Krishna Apurva

+1 (469)-880-8745 | krishnaapurva469@gmail.com | linkedin.com/in/krishnaapurva469 | github.com/krishnaapurva

EXPERIENCE

Data Engineer | Innova Solutions   May 2021 - June 2023
- Architected high-performance data pipelines using stored procedures, **CTEs,** and **partitioning**, optimizing data flow into **Teradata** and boosting efficiency by ~25%
- Refined **ETL** workflows in Talend with parallel processing and incremental loading, reducing data processing time by ~20% and accelerating reporting
- Implemented data partitioning strategy in PostgreSQL, leveraging intermediate tables for recent data and archiving 24 months of historical records in a **Data Warehouse**, improving report generation by ~45%
- Developed a data validation framework in **Talend**, enforcing schema compliance and lookup validations, reducing ~70% of data discrepancies before ingestion, and strengthening data quality
- Optimized ETL performance by tuning SQL queries, indexing, and parallel execution, improving batch efficiency by ~35% and reducing compute costs

Data Engineer | JJ Enterprises   June 2020 - May 2021
- Orchestrated an automated data pipeline using Apache **Airflow, Python**, and Data Build Tool (dbt), minimizing manual processing time, enabling real-time analytics for sales and customer insights
- Applied **K-means clustering** to segment 6k+ customers based on purchase patterns, leading to a ~15% growth in repeat purchases through personalized marketing
- Designed **Tableau** dashboards to monitor **KPIs** across 700K+ sales transactions, refining inventory planning and pricing strategy, driving a ~10% rise in business revenue

ML Engineer Intern | Medicaps University   June 2019 - August 2019
- Leveraged OpenCV for accurate identification and tracking of individuals and moving objects, enabling real-time monitoring of suspicious activities with ~75% accuracy

PROJECTS

Olist E-commerce Azure Data Pipeline | ADF, ADLS, Databricks, Synapse Analytics, MongoDB, MySQL
- Engineered a scalable Azure data pipeline leveraging Data Factory, Data Lake, and Synapse, implementing the Medallion architecture to process 1M+ records and enhance data ingestion, transformation, and analytics
- Transformed and enriched data in Databricks, performing cleaning, joins, and MongoDB-based enrichment, and stored results in Silver level for efficient querying
- Boosted query performance by 40% and decreased processing time by 30% by integrating Azure Synapse with Tableau using serverless SQL pools and CETAS for real-time insights

Car Insurance Claim Prediction | R, EDA, Data Mining, Data Cleaning, Oversampling, Undersampling
- Balanced a highly imbalanced dataset (6% Claim, 94% No Claim) using oversampling techniques
- Employed various predictive modeling techniques, including Logistic Regression, Random Forest, Decision Tree, and Neural Network, for car insurance claims with 82% accuracy and 61% AUC

Credit Card Fraud Detection System | Docker, Kubernetes, Jenkins
- Built and deployed a real-time data processing workflow using Docker and Kubernetes, processing 15K+ transactions while ensuring seamless orchestration and improving system uptime by 35%
- Streamlined deployment cycles using Jenkins CI/CD pipelines, leveraging virtual machines for enhanced backend support and resource management, achieving 92% high availability

SKILLS

Data Tools:	R, Python, Pandas, Databricks, Spark, Linux, Virtual Machine, Docker, Kubernetes, Git
Databases:	SQL, PostgreSQL, Oracle, MySQL, NoSQL, MongoDB, Teradata, PL/SQL
ETL:	Airflow, Talend, Data Build Tool (dbt), Azure, ADF, ADLS, Synapse Analytics, AWS
Data Analytics:	Data Visualization, Tableau, Power BI, MS Excel, Statistical Analysis
Credentials:	AWS Academy Graduate Data Engineer, Dean's Meritorious Scholarship Recipient

EDUCATION

Master of Science, Business Analytics and Artificial Intelligence   **May 2025**
The University of Texas at Dallas   GPA: 3.9/4.0
Bachelor of Technology, Information Technology   **May 2020**
Medicaps University   GPA: 3.6/4.0