Dhruvraj Singh Rathore

📞 (737)206-1179 | ✉ dhruvrajrathore2011@gmail.com | 🔗 Linkedin | 🐙 Github

EDUCATION

Texas A&M University   Aug. 2024 – Dec. 2025
Master of Science in Data Science, CGPA: 4.0   *College Station, TX*
SRM Institute of Science and Technology   Jul. 2018 – May 2022
Bachelor of Technology in Computer Science, CGPA: 3.8   *Chennai, India*

TECHNICAL SKILLS

Programming & Data Science: Python, SQL, Pandas, NumPy, Matplotlib, Scikit-learn, Shell Script
Databases & Cloud Computing: MySQL, NoSQL, Redis, MongoDB, AWS Suite (EMR, S3, EC2, Lambda)
Big Data & Machine Learning: Spark, PySpark, LLMs (Large Language Models), BERT, Llama3.2, LangChain, Predictive Analytics, AWS Sagemaker, RAG
Tools & Platforms: Git/GitHub, CI/CD, Apache Airflow, Docker, Power BI, SnowFlake, Data Built Tool

EXPERIENCE

Data Analyst   Dec. 2022 – Jun. 2024
Draup Business Solutions   *Bangalore, India*

- Designed and deployed high-performance **ETL pipelines** using **PySpark and SQL on AWS EMR**, improving data integrity by 35% and reducing processing time by 30%.
- Developed a data quality monitoring system using Apache Airflow, **automating the processing of 200M+ records daily**, reducing data inconsistencies by 40%, and **cutting manual intervention time by 50%**.
- Implemented **optimized OLAP data models (star/snowflake schemas)** for 20M+ job records, improving query performance by 40% and predictive accuracy by 25%, enabling seamless analytical integration.
- Integrated a serverless data retrieval system using **AWS Lambda, S3, and DynamoDB**, streamlining ad-hoc client data requests and **reducing response time by 30%**.
- Automated JIRA ticketing for data issues, streamlining ETL and model deployment tasks. Reduced manual effort by 40% and sped up issue resolution by 2x, improving workflow efficiency.

Data Scientist   Aug. 2021 – Nov. 2022
HighRadius Corporation   *Hyderabad, India*

- 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%.
- Built machine learning models to predict customer payment dates using **gradient boosting and regression techniques**, achieving an accuracy of 75% and improving cash flow forecasting.
- Enhanced model performance using **GridSearchCV for hyperparameter tuning, Ridge regularization to reduce overfitting, and Adam optimizer**, boosting accuracy by 40%.
- Created Power BI analytical dashboards to visualize customer payment patterns and model outputs, helping non-technical stakeholders make data-driven decisions.

PROJECTS

TravelGenie 🐙 | *Python, REST APIs, BERT, RAG*   Mar. 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.

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.
- Implemented FAISS + SciBERT for efficient document retrieval and fine-tuned BERT for ranking, enhancing search relevance by 40% and reducing research retrieval time by 60%.
- 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 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%