

JINXIANG MA

+1(510) 345-8049 ◇ Sunnyvale, CA

majinxiang2000@gmail.com ◇ [Linkedin](#) ◇ [Portfolio](#)

SUMMARY

Versatile Data Science Professional and Industrial Engineer with hands-on experience across analytics, mathematical modeling, and data engineering. Skilled in building scalable pipelines and translating complex models into actionable insights for decision-making. Proficient in Python, SQL, Power BI, and cloud tools. Strong interest in applying data science and operations research to real-world challenges in AI, intelligent systems, and large-scale technology environments.

EDUCATION

University of Michigan, Ann Arbor, Ann Arbor, MI

Graduated May 2025

Master of Science(M.S.), Industrial and Operations Engineering

GPA: 3.8

Relevant Coursework: Linear Programming, Network Optimization, Integer Programming, Stochastic Process, Queuing Theory, Numerical Optimization, Data Envelopment Analysis, Risk Analysis, Reinforcement Learning, Dynamic Programming, Scheduling, Supply Chain Facility, Statistical Learning

Graduate Certificate, Data Science

Relevant Coursework: Database Systems, Database Application Design, Information Retrieval, Spatial Analysis

University of California, Santa Barbara(UCSB), Santa Barbara, CA

Graduated Jun 2023

Bachelor of Science(B.S.), Statistics and Data Science

GPA: 3.74

Relevant Coursework: R Programming, SAS Programming, Python Programming, Probability Theory, Regression Analysis, Data Science Concept and Analysis, Mathematics of Fixed Income Markets, Bayesian Data Analysis, Non-parametric Statistics, Experimental Design, Financial Market Risk and Modeling, Time Series Analysis, Machine Learning, Predictive Modeling in Linguistic, Deep Learning, Cloud Computing & Big-Data Analytic

Bachelor of Art(B.A.), Mathematics

Relevant Coursework: Linear Algebra, Discrete Mathematics, Vector Calculus, Differential Equation, Real Analysis, Numerical Analysis, Operations Research, Number Theory

EXPERIENCE

Data Management Consultant

Aug 2024 - Now

Westland Management Solution

San Francisco, CA

- Built automated pipelines to synchronize data between Primavera Unifier and external systems, reducing manual updates and ensuring consistency across platforms.
- Developed Python-based QA/QC validation scripts to identify data entry issues and enforce schema integrity, improving input reliability for downstream reporting.
- Created interactive dashboards in Power BI for senior management, tracking project milestones, financial metrics, and schedule adherence.

Process Improvement Consultant

Aug 2024 - Dec 2024

Michigan Medicine

Ann Arbor, MI

- Applied SARIMA models to forecast infusion workloads, improving accuracy and identifying peak periods.
- Proposed a hybrid flow shop scheduling model to optimize the chemotherapy preparation workflow, integrating the newly introduced ultrasound sonication step that operates concurrently with other tasks, reducing makespan and bottlenecks and enhancing sterile compounding accuracy by 5–10%

Engineering Design Trainee - Data Science(Promoted to Intern Lead)

Jun 2022 - Aug 2024

San Francisco Public Utilities Commission

San Francisco, CA

- Overhauled capital-project tracking: consolidated 200+ spreadsheets into a Power App, SharePoint & Oracle Database workflow, cutting manual data entry by 70% and trimming three days off each monthly reporting cycle.

- Built an enterprise Master Project List by designing a Pentaho Kettle ETL pipeline and an NLP ensemble (BM25, TF-IDF, Transformer) that reached 92% precision in entity matching across three legacy sources.
- Automated budget analytics for a \$6B Bay-Area infrastructure portfolio; Power BI dashboards flagged phase-level cost variances and informed reallocations in FY 23.
- Built a Python–Django forecasting tool that visualized staff availability and workload projections 12 weeks ahead, helping Project Managers reduce last-minute reassignments and improve planning accuracy.
- Led intern training and knowledge transfer: produced SOPs and live coding workshops for six interns, reducing ramp-up time from two weeks to three days.

Data Analyst

Feb 2022 - Jun 2023

GoGaucho

Santa Barbara, CA

- Engaged in the app development process of GoGaucho, a mobile app serving over 10,000 users monthly. Retrieved academic curriculum data using UCSB API, analyzed historical course enrollment, and shared results with the GoGaucho team

Student Assistant

Sep 2021 - Jun 2023

Data, Research, Exploration, Access & Methods (DREAM) Lab , UCSB Library

Santa Barbara, CA

- Assisted library users in cleaning and preparing data for analysis, improving data accuracy and usability
- Maintained statistics on the lab and identified IT issues, ensuring smooth operation and timely issue resolution
- Answered student’s questions in the Data Science Carpentry workshop and develop workshop materials.
- Delivered programming projects on speech processing and web-scraping, enhancing research and data efficiency.

RESEARCH

Student Researcher

Jan 2024 - Now

SOCR Lab, University of Michigan, Ann Arbor

Ann Arbor, MI

Research Project: Compressive Big Data Analytics (CBDA)

Advisor: Ivo D. Dinov (Professor)

- Developed Python scripts to implement efficient sub-sampling techniques, including reservoir sampling and chunk-based methods, enabling analysis of large datasets (40GB) without loading the entire dataset into memory.
- Applied the SuperLearner algorithm to small data samples, evaluating and ranking prediction models based on accuracy metrics.

