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 Management Systems, 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.
- Applied various clustering algorithms to identify Cortical Activity Patterns (CAPs) in brain across multiple studies
- Extensively applied statistical methods to verify various research hypotheses and validate findings in ongoing experiments
- Continuously produced visualizations like heat maps, 3D cortical activity maps, etc. for research publications and presentations

Tata consultancy Services **Pune, Maharashtra, India**
Analyst *February 2021 - July 2023*
- Implemented RPA solutions with Automation 360, VBA, and SQL to automate tasks across SAP and mainframe systems
- Developed and deployed Machine Learning models and data transformation pipelines using Python to forecast key business metrics.
- Collaborated directly with cross-functional teams in an agile environment to gather requirements, analyze business processes, and deliver scalable solutions.
- Received the 'On the Spot' Award for consistently delivering projects and eliminating 80% of manual effort in routine SAP/mainframe steps.

PROJECTS

AI Research Paper Assistant - *Link to project*
- Developed a LLM-based research chatbot with Retrieval Augmented generation implemented with FAISS, Ollama and python
- Used PyMuPDF and Regex for text extraction and segmentation, evaluated model performance using ROUGE and BERT language model.
- Built a UI in Streamlit for seamless interaction with the system and optimized the model through prompt engineering.

Predicting Customer Revenue - *Link to project*
- 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.
- Deployed the top performing model to predict customer revenue from website traffic data

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.
- Developed a user-friendly application through Flask with an API-driven prediction and data transformation pipeline

Financial Analysis of a Supermarket Chain - *Link to project*
- An interactive tableau story to explore the financial performance and key performance indicators of a supermarket chain.
- Provided stakeholders with actionable insights that will enhance sales strategies.
- Identified key products and target demographics to optimize future marketing and forecasting sales.

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 diagrams, converted to relational database, and optimized storage structures
- Implemented indexing for query performance optimization

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, Correlation Matrices, Outlier Detection
- **Database:** MySQL, Database Design, JDBC (Java Database Connectivity), Microsoft Azure, MongoDB, ChromaDB, FAISS
- **Data Analysis tools:** Tableau, Microsoft Power BI, Excel
- **Natural Language Processing:** Tokenization, Named Entity Recognition (NER), TF-IDF, Word Embeddings (Word2Vec)
- **Tools and technologies:** LangChain, Ollama, Retrieval-Augmented Generation (RAG), Hugging Face, Git

CERTIFICATIONS

- Google Data Analytics Professional Certification
- Advanced C++ certification (Udemy)