

Dhruvraj Singh Kathore

☎ (737)206-1179 | ✉ dhruvrajrathore2011@gmail.com | [🌐 LinkedIn](#) | [🐙 Github](#)

EDUCATION

Texas A&M University

Master of Science in Data Science, CGPA: 4.0

Aug. 2024 – Dec. 2025

College Station, TX

SRM Institute of Science and Technology

Bachelor of Technology in Computer Science, CGPA: 3.8

Jul. 2018 – May 2022

Chennai, India

TECHNICAL SKILLS

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

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

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

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

EXPERIENCE

Student Research Assistant

Texas A&M University

Oct. 2024 – Present

College Station, TX

- Conducted **descriptive statistical analysis** on Medicaid data to derive actionable insights using Python and SAS.
- Streamlined graph generation for project presentations with R, **reducing processing time by 30%**.
- Performed **hypothesis testing using T-tests and Pearson correlation** to identify relationships between target columns, and **utilized Q-Q plots** to analyze distribution similarities across datasets over multiple years.

Data Analyst

Draup Business Solutions

Dec. 2022 – Jun. 2024

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 **minimizing end-to-end processing time by 45%** while maintaining data integrity.
- Integrated a serverless data retrieval system configured **AWS Lambda, S3, and DynamoDB**, streamlining ad-hoc client data requests and **reducing response time by 60%**.

Data Scientist

HighRadius Corporation

Aug. 2021 – Nov. 2022

Hyderabad, India

- Built **machine learning models**, utilized **LightGBM and Random Forest** to predict customer payment dates, improving cash flow forecasting accuracy by 40%.
- Proactively assessed and **optimized deployed models for 50+ accounts**, collaborating with senior analysts to enhance performance and accuracy.
- Created Power BI dashboards to visualize predictions, successfully communicating insights to all the non-technical stakeholders.

PROJECTS

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

Dec. 2024 – Jan. 2025

- Built an academic research assistant using RAG & LangChain for fast paper retrieval and summarization.
- Used FAISS + SciBERT for retrieval and fine-tuned BERT for ranking, improving search relevance by 40%.
- Integrated Ollama 3.2 for multi-turn summaries, reducing research time by 60%.

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

Nov. 2024 – Dec. 2024

- Developed a U-Net 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%

Metro Interstate Traffic Volume 🤖 | Machine Learning, Statistical Techniques

Oct. 2024 – Nov. 2024

- Developed a traffic congestion prediction model using scikit-learn, feature scaling, and time-series analysis.
- Evaluated Random Forest, Lasso, Ridge, and Polynomial Regression, optimizing with cross-validation.
- Improved traffic volume prediction accuracy by 30%, enhancing commute time estimates and reducing delays.