Student Assistant

Sep 2021 - Jun 2023

Data, Research, Exploration, Access & Methods (DREAM) Lab , UCSB Library

Santa Barbara, CA

- Assisted library users in cleaning and preparing data for analysis, improving data accuracy and usability
- Maintained statistics on the lab and identified IT issues, ensuring smooth operation and timely issue resolution
- Answered student’s questions in the Data Science Carpentry workshop and develop workshop materials.
- Delivered programming projects on speech processing and web-scraping, enhancing research and data efficiency.

SELECTED PROJECTS

Impact of Memory Size on Quasi-Newton Methods A comprehensive software package focusing on solving unconstrained optimization problems in Matlab. Explored and implemented a set of optimization methods, particularly focusing on answering the question of how the memory impact the performance of quasi-Newton methods ([Link](#))

Lazymind - AI Planning Assistant Built a modular task planning application that transforms ambiguous user goals into structured, multi-step execution roadmaps. Designed core logic for roadmap generation with attention to execution traceability, fallback handling, and future scheduling extensibility. Applied systematic testing and interface isolation to ensure reliability across roadmap types and user inputs. Frontend developed using TypeScript and Next.js

Bayesian Spatial Modeling of Birth Weight in Bangladesh Developed Full GP, Vecchia approximation, Nearest Neighbor GP, and Meshed GP models to predict child birth weight in unseen regions based on maternal and household-level covariates. Used exponential covariance functions, applied MCMC (Gibbs, MH) for posterior inference, and evaluated spatial generalization through prediction maps and semivariograms. ([Link](#))

Time Series Forecasting on SFO Air Passenger Statistics. Conducted a comprehensive time series analysis on San Francisco International Airport (SFO) passenger data from 2005-2019 to forecast future passenger traffic. The project involved extensive data cleaning, transformation, and model diagnostics, leading to the selection of an optimal SARIMA model based on AICc criteria. [\(Link\)](#)

Named Entity Recognition with Transformer Developed and trained a Transformer model for Named Entity Recognition (NER) using the PyTorch framework, focusing on identifying and classifying named entities such as persons, locations, and organizations from text. [\(Link\)](#)

Python Code Autocomplete and Suggestions Chatbot Developed a semantic search and RAG-based system using FAISS and GPT-3.5-turbo to retrieve and generate Python code from natural language queries, improving search precision by 58%. Built a FastAPI backend and Streamlit UI for real-time code suggestions. [\(Report\)](#)

Optimization of Wind Power Plant Locations Determined optimal locations for wind power plants in Iran using Data Envelopment Analysis (DEA). Apply multiple DEA models, including CCR, BCC, Super Efficiency, and Non-discretionary models, to assess and rank potential sites based on critical factors such as wind speed, land availability, and proximity to power grids. [\(Link\)](#)

Project Management Tool for Infrastructure Projects Developed a Django-based CRUD application to track and manage the status of infrastructure projects. Utilized Bootstrap for the frontend interface to ensure a responsive and user-friendly experience. Implemented PostgreSQL for robust backend data storage solutions, enhancing data integrity and accessibility. Designed and built ETL pipelines using Pentaho Kettle to facilitate efficient batch updates and data retrieval from the PostgreSQL database, optimizing the flow of project information and improving decision-making processes.[\(Link\)](#)

Risk Analysis of Deepwater Horizon Blowout Conducted risk analysis case study on the Deepwater Horizon blowout, applying Functional Block Diagram and Fault Tree Analysis to model the Blowout Preventer (BOP) system and trace root causes of failure. Analyzed technical reports and CSB animations to identify cascading mechanical, electrical, and control system failures

Multilingual Audio Transcription with Azure Speech to Text SDK Performed Batch Transcription using REST API. Customized continuous recognition with language detection and speaker diarization features to generate transcription for bilingual audio. Trained and tuned custom speech model in Azure Speech Studio, yielding a new model that has a lower word error rate(WER) compared to the original model.

Dianping Business Scrape Scraped Children’s Libraries information in China from DianPing.com (the Chinese equivalent of Yelp) with HTTP requests and sessions, extracted website cookies using selenium web-driver and bypassed Captcha with a modified request header. Retrieved a list of children’s libraries in over 2000 Chinese cities with the URL of their respective web pages. Parsed HTML code iteratively into BeautifulSoup object, located and scraped relevant information with XPath, and exported the final result into CSV files. [\(Link\)](#)

SKILLS

Programming Skills	Python(Pandas, SciPy, Matplotlib, Scikit-learn, PyTorch, XGBoost), R(dplyr, ggplot2), C++, Scala, SQL, SAS, JavaScript, HTML, CSS, Stan, VBA, Git, Bash, L ^A T _E X
Computer Software	Jupyter Notebook, RStudio, Matlab, PostgreSQL, PowerBI(DAX), Power Query PowerApp, Power Automate, SharePoint, Premavera Unifier, Microsoft Azure, AWS, Pentaho Kettle, Django, Gurobi, Docker, Kubernetes
Languages	English(Fluent), Cantonese(Native), Mandarin(Native)