(+44) 078-21824835
Glasgow, UK
majiahui821@gmail.com

Jiahui Ma

jiahui-portfolio.com
github.com/dsnasn
linkedin.com/in/jiahui821

Skilled in building end-to-end data solutions and deploying models on Azure and AWS. Proficient in Python, SQL, and ML frameworks, with expertise in statistical analysis, NLP, recommendation systems, and data-driven decision-making.

Skills

Programming	Python, SQL, Bash, Java
Frameworks	PyTorch, TensorFlow, MapReduce, Apache Spark
Cloud Platforms	Microsoft Azure, AWS, Google Colab, Databricks
Tools & Visualization	Streamlit, Tableau, Power BI, Matplotlib, Seaborn, Docker, Git, Jenkins

Education

Master of Computer Science, The University of Glasgow — Expected Sep 2025
Bachelor of Software Engineering, Beijing Jiaotong University | UK 2:1 Equivalent — 2020 — 2024
Relevant Coursework: Machine Learning & AI, Data Science and Systems, Big Data, Deep Learning

Work Experience

Research Intern — **Mar 2025 — Present | Glasgow, UK**
University of Glasgow
- Analyzed and visualized data transmission paths in IoT systems, enabling reliable upstream/downstream data flow to support data modeling and analytics.
- Developed a real-time dashboard with Grafana and InfluxDB for visualizing GPS data from LoRaWAN-based drones delivering medicine in remote mountain regions.

Machine Learning Intern — **Jan 2024 — Jun 2024 | Beijing, China**
CNBM Information Technology Co., Ltd.
- Improved data cleaning efficiency by **30%** through SQL optimization and managing **500GB+** of industrial data.
- Engineered features (power-output ratio, rolling mean energy, peak-hour flags) to boost model performance.
- Built energy and production forecasting models (Random Forest, XGBoost), increasing AUC from **0.75** to **0.85**, enabling smarter power allocation and reduced energy waste.
- Built KPI dashboards using Matplotlib and Seaborn, enhancing decision-making across departments.

Data Scientist Intern — **Jun 2023 — Dec 2023 | Beijing, China**
ByteDance
- Designed real-time dashboards to monitor key ad metrics (PV, UV, conversion rates) during promotions, assessing strategy impact and detecting issues. Improved click-through by **5%** and conversion by **3%** via A/B testing.
- Built churn prediction model (XGBoost) to detect at-risk users, achieving F1 score of **0.82** and reducing churn by **8%**. Enabled proactive retention strategies, such as targeted promotions and personalized recommendations.
- Conducted user traffic analysis during promotions, increasing active users by **7%** and ARPU by **2%**.

Projects

Sentiment Analysis using NLP and LLMs | GitHub — **Jun 2024 — Aug 2024**
- **Model Development**: Fine-tuned a BERT-base-Chinese model on **30,000+** user reviews for sentiment analysis, achieving **92%** accuracy and F1 score of **0.89**.
- **Cloud Deployment**: Deployed a Streamlit app with Docker, hosted on Azure and AWS for real-time API-based sentiment analysis.

Breast Cancer Histopathology Diagnosis | GitHub — **Sep 2024 — Dec 2024**
- **Model Development**: Fine-tuned DenseNet121 for multi-class classification of breast cancer images, achieving **95%** accuracy.
- **Deployment**: Built Streamlit web application for real-time image upload and diagnosis, improving user accessibility.

Financial Asset Ranking with Spark — **Feb 2025 — Apr 2025**
- Designed a Spark pipeline in Java using MapReduce-style transformations to rank financial assets from **24M+** US stock records.
- Delivered top-5 investment recommendations via scalable batch analytics and efficient distributed filtering.

Spotify Listening Analysis with Tableau | Explore My Visualizations
Explored personal listening behavior using clustering to generate playlist recommendations; visualized global trends.

NHS Healthcare Dashboard with Power BI | Explore My Visualizations
Built a dashboard analyzing NHS Scotland patient flow and hospital efficiency to uncover service bottlenecks.

Certifications

DP-100, DP-203 (Azure Certifications), IBM Data Science Professional Certificate, AWS Certified Cloud Practitioner
Google Data Analytics, IELTS 8.0 (fluent), Chinese(native)