

# 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

Master's, Data Science

August 2023 - May 2025

GPA: 3.97

### Savitribai Phule Pune University

Bachelor's, Electronics and Telecommunication Engineering

June 2016 - May 2020

GPA: 7.6

## PROFESSIONAL EXPERIENCE

---

### University of Oklahoma

Graduate Research Assistant

Oklahoma, USA

February 2024 - Present

- Developed automated pipelines for transforming EEG and MRI data into BIDS-compliant structures in Python and MATLAB, increasing research efficiency by 40%.
- Applied unsupervised ML 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

Software Engineer

Pune, Maharashtra, India

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.
- Collaborated directly with clients and cross-functional teams to gather requirements, analyze business processes, and deliver scalable solutions.
- Received the 'On the Spot' Award for consistently delivering projects and achieving a 30% increase in operational efficiency.

## PRACTICUM PROJECT (ONGOING)

---

### Intelligent Research Paper Analysis with Large Language Models

University of Oklahoma

- Developing a LLM-based research analysis system with RAG architecture using LangChain, FAISS and ChromaDB
- Processed literature PDFs into structured datasets for fine-tuning using PyMuPDF and SpaCy for text extraction and segmentation
- Built backend and U.I. for seamless interaction with the system and optimized the model through prompt engineering.
- Fine-tuning the Large Language Model for academic writing and evaluating performance with BLEU, ROUGE and BERTScore.

## PROJECTS

---

### Fake Job Detection Portal - [Link to project](#)

- Applied Natural Language Processing to vectorize job descriptions for training Machine Learning models
- Trained various predictive models using Sci-kit Learn and performed statistical analysis for model evaluation.
- Developed a user-friendly application through Flask with an API-driven prediction and data transformation pipeline

### Predicting Customer Revenue - [Link to project](#)

- Performed data preprocessing and built 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

### Financial Analysis of a Supermarket Chain - [Link to project](#)

- An interactive tableau story to explore the financial performance of a supermarket chain, Wearmart INC
- Provided stakeholders with actionable insights that will enhance sales strategies.
- Identified key products and target demographics to optimize future marketing and sales efforts

### 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 schema, and optimized storage structures
- Implemented indexing for query performance optimization

## SKILLS

---

- **Programming Languages:** Python, R, SQL, HTML/CSS, C/C++, MATLAB
- **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
- Automation 360: Advanced Certification
- C++ Programming from beginner to Advanced 2023(Udemy)