Dhruvraj Singh Rathore

737-206-1179 | dhruvrajrathore2011@gmail.com | Linkedin | Github

Education

Texas A&M University — College Station, TX
Master of Science in Data Science — *Aug. 2024 – Present*

SRM Institute of Science and Technology — Chennai, India
Bachelor of Technology in Computer Science Engineering — *Jul. 2018 – May 2022*

Technical Skills

Programming & Data Science: Python, SQL, R, SAS, Pandas, NumPy, Matplotlib, Scikit-learn, Shell Script
Databases & Cloud Computing: MySQL, DynamoDB, Snowflake, AWS Suite (EMR, S3, EC2, Glue, Lambda)
Big Data & Machine Learning: Spark, PySpark, TensorFlow, PyTorch, LLMs (Large Language Models), BERT, Ollama 3.2, LangChain, Statistical Modeling, Hypothesis Testing
Tools & Platforms: Git/GitHub, CI/CD, Hugging Face Transformers, Apache Airflow, Docker, Power BI

Experience

Student Research Assistant — Oct. 2024 – Present
School of Public Health, Texas A&M University — *College Station, TX*

- Conducted **descriptive statistical analysis** on Medicaid data to derive actionable insights using Python and SAS.
- Automated graph generation for project presentations using 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 — Dec. 2022 – Jun. 2024
Draup Business Solutions — *Bangalore, India*

- Implemented **scalable and robust ETL** data injestion pipelines with **PySpark and SQL in AWS EMR jupyter notebooks**, enhancing reliability and resulted in a **35% improvement** in end-to-end data integrity.
- Led a **cross-functional initiative** to architect a dashboard, processing 200 million **data checks through Airflow to generate 100+ exceptions**; rapid resolution of production issues.
- Integrated **AWS Lambda** with **S3 and DynamoDB** to support ad-hoc client data request.

Data Analyst — Jul. 2022 – Nov. 2022
HighRadius Corporation — *Hyderabad, India*

- Transformed data gathering phase by **simplifying data extraction and preprocessing steps** with **Python and SQL, decreasing total time for analysis by 50%.**
- Developed a keyword matching algorithm to automate matching of claims to deductions, yielding **3x increase in net recovery rates and resulting in savings of approximately $50M.**

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

- Collaborated with multiple Fortune 500 CPG companies to facilitate AR work distributions utilizing **time series data in Python, reducing manual efforts by 4 times.**
- Created predictive models for customer payment date patterns leveraging **machine learning regression models, Bagging, and boosting algorithms like LightGBM**, resulting in a **70% increase in model accuracy.**
- Optimized deployed ML models, boosting automation efficiency by 25% and achieving 35% revenue savings

Projects

Cotton Field Detector | *Python, CNN, U-NET, Pytorch, Computer Vision* — Oct. 2024 – Nov. 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.

Metro Interstate Traffic Volume | *Machine Learning, Statistical Techniques* — Nov. 2024 – Dec. 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.