at statistical analysis, using predictive data modeling, and analyzing data mining algorithms to deliver insights and implement action-oriented solutions to crucial real-life problems.

## SKILLS

Languages: Python, R, Java, SQL, MATLAB, SAS

Tools & Framework: Azure, Hadoop, Spark, Hive, PIG, Node.js, ArcGIS

## WORK EXPERIENCE

### **Research Assistant: Transportation Safety Data Analysis**

**Aug. 2019 – Jan. 2021**

- Extracted and statistically analyzed ten-year transportation safety data: classification methods on injury severity and count models on crash frequency
- Designed metrics and assessed safety performance of rural stop-controlled state highway systems
- Corrected 2,241 values in Federal Rail Administration's database and improved crash prediction accuracy by 80% over USDOT's model

### **Algorithm Engineer Intern: Ping An Technology**

**May 2019 – Aug. 2019**

- Estimated and deployed models predicting risks of Hyperuricemia using XGBoost and logit regression, and eventually made available to millions of users
- Built a deep recurrent survival analysis model on Anzhen hospital's time-to-event data
- Achieved 11<sup>th</sup> place in KDD Cup 2019 Reinforcement Learning competition track on malaria control

### **Teaching Assistant: Urban Transportation Planning**

**Sep. 2018 – May 2019**

- Worked as graduate teaching assistant for two semesters
- Provided educational materials and mentored 30+ students, including weekly assignments and a final project
- Delivered more than 10 full lectures in classroom and lab

## PROJECTS

### **Research on School Zone Safety**

**Jun. 2018 – Jun. 2020**

- Collected speed data of over 378,000 vehicles at 18 school zones in Lincoln
- Constructed data pipeline to extract and map five-year crash data related to school zones from the DOT
- Designed metrics, quantified safety effects of school zone establishment and provided guidance for local government

### **Sepsis Prediction Using Big Data Analytics**

**Aug. 2018 – Dec. 2018**

- Built data pipelines with PostgreSQL to extract information from MIMIC database
- Applied tools including Hadoop, Hive and PIG for data ETL processing
- Used various models including logistic regression, CNN, RNN to eventually achieve an accuracy of 95%

### **Interoperability between EHR and Lab Applications**

**Aug. 2018 – Dec. 2018**

- Created two healthcare systems for patients and labs respectively, using MongoDB and Node.js
- Developed back-end demo and enabled data communication between the two systems

### **NLP-based Recommendation System**

**Jun. 2018 – Aug. 2018**

- Conducted exploratory data analysis and word clouds to provide insights on DonorsChoose data
- Made Sankey plots to demonstrate the flow of donations over \$100,000
- Utilized cosine similarities of TF-IDF vectors to make content-based recommendations to potential donors

### **Online Tool Rental System**

**Aug. 2017 – Dec. 2017**

- Built a web application using Node.js for small tool rental business

- Constructed back-end demo to help provide a unified system for customers and employees
- Created an efficient conceptual SQL model that reflects the relationships between the products and services

### **Statistical Analysis on FRA Crash Data**

**May 2016 – Aug. 2017**

- Used PCA, K-means and latent class clustering to deal with unobserved heterogeneity within data
- Identified significant contributing factors using logistic regression models and obtained model interpretation
- Modeling on clustered data revealed more information than models with unclustered data

## **EDUCATION**

### **University of Nebraska Lincoln, 2021**

Ph.D. in Transportation Engineering  
Minor in Statistics

### **Georgia Institute of Technology, 2019**

M.S. in Computer Science

### **Beijing Jiaotong University, 2014**

M.S. & B.S. in Transportation Engineering

## **SELECTED COURSEWORK**

- UNL: Urban Transportation Planning, Multiple Regression Analysis, Categorical Data Analysis, Applied Multivariate Statistical Analysis, Nonlinear Optimization
- Georgia Tech: Machine Learning, Computer Vision, Database System Concepts & Design, Software Development Process, Big Data Analytics for Healthcare, Knowledge-based AI

## **PUBLICATIONS & TECHNICAL REPORTS**

- Khattak, A. J., & **Kang, Y.** (2020). Research on School Zone Safety (No. M092). Nebraska. Department of Transportation.
- Khattak, A., **Kang, Y.**, & Liu, H. (2020). Nebraska Rail Crossing Safety Research (No. SPR-P1 (19) M091). Nebraska. Department of Transportation.
- Li, Y., Jia, W., **Kang, Y.**, Chen, T., Li, X., Du, X., ... & Xie, G. (2020). DeepComp: Which Competing Event Will Hit the Patient First? *In 2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)* (pp. 629-636). IEEE.
- **Kang, Y.**, Iranitalab, A., & Khattak, A. (2019). Modeling railroad trespassing crash frequency using a mixed-effects negative binomial model. *International Journal of Rail Transportation*, 7(3), 208-218.
- Iranitalab, A., **Kang, Y.**, & Khattak, A. (2018). Modeling the Probability of Hazardous Materials Release in Crashes at Highway–Rail Grade Crossings. *Transportation Research Record*, 2672(10), 28-37.
- Khattak, A., & **Kang, Y.** (2018). Offset Right-Turn Lanes on State Highway Systems (No. 26-1121-0030-001). Nebraska. Department of Transportation.
- **Kang, Y.**, & Khattak, A. (2017). Cluster-based approach to analyzing crash injury severity at highway–rail grade crossings. *Transportation Research Record*, 2608(1), 58-69.