TUSHAR JAYENDRA MHATRE

+1 (405) 223-8789 | tushar.jayendra.mhatre-1@ou.edu | linkedin.com/in/tushar-mhatre24/ | github.com/TusharM24

EDUCATION

University of Oklahoma   **August 2023 - May 2025**
Master's, Data Science   *GPA: 3.97*

Savitribai Phule Pune University   **June 2016 - May 2020**
Bachelor's, Electronics and Telecommunication Engineering   *GPA: 7.6*

PROFESSIONAL EXPERIENCE

University of Oklahoma   **Oklahoma, USA**
Graduate Research Assistant   *February 2024 - Present*
- Developed automated pipeline for transforming EEG and MRI data into BIDS-compliant structures, 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   **Pune, Maharashtra, India**
Software Engineer   *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 enhance predictive analytics
- Collaborated directly with clients 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   **January 2025 - May 2025**
University of Oklahoma
- Developing a LLM-based research analysis system with RAG architecture using Deepseek-R1, LangChain and ChromaDB
- Processed literature PDFs into structured datasets for finetuning using PyMuPDF and SpaCy for text extraction and segmentation
- Built backend and UI for seamless interaction with the analysis system
- Fine-tuning DeepSeek-R1 model for academic writing and evaluating performance with BLEU, ROUGE and BERTScore.

PROJECTS

Fake Job Detection Portal
- Applied Natural Language Processing to vectorize job descriptions for training supervised ML models
- Trained various classification 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.

Financial Analysis of a Supermarket Chain
- 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.

Predicting Customer Revenue
- 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

Database Design for Patient Assistance Network(PAN)
- 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
- **R libraries:** dplyr, tidyr, knitr, ggplot2, caret, h2o, readR, tibble, plotly
- **Statistical Analysis:** Hypothesis Testing, Confidence Intervals, Correlation Matrices, Outlier Detection
- **Database:** MySQL, Database Design, JDBC (Java Database Connectivity), Microsoft Azure, MongoDB, ChromaDB
- **Data Analysis Tools:** Tableau, Microsoft Power BI, Excel
- **Natural Language Processing:** Tokenization, Named Entity Recognition (NER), TF-IDF, Word Embeddings (Word2Vec)
- **AI & NLP Tools:** LangChain, Ollama, Retrieval-Augmented Generation (RAG), Hugging Face
- **Deep Learning:** Image Segmentation, Convolutional neural networks, Object detection

CERTIFICATIONS

- Google Data Analytics Professional Certification
- Automation 360: Advanced Certification
- C++ Programming from beginner to Advanced 2023(Udemy)