Tanishq Tanmay

tanishq51099@gmail.com • (517) 980-4684
LinkedIn • East Lansing, MI 48823

Data Scientist

Detail-oriented Associate Data Scientist with expertise in mathematical modeling, exploratory data analysis, and data visualization. Skilled in Python, SQL, and cloud-based data science solutions. Passionate about leveraging data-driven insights to enhance decision-making and improve operational efficiency. Strong problem-solving and analytical skills, with experience in both structured and unstructured data processing.

Technical Proficiencies

Data Science: Machine Learning, Generative AI, NLP, LLM, AI Agents, Computer Vison, Feature engineering, AWS Cloud-Based AI/ML Services (Redshift, S3, Glue, Sagemaker, lambda), Google Dataflow, Google Vertex AI, ETL Pipelines, Time Series Forecasting, Web Scraping, Statistics and Probability, Hypothesis Testing, ML Model Deployment, Docker, Optimization, Big Data, Git, VS Code, Signal Processing.

Programming: Python, R, SQL, MATLAB, C/C++, Linux.

Libraries: Matplotlib, Pandas, TensorFlow, PyTorch, ONNX, Scikit-learn, LangChain, Llama-Index, Hugging Face, NLTK, ggplot2, Streamlit, GDAL, Geopandas, Rasterio.

Cloud Computing: AWS (Sagemaker, Bedrock), GCP (Dataflow, Vertex AI).

Databases: MySQL, MongoDB, PostgreSQL, Google BigQuery, AWS DynamoDB, AWS RDS, AWS RedShift.

Visualization Tools: Tableau, Looker Studio, Plotly, Seaborn, Altair.

Other: Remote Sensing, ArcGIS, ERDAS, High Performance Computing, Microsoft Excel including Power Query, Word, PowerPoint, Data Structures and algorithms.

Career Experience

Global Observatory for Ecosystem Services (Michigan State University), East Lansing, US **2023 – Present**
Data Scientist

- Collaborate with NASA and WRI on a high-impact initiative to map trees by using remote sensing, contributing to 2030 net-zero carbon emissions goal by implementing U-Net CNN architecture for tree crown segmentation from high-resolution satellite imagery.
- Achieved publication in IOP Science for groundbreaking research on measuring trees outside forests as nature-based solution for net-zero emissions in South Asia, utilizing remote sensing and advanced machine learning techniques. Link.
- Enhanced geospatial analysis precision and carbon evaluation accuracy by 14% by integrating K-Means clustering with advanced computer vision techniques, including edge detection and contour plotting algorithms.
- Developed automated data pipelines for or preprocessing satellite imagery, reducing manual effort by 40% and accelerating tree mapping.
- Developed interactive Tableau dashboards to visualize geospatial data, enabling intuitive exploration of tree cover area output distribution and carbon sequestration potential.

Key Projects and Research Work

Siemens Capstone Project: Multi-Objective Optimization Modeling, East Lansing, US **2023**

- Optimized multi-objective engineering designs in partnership with Siemens by developing rule sets and ML models to define Pareto and rank-based solutions that improved design efficiency by 35%.
- Identified Random Forests as most effective model for predicting Pareto designs through evaluation of Decision Trees, SVMs, and Feedforward Neural Networks, culminating in a detailed analytical report.

Tableau Project: British Airways Review

- Developed an interactive Tableau dashboard analyzing 30,000+ British Airways reviews, providing insights into customer satisfaction, service performance, and sentiment trends.
- Published the project on Tableau Public, making it accessible for visualization and analysis. Project Link

Project: Human Emotion Classification using Machine Learning techniques with Mr. Shashi Kumar G S 2021 – 2022

Research Assistant, Manipal Institute of Technology, Manipal, India

- Boosted emotion prediction accuracy from 70% to 90% by training ML models, including SVM, KNN, Trees, and Logistic Regression, on wavelet-transformed EEG signal data.
- Published research in Measurement Sensors on wavelet-based ML models for classifying human emotions using EEG signals, contributing to advancements in emotion recognition technology. Link

Generative Question Answering with Retrieval Augmented Generation (RAG)

- Enhanced generative question-answering performance by integrating RAG framework with Transformer-based models (T5, Llama 2) and fine-tuning T5 with prompts to achieve a 0.91 F1 score (SOTA) using Pytorch.
- Improved passage retrieval efficiency by 17% by deploying DPR with DistilRoBERTa and MPNet encoders, leveraging ChromaDB for optimized embedding storage and retrieval.
- Elevated retrieval accuracy by 33% through Cross-Encoder ranker integration and implemented query expansion using MISTRAL LLM for advanced text generation.

Bluetooth Device Localization Estimation with Dr. Huacheng Zeng

- Achieved precise indoor device localization using BLE and time series data by applying unsupervised learning techniques, including KMeans, DBSCAN, GLM, GAM, Gaussian Mixture Models, and Autoencoders.
- Enhanced tracking accuracy and robustness in static and dynamic environments through RSSI-based position estimation from multiple BLE beacons.

Education

Master of Science in Data Science, May 2024 | GPA: 3.80/4.0
Michigan State University, East Lansing, MI

Bachelor of Technology, Major in Electronics & Communication & Minor in Embedded System Design, Jul 2022 | GPA: 8.43/10
Manipal Institute of Technology, Manipal, Karnataka, India

Virtual Exchange Program, Subject – Introduction to Biomedical Engineering, May 2020
University of New Brunswick, Canada, Online

Certifications

AWS Certified AI Practitioner – Amazon Web Services (AWS)

Certification course of Google Cloud Big Data and Machine Learning Fundamentals - Coursera

Certification course of AWS Cloud Technical Essentials, and Introduction to Designing Data Lakes on AWS - Coursera

Certification course of Fundamentals of Visualization with Tableau - Coursera.

Certification courses of Machine Learning foundations, Neural Networks & Deep Learning, Improving Deep NN: Hyperparameter Tuning

Regularization & Optimization, and Convolutional Neural Networks - Coursera.