DHRUVRAJ SINGH RATHORE

Austin, Texas | (737)-206-1179 | dhruvrajrathore2011@gmail.com | Portfolio | LinkedIn | GitHub

Data Scientist with 2+ years' experience in machine learning, deep learning, and statistical modeling. Skilled in Python, PyTorch, TensorFlow, and AWS, with expertise in NLP, computer vision, and time-series forecasting.

EDUCATION

Texas A&M University Aug. 2024 – Dec. 2025
Master of Science in Data Science, GPA: 4.0 *College Station, TX*

SRM Institute of Science and Technology Jul. 2018 – May 2022
Bachelor of Technology in Computer Science, GPA: 3.8 *Chennai, India*

PROJECTS

TravelGenie | **GitHub** | *Python, REST APIs, BERT, RAG* March 2025 – May 2025

- Built an automated itinerary planner using Python, REST APIs, and LLMs to generate budget-friendly travel plans.
- Fetched real-time flight, hotel, and attraction data using parallel API calls and ranked results with BM25+embeddings.
- Used RAG to generate itineraries with cost breakdowns, reducing travel planning time from hours to under 30 seconds.

Cotton Field Detector | **GitHub** | *Hackathon, U-NET, Pytorch, Computer Vision* November 2024 – December 2024

- Developed a U-Net **deep learning** model in **PyTorch** to detect cotton fields from satellite images with 92% IoU.
- Applied segmented **image classification** to isolate cotton crops from other vegetation for accurate mapping.
- Achieved 88% segmentation accuracy, automating crop area estimation and reducing manual inspections by 50%

Matastatic Cancer Detection | **GitHub** | *Deep Learning, CNN, Pytorch* November 2024 – December 2024

- Integrated deep learning model using CNNs to classify metastases in histopathological images from the PatchCamelyon (PCam) dataset.
- Applied **data augmentation and batch processing** to upgrade model generalization and training efficiency.
- Achieved an F1 score of 0.8768, demonstrating high accuracy in cancer metastasis detection.

Metro Interstate Traffic Volume | **GitHub** | *Machine Learning, Statistical Techniques* October 2024 – November 2024

- Built a **machine learning predictive models** (Random Forest, Ridge, Lasso, Polynomial Regression) for traffic congestion forecasting, leveraging feature scaling, lag variables, and **time-series decomposition**.
- Engineered temporal and weather features to enhance model interpretability and capture complex traffic dynamics.
- Optimized models using **cross-validation and hyperparameter tuning**, improving MAE by 30%.

EXPERIENCE

Data Engineer December 2022 – June 2024
Draup Business Solutions *Bangalore, India*

- Built **ETL pipelines with PySpark** & SQL on EMR, improving data integrity by 35% and processing speed by 30%.
- Developed an Apache Airflow + Great Expectations pipeline for 200M+ daily logs, cutting inconsistencies by 40% and manual checks by 50%.
- Created a serverless AWS Lambda + DynamoDB system for S3 data retrieval, reducing client response time by 30%.

Data Scientist August 2021 – November 2022
HighRadius Corporation *Hyderabad, India*

- Built **machine learning models** to predict customer payment dates using **gradient boosting**, linear regression, and cross-validation, achieving 75% accuracy and improving cash flow forecasting through stakeholder-aligned insights.
- Improved models with **GridSearchCV, Ridge regularization, and Adam optimizer**, boosting accuracy by 40%.
- Transformed data gathering using **SQL indexing and window functions** for faster retrieval and aggregations, while leveraging lazy evaluation in Python to reduce memory usage, cutting processing time by 50%.

TECHNICAL SKILLS

Programming & Data Science: Python, SQL, Pandas, NumPy, Matplotlib, Seaborn, Scikit-learn, PyTorch, TensorFlow, Statsmodels, NLTK, SpaCy, Shell Scripting

Machine Learning & AI: Model Deployment (MLflow, AWS SageMaker), Neural Networks, RAG, NLP, Time-Series Forecasting, Clustering (K-Means, DBSCAN), Dimensionality Reduction (PCA, t-SNE), A/B Testing

Big Data & Cloud Infrastructure: Spark, PySpark, AWS (EMR, S3, EC2, Lambda, Redshift), Databricks, Redis

Tools & Platforms: Git/GitHub, CI/CD, Apache Airflow, Docker, JIRA, Power BI, Snowflake, DBT