

Krishna Apurva

(469)-880-8745 | krishnaapurva469@gmail.com | <https://www.linkedin.com/in/krishnaapurva469/> | <https://github.com/krishnaapurva>

EDUCATION

Master of Science, Business Analytics and Artificial Intelligence

May 2025

The University of Texas at Dallas

GPA: 3.9/4.0

Courses: Advanced Statistics for Data Science, Machine Learning, Deep Learning, Predictive Analytics, Big Data, Econometrics and Time Series Analysis, Business Analytics with R, AWS Cloud Analytics

Bachelor of Technology, Information Technology

May 2020

Medicaps University

GPA: 3.8/4.0

SKILLS

Languages: SQL, R, Python (Pandas, NumPy, Matplotlib, PySpark), Spark, Shell Scripting, PL/SQL
Databases: PostgreSQL, MySQL, Oracle, MongoDB, Teradata, Snowflake
ETL & DevOps Tools: Apache Airflow, Talend, SSIS, DataStage, Informatica, Docker, Kubernetes, Jenkins, Git
Cloud: Azure (Synapse Analytics, Data Factory, Data Lake), AWS (Redshift, Glue, S3, EC2)
Data Science: Regression, Classification, Clustering, Neural Network, ANN, Decision Tree, Random Forest
Statistics & Analytics: Tableau, Power BI, Statistical Analysis, A/B Testing, Hypothesis Testing, MS Excel
Certifications: AWS Academy Graduate Data Engineer [Link]

PROFESSIONAL EXPERIENCE

Data Engineer | Innova Solutions

May 2021 - June 2023

- Optimized **ETL pipeline** efficiency by implementing fine-tuned stored procedure based on CTE and Window functions that filtered zero-payment claims pre-staging, cutting processing time by 37.5% per run
- Engineered ETL workflows in Talend with **parallel processing**, optimizing resource utilization to handle peak loads of **40M records** per run, reducing processing time per million records from 8 to 2.5 minutes
- Automated processes using **PL/SQL procedures** to streamline daily tasks such as data backups, indexing, and performance monitoring, resulting in an 18% improvement in database operational efficiency
- Integrated lookup validation framework in Talend, enforcing data consistency by cross-referencing predefined reference tables, preventing invalid records, and reducing data discrepancies
- Containerized the pipeline with Docker, improving deployment efficiency, ensuring consistency across virtual environments, and enabling seamless scalability to handle future data growth

Data Engineer | JJ Enterprises

June 2020 - May 2021

- Built an Airflow ETL pipeline in **Python** to transform and load data into a **Data warehouse**, creating structured views for quick access to 24 months of historical data, accelerating business decision-making
- Designed **Tableau** dashboards to monitor **KPIs** across 700K+ sales transactions, refining inventory planning and pricing strategy, driving a 1.2% rise in business revenue
- Applied **K-means clustering** to segment 6k+ customers based on purchase patterns, leading to a 12% growth in repeat purchases through personalized marketing
- Collaborated with sales and marketing teams to analyze customer feedback and conducted data-driven market research to introduce new products, driving an \$11,000 increase in quarterly sales

ML Engineer Intern | Medicaps University

June 2019 - August 2019

- Leveraged OpenCV for accurate identification and tracking of individuals and moving objects, achieving 75% accuracy in real-time tracking of suspicious activities

PROJECTS

Car Insurance Claim Prediction (R, Models - Regression, Random Forest, Decision Trees, Neural Network)

- Implemented various ML models in R to predict car insurance claims, achieving 82% accuracy with an AUC of 0.61 AUC while addressing class imbalance using oversampling techniques for improved model performance

Olist E-commerce Azure Data Pipeline (ADF, ADLS, Databricks, Synapse Analytics, MongoDB, MySQL)

- Designed and implemented a scalable Azure data pipeline leveraging Medallion architecture to efficiently ingest, transform, and analyze 1M+ records, optimized data storage with ADLS, enabling real-time insights through Azure Synapse and Tableau

HONORS AND ACHIEVEMENTS

Dean's Meritorious Scholarship: Awarded for exceptional academic excellence and leadership in management

Top 1% of 500 groups: Finalist in the national-level Smart India Hackathon