Dhruvraj Singh Rathore

dhruvrajrathore2011@gmail.com| LinkedIn | 7372061179

EDUCATION

Texas A&M University, College Station, TX **August 2024 – December 2025**
Master of Science in Data Science

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

WORK EXPERIENCE

Draup Business Solutions, Bangalore **December 2022 – June 2024**
Data Analyst

- Implemented **scalable and robust ETL** data validation frameworks utilizing **PySpark and SQL in AWS**, enhancing reliability and resulted in a **35% improvement in end-to-end data integrity.**
- Engineered a dashboard by working with cross-functional teams and created **100+ data exceptions by processing 200 million data** leveraging **PySpark and Airflow**, enabling stakeholders to swiftly resolve production data 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 **JIRA** with command center dashboard to optimize ticket creation; assigned data issues to respective data owners, **minimizing the manual effort by 70%** and **improving response time by 3x.**

HighRadius Corporation , Hyderabad **July 2022 – November 2022**
Data Analyst

- 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%.**

HighRadius Corporation , Hyderabad **August 2021 – June 2022**
Data Science Trainee

- Collaborated with multiple Fortune 500 CPG companies to facilitate AR work distributions utilizing **time series data** in **Python**, **reducing manual efforts by 4 times.**
- Analyzed around 50M AR data using **python scripts** and performed **Exploratory Data Analysis for** data exploration.
- 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.**

SKILLS

Programming:	Python, R Programming, Pyspark, Pandas, Numpy, SAS
Visualization:	Power BI, Matplotlib, Seaborn, Tableau, Microsoft Excel
DataBase:	SQL, Postgres, NoSQL, MongoDB, Data Warehouse, Redis
Cloud computing:	AWS Suite (EMR, IAM, S3, EC2, Glue, Lambda), Apache Spark (PySpark), DevOps
Other:	Git/GitHub, CI/CD, Version Control, Airflow, Map-Reduce, LLM
Certification:	Python for Data Analysis and Visualisation, Statistics for Data Science, Machine learning with Scikit Learn, MySQL for Data Analytics and business analytics

PROJECTS

- **Cotton Field Detector**
  In this Hackathon project, the goal is to create a method to automatically identify and map agricultural boundaries and calculate the area covered by cotton crop land from **a satellite generated image of the United States.** We used **UNET Algorithms in PyTorch to perform segmented image classification** and masked only the cotton area ignoring other crops then used the pixels/area covered by the satellite image; we calculated the cotton crop area covered in acres.
- **Credit Card Fraud Detection.**
  Devised a model to **predict fraudulent credit card transaction**s by detecting anomalies in transaction amounts, locations, and other relevant data. Utilized Random Forest Classifier and Logistic Regression, achieving a ROC-AUC score of 97.96% through **hyperparameter tuning** with GridSearchCV.