

# Jiayi Xu

Tel: +86-18602735011 | Email: [jiayixu17@gmail.com](mailto:jiayixu17@gmail.com) | Blog: <https://jiayi-blog.vercel.app>

## EDUCATION

**China University of Geosciences (CUG), Wuhan, China**

Sept 2019 - July 2023

Bachelor of Management in Information Management and Information Systems

**GPA: 91.27/100**

**Relevant Coursework:** Applied Statistics, Machine Learning and Data Mining, Operations Research, Advanced Mathematics, Linear Algebra, Probability Theory and Mathematical Statistics.

## RESEARCH EXPERIENCE

**China University of Geosciences - Center for International Cooperation in E-Business**

Mar 2021 - Present

*Research Assistant* | Advisor: [Prof. Zhu](#)

Project: Knowledge Graph Construction and Attribute Value Extraction for Large-scale Textual Data of Tourism Products

- Used Python to crawl and preprocess product textual data from Ctrip.com.
- Constructed a domain ontology for tourism products, incorporating 9 feature dimensions.
- Extracted tourism product features using BERT-BiLSTM-CRF and merged entities with highly semantic similarity.
- Imported entities and relations into a Neo4j database for graph visualization.
- Obtained product similarity through link prediction algorithms (Adamic-Adar) and modified cosine similarity.

Project: Research on Product Homogeneity Characteristics and Performance of Complementors in E-commerce Platforms

- Measured the homogeneity level of products between platform owner and complementors by constructing graph networks and calculating graph node similarity.
- Introduced 42 additional feature variables for the complementors and trained the sales forecasting model using XGBoost.
- Generated parameter interpretation by SHAP feature importance derived from the XGBoost model.
- Devised the plan to use causal inference methods to facilitate the interpretation of treatment variables.

**Tongji University**

Apr 2022 - Sept 2022

*Research Intern* | Advisor: [Prof. Lu](#)

- Compiled journal articles related to geopolitical risks in the past 20 years based on TF-IDF and counted the distribution of the number of articles by year.
- Measured the similarity between datasets using Pearson's correlation coefficient, cosine similarity, and KL divergence.
- Implemented term set expansion, including training word embedding mode (Word2vec), manual annotations, and building classification models using Random Forest.

## PUBLICATION

**Xu, Jiayi;** Zhang, Shuang; Zhu, Zhen; Zou, Lincan; and Yang, Mengting, "Exploratory Research on Knowledge Graph Construction and Attribute Value Extraction for Large-scale Textual Data of Tourism Products" (2022). WHICEB 2022 Proceedings. 19. <https://aisel.aisnet.org/whiceb2022/19>

## INTERNSHIPS

**Didi Chuxing (Beijing) Co., Ltd.**

Beijing, China

*Data Analyst Intern - Strategic Development Division*

Sept 2022 - Dec 2022

- Utilized SQL, Excel, and Tableau to collect, clean, analyze, and visualize driver and passenger data across 45 cities.
- Performed Attribution Analysis for weekly data fluctuation and anomalies and presented analysis reports.
- Involved in designing randomized controlled experiments (e.g., A/B tests, switchback experiments) and conducting evaluations of experimental effectiveness.
- Constructed virtual control groups (e.g., DID, PSM) to evaluate experiments conducted for all users by Python and R.

**PricewaterhouseCoopers (PwC) Acceleration Center**

Shanghai, China

*Assistant Data Engineer Intern - P&T Digital*

June 2022 - Aug 2022

- Designed data models and maintained SQL Server databases to achieve business needs.
- Contributed to improving a risk assessment algorithm by constructing the Bayesian Network.
- Championed the Data Scientists team in achieving AI-driven automatic parsing tasks for PDF documents, including text data annotation and preprocessing.

- Conducted new product demand analysis for a music company, including collecting secondary data and unearthing user requirements by focusing on finance, educational resources, communication platforms, social culture, etc.
- Built user profiles by user clustering according to the full lifecycle process of musicians from entry to exit.
- Led a team for market positioning and competitive analysis by using SWOT and PEST analysis.

## SELECTED PROJECTS

---

### **The 2<sup>nd</sup> Campus Geeks Contest of Digital China Holdings Ltd**

May 2022

#### **Team member | Project: Warehouse management path picking optimization**

- Developed a mathematical model for the VRP problem and used a genetic algorithm by Java for its solution.
- Improved the computation time to less than 20 minutes, securing a top 5% rank in the competition.

### **The Mathematical Contest in Modeling 2022 (MCM/ICM)**

Feb 2022

#### **Team member | Project: Optimal trading strategy for Bitcoin**

- Employed LSTM neural networks to predict the daily trading prices of gold and Bitcoin.
- Designed a multi-stage investment portfolio optimization model based on dynamic programming algorithms and implemented a dynamic stop-loss model as a secondary risk prevention.

### **2021“Haihui Cup” Smart City and Intelligent Construction Innovation & Entrepreneurship Competition Aug. 2021**

#### **Team Leader | Project: Green Retrofit Decision Support System**

- Designed a decision support system by creating UML diagrams and developing prototypes.
- Assumed responsibility of front-end development tasks using HTML.

## AWARDS

---

College Students' Innovative Entrepreneurial Training Plan Program (National) - Excellent Level (top 5%)

2022

National Hot Chainers Inaugural University Blockchain App Design Competition (National) - 3rd Prize

2021

## SKILLS

---

**Data and Statistical Skills:** Data Collection and Cleaning, Database Design and Management, Data Analysis and Visualization, Predictive Analytics, Statistical Modeling and Inference

**Programming & Software:** Python, R, SQL, Stata, Java, C, Tableau, Cypher, Lingo, Excel, Android Studio, LaTeX