Dhruvraj Singh Rathore

737-206-1179 | dhruvrajrathore2011@gmail.com | linkedin

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*

Experience

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

- Implemented **scalable and robust ETL** data pipelines utilizing **PySpark and SQL in AWS EMR**, 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.
- Standardized **data mining and data wrangling pipelines**, built a model predicting median base pay for 20M+ job roles and locations, **improving accuracy by 40%** over previous models.
- Integrated **AWS Lambda** with **S3 and DynamoDB** to streamline real-time data ingestion and storage for **scalable analytics workflows.**

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

- Transformed data gathering phase by **simplifying data extraction from relational databases and preprocessing steps** with **SQL and Python, 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.**
- Designed a **Power BI dashboard** to share insights with stakeholders, **reducing decision-making time by 40%.**

Data Science Trainee — 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 **ML regression models, Bagging, and boosting algorithms**, resulting in a **70% increase in model accuracy.**
- Refactored already deployed ML models by optimizing accuracy, precision and recall with hyperparameter tuning, leading to a **25% improvement in automation efficiency and 35% revenue savings.**

Projects

Cotton Field Detector | *Python, CNN, U-NET, Pytorch* — 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.

Metastatic Cancer Detection | *Python, CNN, Pytorch, Git/Github* — 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 enhance model generalization and training efficiency.
- Achieved an F1 score of 0.8768, demonstrating high accuracy in automated cancer metastasis detection.

Skills

Programming: Python, SQL, Pyspark
Database: Postgres, MySQL, MongoDB, DynamoDB, Redis, SnowFlake, DBT
Libraries: Pandas, NumPy, Matplotlib, Scikit-learn
Miscellaneous: Git/GitHub, AWS (EMR, S3, EC2, Lambda), Docker, Apache Spark, Postman, Bash, JIRA