Ayush Daga

LinkedIn | ayush.01.daga@gmail.com | +15719929782

Education

Master of Science in Data Analytics **Aug 2024 - Present**
George Mason University, Virginia - **GPA: 4.0/4.0**
Relevant Coursework:
Applied Statistics, Principals of Data Mining, Big Data Essential, Operation Research, Health Data Integration

Bachelor of Technology in Computer Science and Engineering **Aug 2018 - May 2022**
R.V College Of Engineering, Banglore

Technical Skills

Programming Languages:	Python, R, SQL, Javascript, React, Angular C/C++
Data Analytics and Visualization:	Regression Analysis, Time Series Forecasting (ARIMA), Random Forest, LASSO, Hypothesis Testing, A/B Testing, Feature Engineering **Visualization:** Power BI, Tableau, Apache Superset, matplotlib, seaborn, ggplot2, Abode Photoshop
Big Data and Data Engineering	Databricks, PySpark,Langchain, Hadoop, Hive, Spark MLlib, HDFS, ETL Pipelines, SQL Optimization, ER Modeling, PostgreSQL
Domain Expertise:	Health Analytics, Optimization (Linear/Integer Programming, Simplex), Data-Driven Decision Making

Work Experience

Student Trainee — Tech Mahindra, Hyderabad, India **April 2022 - Sept 2022**
- Designed and implemented a scalable data visualization platform using MySQL Workbench and Apache Superset, improving insight delivery for over 1,000 users
- Automated Excel-MySQL data pipelines, accelerating data retrieval by 40% and enhancing accuracy for project teams.
- Applied time series analysis and anomaly detection to identify irregularities, reducing anomalies by 25% and improving dataset reliability for strategic planning.
- Managed routine data cleaning and database maintenance, ensuring 99% data integrity and building interactive ML-powered dashboards that increased user engagement by 25%.

Intern — Centre of Excellence, RV College of Engineering, Bangalore, India **Sept 2021 - Nov 2021**
- Developed an Air Canvas application using Python and OpenCV to enable real-time air drawing through motion tracking.
- Enhanced accuracy by converting BGR to HSV color space and applying thresholding, reducing false positives by 30% in varying light conditions.
- Implemented contour detection and coordinate tracking with 95% precision, enabling smooth virtual drawing on canvas.
- Designed a user interface with adjustable ink colors and tools, increasing interactivity and user satisfaction by 20%.

Publications

Development of Internal Data Visualization Platform — India **April 2022 - Sept 2022**
- Developed project management dashboard with MySQL and Apache,boosting oversight efficiency by 30%.
- Implemented dashboard with minimal data pre-processing, reducing development time by 40%.

Projects

Wildfire Risk Prediction and Intensity Assessment — USA **Jan 2025 - May 2025**
- Built a scalable wildfire prediction model using PySpark and Spark MLlib on Databricks, processing 150K+ records.
- Engineered 10+ domain-specific features and generated 60K+ synthetic samples for model robustness.
- Achieved 91% accuracy and 0.73 AUC-ROC using Random Forest; performed full data pipeline development and tuning.
- Utilized Databricks DBFS (NoSQL) for scalable data storage, integration, and retrieval.

Comprehensive Framework for ARV Pricing and Supply Chain Analysis — USA **Aug 2024 - Dec 2024**
- Engineered a relational database using MySQL on AWS RDS to structure 33+ variables from ARV shipment and pricing data, supporting scalable analysis and model integration
- Streamlined data preprocessing in R with automated imputation and formatting pipelines, boosting data reliability and reducing manual prep workload by 40%.
- Built ARIMA and Random Forest models to forecast ARV pricing trends and identify key cost drivers.
- Extracted pricing patterns and logistics insights to support data-driven supply chain decisions.