TUSHAR JAYENDRA MHATRE

+1 (405) 223-8789 | tushar.jayendra.mhatre-1@ou.edu | Norman, OK, USA | tusharm24.github.io/TusharM_Portfolio/

PROFESSIONAL SUMMARY

Data Scientist/ML Engineer with 3.5 years of experience in designing, deploying, and optimizing end-to-end machine learning and AI solutions. Skilled in MLOps, predictive modeling, and ETL pipelines, with a strong foundation in A/B testing, statistical analysis, and cloud-based ML deployment.

PROFESSIONAL EXPERIENCE

The Aether Loop **Remote**
AI Engineering Intern *September 2025 - Present*
- Building multi-Agent AI system with Crew AI to build a robust supplement recommendation system based on patient blood biomarkers.
- Using Langchain to implement RAG based retrieval system for ingredient search based on patient profile.

Optimal living Systems **Remote**
Data Architect Intern *July 2025 - September 2025*
- Developed and orchestrated ETL pipelines in Airflow that processed 1,000+ research PDFs into structured synthetic datasets through OCR(Docling) and open-source LLMs for LLM fine-tuning.
- Optimized synthetic data storage in GCP BigQuery with schema-structured collections, cutting LLM training data prep time by 40% (from 5h to 2h per batch).

University of Oklahoma **Oklahoma, USA**
Graduate Research Assistant *February 2024 - May 2025*
- Developed automated pipelines for transforming EEG and MRI data into BIDS-compliant structures in Python and MATLAB, cutting the manual conversion time by 40% and streamlining lab-wide data workflows.
- Maintained and updated large-scale neuroimaging datasets, ensuring seamless accessibility for a lab with 15 researchers.
- Applied clustering algorithms to identify Cortical Activity Patterns (CAPs) in brain EEG data across 10+ studies.
- Conducted statistical analysis (hypothesis testing, causal inference) on neuroimaging data to validate numerous research findings.
- Developed WSL-based Bash scripts to bridge Windows/MATLAB tools with Linux-optimized neuroimaging pipelines (FSL, AFNI), cutting preprocessing runtime by 40% through parallel job scheduling.
- Presented findings for research papers and presentations, translating complex findings for non-technical audiences.

Tata consultancy Services **Pune, Maharashtra, India**
Analyst *February 2021 - July 2023*
- Designed Power BI dashboards which tracked 15+ KPIs (sales, inventory), enabling real-time decision-making for clients.
- Collaborated with cross-functional teams in Agile sprints to gather requirements, define project scope, and deliver data solutions aligned with business needs.
- Used MLflow to track model performance, hyperparameters, and artifacts across 20+ forecasting experiments, enabling reproducible model selection and reducing deployment cycle time by 30%.
- Designed and deployed ETL pipelines (airflow) to aggregate data from multiple sources (SAP, mainframes) for forecasting models.
- Diagnosed and resolved 30+ production incidents in live forecasting pipelines (Python, Airflow), implementing fixes that reduced error rates by 30% and improved system uptime significantly.
- Developed and containerized ML models (Docker) and exposed as microservices via Flask, reducing prediction latency from 2secs to 300ms while handling 1K+ requests per minute (RPM).
- Conducted rigorous A/B testing on model iterations, demonstrating 15% lower RMSE and 20% better bias than previous models.

EDUCATION

University of Oklahoma **August 2023 - May 2025**
Master's, Data Science *GPA: 3.97*
- Courses: Data mining, Machine Learning Practice, Data Visualization, Database,data structures, Healthcare analytics, Statistics

PROJECTS

AI Research Paper Assistant - *Link to project*
- Developed a LLM-based multimodal research chatbot with Retrieval Augmented generation implemented with FAISS, Ollama, Langchain, and Python.
- Used Unstructured and docling for OCR and text extraction, created textual description and embeddings for figures and tables, and implemented hybrid retrieval for accurate retrieval.
- Built a UI in Streamlit for seamless interaction with the system and optimized the model through prompt engineering.

SKILLS

- **Programming Languages:** Python, R, SQL, C/C++, MATLAB, Java
- **Python Libraries:** NumPy, Pandas, matplotlib, Scikit-learn, Tensorflow, XGboost, Pytorch, Seaborn, imblearn, Flask
- **Statistical Analysis:** Hypothesis Testing, Confidence Intervals, Outlier Detection, Causal Inference
- **Database:** MySQL, Database Design, JDBC (Java Database Connectivity), Microsoft Azure, FAISS, PostgreSQL, GCP BigQuery
- **Business Intelligence tools & Technologies:** Tableau, Microsoft Power BI, Excel/Numbers/Sheets, Git, Jupyter
- **Natural Language Processing:** Tokenization, Named Entity Recognition (NER), TF-IDF, Word Embeddings (Word2Vec)
- **Generative AI:** LangChain, Ollama, Langraph, Crew AI, Hugging Face, RAG, Unsloth AI