

Dhruvraj Singh Rathore

☎ (737)206-1179 | ✉ dhruvraj_16@tamu.edu | [🌐 LinkedIn](#) | [🐙 Github](#)

SUMMARY

Passionate data science professional with expertise in **machine learning, deep learning, and big data**. Proficient in machine Learning frameworks like **Scikit-learn, TensorFlow, PyTorch, and LLMs**. Experienced in developing AI applications, including personalized research assistants using **RAG and LangChain**. Strong background in ETL automation pipelines, predictive modeling, and data visualization.

EDUCATION

Master of Science in Data Science

Texas A&M University, GPA: 4.0

Aug. 2024 – Dec. 2025

College Station, TX

Bachelor of Technology in Computer Science

SRM Institute of Science and Technology, GPA: 3.8

Jul. 2018 – May 2022

Chennai, India

PROJECTS

Personalized Academic Research Assistant 🤖 | *NLP, RAG, Langchain, LLM*

Dec. 2024 – Jan. 2025

- Built an academic research assistant using RAG and LangChain to retrieve, rank, and summarize papers.
- Used FAISS with SciBERT embeddings for retrieval and fine-tuned BERT for ranking.
- Integrated Ollama3.2 for summarization and multi-turn conversational queries.

Cotton Field Detector 🤖 | *Python, Deep Learning, U-NET, Pytorch, Computer Vision*

Nov. 2024 – Dec. 2024

- Developed an automated method to identify and map cotton crop areas from satellite imagery of the United States using UNET algorithms in PyTorch.
- Performed segmented image classification to isolate cotton crops from other vegetation.
- Calculated total cotton acreage by analyzing pixel coverage of masked areas, providing accurate crop area measurements.

Matastatic Cancer Detection 🤖 | *Machine Learning, Deep Learning, CNN, Pytorch*

Nov. 2024 – Dec. 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 🤖 | *Machine Learning, Statistical Techniques*

Oct. 2024 – Nov. 2024

- Built a traffic congestion model using scikit-learn, with feature scaling, one-hot encoding, and time-series analysis.
- Applied Random Forest, Lasso, Ridge, Linear, and Polynomial Regression with cross-validation.
- Tuned models with GridSearchCV, and found Polynomial Regression most accurate with RMSE and R-squared.

EXPERIENCE

Data Analyst

Dec. 2022 – Jun. 2024

Draup Business Solutions

Bangalore, India

- **Led cross-functional initiative** for the development of a **data quality monitoring system**, leveraging **Airflow** to process 200M+ records daily, resulting in 40% improved data accuracy.
- Engineered robust ETL pipelines using **PySpark on AWS EMR**, optimizing data ingestion workflows and **reducing end-to-end processing time by 45%** while maintaining data integrity.

Data Scientist

Aug. 2021 – Nov. 2022

HighRadius Corporation

Hyderabad, India

- Automated data extraction and preprocessing using **Python and SQL**, reducing analysis time by 50%.
- Built **machine learning models using LightGBM and Random Forest** to predict customer payment dates, improving cash flow forecasting accuracy by 40%.

TECHNICAL SKILLS

Programming & Data Science: Python, SQL, R, SAS, Pandas, NumPy, Matplotlib, Scikit-learn, Shell Script

Big Data & Machine Learning: Spark, PySpark, TensorFlow, PyTorch, LLMs (Large Language Models), BERT, Llama3.2, LangChain, Statistical Modeling, Hypothesis Testing, RAG

Databases & Cloud Computing: MySQL, DynamoDB, Redis, AWS Suite (EMR, S3, EC2, Glue, Lambda)

Tools & Platforms: Git/GitHub, CI/CD, Hugging Face Transformers, Data Build Tool (DBT), Apache Airflow, Docker, MS Excel, Power BI, Snowflake