MohammadJavad AghababaieBeni

75 Waterworks Rd, Birmingham, B169AL
mj.babaie@gmail.com | 07508487847

Applied Data Scientist with a strong ML and AI engineering background, specialising in production-grade predictive models and data-driven systems. Experienced in taking models from experimentation to deployment, with hands-on expertise in Python, SQL, MLOps, CI/CD, and model monitoring, delivering measurable business impact across regulated domains. UK Graduate Visa holder (valid until February 2027).

GitHub: https://github.com/Muh76
LinkedIn: https://www.linkedin.com/in/mohammadbabaie/

Professional Experience

Outlier AI Remote/US **10/24 - Present**
LLM Evaluation Analyst

- Evaluated and scored 600+ LLM responses across legal QA, general knowledge, and creative tasks, identifying reasoning errors, hallucinations, and knowledge gaps.
- Designed and authored data science–focused coding tasks (data preprocessing, modeling, evaluation, and error analysis) to assess LLM reasoning quality and alignment with real-world analytical workflows.
- Implemented structured evaluation frameworks and content guardrails, improving response correctness by ~20% across subsequent model iterations.
- Collaborated with AI engineers by delivering detailed evaluation reports and actionable feedback, directly informing data and model improvements.
- Increased evaluation coverage by 30% by designing edge-case and stress-test scenarios, strengthening robustness across diverse input distributions.

Behineh Bana Zagros Isfahan/Iran **02/20 - 06/23**
Data Scientist

- Analyzed operational, cost, and project datasets across multiple construction and industrial initiatives to support planning and data-driven decision-making.
- Built regression and time-series models to forecast project timelines, material demand, and costs, reducing manual estimation effort by ~25%.
- Developed automated data processing pipelines and analytical reports using Python and SQL, cutting recurring reporting time by ~30%.
- Designed and maintained Power BI dashboards tracking KPIs (progress, cost variance, resource utilisation) for both technical and non-technical stakeholders.
- Cleaned, structured, and maintained large-scale datasets covering procurement, schedules, and supplier performance.
- Applied statistical analysis to uncover patterns, anomalies, and operational risks, supporting measurable improvements in project execution.

Education

Aston University **09/23 - 10/24**
MSc in Computer Science

Azad University of Sharekord **09/16 - 01/20**
BSc in Computer Science

Key Skills

- Programming & Data: Python, NumPy, Pandas, scikit-learn, TensorFlow, PyTorch, SQL, Git
- Machine Learning: Predictive Modeling (Regression, Classification, Time Series Forecasting, Neural Networks), Hyperparameter Tuning, Clustering (KMeans, DBSCAN), Ensemble Methods (Random Forest, XGBoost, LightGBM, CatBoost)
- Analytics & Statistics: Exploratory Data Analysis (EDA), Feature Engineering, Hypothesis Testing, Correlation Analysis, Outlier Detection
- LLMs & NLP: LLM Applications (RAG pipeline, Embedding, Prompt Engineering, LangChain), Hugging Face Transformers, Text Classification
- MLOps & Deployment: Docker, MLflow, DVC, CI/CD (GitHub Actions), FastAPI
- Cloud & Analytics: GCP, Azure ML, Power BI, Excel

Projects

Diabetes Hospital Readmission Prediction **02/25 - 06/25**

- Predictive analytics model to identify patients at risk of 30-day hospital readmission (130k+ records)
- Engineered 300+ clinical and demographic features with grouped cross-validation to prevent patient-level leakage
- Trained and compared multiple models (Logistic Regression, LightGBM, XGBoost, CatBoost) with hyperparameter tuning
- Achieved ROC-AUC 0.67 and 67.9% accuracy on held-out data
- Estimated ~£6.5M annual cost savings with ~3-month break-even if deployed at hospital scale
- Deployed model with experiment tracking and monitoring dashboards to assess performance and business impact

Contract Review & Risk Analysis System (CUAD) **04/25 - 08/25**

- NLP system for automated clause segmentation, contract classification, and risk scoring
- Built on the CUAD dataset with production-grade evaluation
- Micro-F1 0.837 and 97.4% contract-type accuracy
- Sub-500 ms inference latency per contract
- End-to-end pipeline: TF-IDF + Logistic Regression and DistilRoBERTa with probability calibration
- Model serving via FastAPI with batch and dashboard interfaces
- Deployed on GCP Cloud Run with Dockerised services
- MLOps practices: DVC/MLflow tracking, group-by-contract splits
- ROI analysis showing break-even at ~14 contracts/month (≈6.7-month payback)

Linear B-Cell Epitope Prediction (MSc Thesis) **02/24 - 09/24**

- Peptide sequence classification models for linear B-cell epitope prediction in vaccine research
- Large-scale immunoepitope dataset (~601k peptides; 393 biochemical features from ESM-1b embeddings)
- Coronavirus-specific subset with peptide filtering (8–25 amino acids) and stratified train/validation/test splits
- 77% feature dimensionality reduction using Boruta and Genetic Algorithms to reduce overfitting
- Trained Feedforward Neural Networks and XGBoost models
- Achieved AUC 0.994, F1 0.88, MCC 0.789 on target dataset; AUC 0.981+ on general dataset
- Azure ML–based training with experiment tracking and Bayesian optimisation

Legal Q&A RAG Chatbot (UK Legal Domain) **09/25 - Present**

- Production-ready RAG system over 131k+ legal document chunks
- Hybrid retrieval: BM25, FAISS, OpenAI embeddings, RRF, cross-encoder reranking
- FastAPI backend with role-based access (Solicitor / Public)
- Sub-3s average latency with 680+ embeddings/sec throughput
- 15–20% retrieval accuracy improvement via reranking
- 40% hallucination reduction using domain filtering and citation guardrails
- End-to-end deployment: Docker, PostgreSQL, Streamlit