TUSHAR JAYENDRA MHATRE

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

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

PROFESSIONAL EXPERIENCE

University of Oklahoma **Oklahoma, USA**
Graduate Research Assistant *February 2024 - Present*
- 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 various clustering algorithms to identify Cortical Activity Patterns (CAPs) in brain EEG data across multiple studies.
- Conducted statistical analysis (hypothesis testing, causal inference) on neuroimaging data to validate numerous research findings.
- Created visualizations (heatmaps, 3D cortical maps) for research papers and presentations, translating complex findings for non-technical audiences.
- Documented data workflows to ensure reproducibility and compliance with research standards and maintained detailed logs.
- Implemented 3D CNN architectures (PyTorch) to classify brain states from MRI scans, achieving 89% accuracy in detecting early-stage neurological biomarkers.

Tata consultancy Services **Pune, Maharashtra, India**
Analyst *February 2021 - July 2023*
- Built RPA solutions (Automation 360, VBA, SQL) to automate SAP/mainframe tasks, reducing manual effort by 80% (Awarded for excellence).
- Collaborated with cross-functional teams in Agile sprints to gather requirements, define project scope, and deliver RPA/AI solutions aligned with business needs.
- Developed demand forecasting models (Python, XGBoost), achieving 92% MAPE accuracy and reducing stockouts by 25%.
- Designed and deployed ETL pipelines to aggregate data from multiple sources (SAP, mainframes) for forecasting models.
- Diagnosed and resolved 15+ production incidents in live forecasting pipelines (Python, Airflow), implementing fixes that reduced error rates by 30% and improved system uptime significantly.
- Containerized ML models (Docker) and exposed as microservices via Flask , reducing prediction latency from 2s '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 legacy systems.

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, implemented hybrid retrieval for accurate retrieval.
- Built a UI in Streamlit for seamless interaction with the system and optimized the model through prompt engineering.

Predicting Customer Revenue - *Link to project*
- Engineered end-to-end ML pipeline: Performed EDA, data cleaning (handled missing values, normalized features), and applied RFE for feature selection.
- Implemented regression models using R's caret library (Linear, Lasso, Ridge, MARS, Elastic Net).
- Utilized cross-validation to reduce overfitting and performed statistical analysis using t-tests for model evaluation.

Database Design for Patient Assistance Network(PAN) - *Link to project*
- The system is built using Azure SQL Database and a Java application that interacts with the database via JDBC.
- Designed an ER diagram and converted it to a normalized relational database, reducing redundancy by 40% vs. the initial draft.
- Optimized SQL queries via indexing, improving query performance by 35% (Azure SQL).

Fake Job Detection Portal - *Link to project*
- Applied Natural Language Processing to vectorize job descriptions for training supervised Machine Learning models
- Trained various classification models using Sci-kit Learn and performed model evaluation through statistics using t-tests and ANOVA.
- Deployed Flask API with <1s latency per prediction and robust transformation pipeline in the backend

SKILLS

- **Programming Languages:** Python, R, SQL, C/C++, MATLAB, Java
- **Python Libraries:** NumPy, Pandas, matplotlib, Scikit-learn, Tensorflow, XGboost, Pytorch, Seaborn, imblearn, Flask
- **R libraries:** dplyr, tidyr, knitr, ggplot2, caret, h2o, readR, tibble, plotly, shiny
- **Statistical Analysis:** Hypothesis Testing, Confidence Intervals, Outlier Detection, Causal Inference
- **Database:** MySQL, Database Design, JDBC (Java Database Connectivity), Microsoft Azure, ChromaDB, FAISS, PostgreSQL
- **BI Tools and technologies:** LangChain, Ollama, Git, Jupyter Notebook, github, Excel/Numbers/Sheets, Tableau, Power BI, Word/Pages/Docs
- **Certifications:** Google Data Analytics Professional Certification, Advanced C++ certification (Udemy)