

# Dhruvraj Singh Rathore

737-206-1179 | [dhruvrajrathore2011@gmail.com](mailto:dhruvrajrathore2011@gmail.com) | [linkedin](#)

## EDUCATION

---

### Texas A&M University

*Master of Science in Data Science*

College Station, TX

*Aug. 2024 – Present*

### SRM Institute of Science and Technology

*Bachelor of Technology in Computer Science and Engineering*

Chennai, India

*Jul. 2018 – May 2022*

## EXPERIENCE

---

### Data Analyst

*Draup Business Solutions*

Dec. 2022 – Jun. 2024

*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

*HighRadius Corporation*

Jul. 2022 – Nov. 2022

*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**.
- Designed a **Power BI dashboard** to share insights with stakeholders, **reducing decision-making time by 40%**.

### Data Science Trainee

*HighRadius Corporation*

Aug. 2021 – Jun. 2022

*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**, resulting in a **70% increase in model accuracy**.
- Refactored already deployed ML models by optimizing accuracy 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.

### Metro Interstate Traffic Volume | *Python, Statistical Techniques, Model Development*

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.
- Used pandas, NumPy, Matplotlib, and Seaborn for data analysis; tuned hyperparameters with GridSearchCV.
- Evaluated with RMSE and R-squared; Polynomial Regression showed the best accuracy.

## SKILLS

---

**Programming:** Python, SQL, Pyspark, R Programming

**Database:** Postgres, MySQL, MongoDB, DynamoDB, Redis, Snowflake, DBT

**Libraries:** Pandas, NumPy, Matplotlib, Scikit-learn, PyTorch, TensorFlow

**Miscellaneous:** Git/GitHub, AWS Suite (EMR, S3, EC2, Glue, Lambda), Airflow, Docker