Krishna Apurva

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

EDUCATION

The University of Texas at Dallas **May 2025**
Master of Science - Business Analytics and Artificial Intelligence GPA: 4.0
Courses: *Big Data, Advanced Statistics, Database Foundation for Business Analytics, AWS Cloud Analytics, Applied Machine learning, Business Data Warehouse, Predictive Analytics for Data Science*

Medicaps University **May 2020**
Bachelor of Technology - Information Technology GPA: 3.6

SKILLS

Languages: R, Python (NumPy, Pandas, Matplotlib, Scikit-learn), Apache Spark, Java, Linux, VM, Git
Databases: SQL, PostgreSQL, Oracle, NoSQL, MongoDB, PL/SQL, ER Model
ETL: Airflow, Talend, ETL Pipeline, Apache Kafka, Docker, Kubernetes, DataStage, TAC Server
Big Data/Cloud: Hadoop, Hive, Databricks, Teradata, AWS, Azure, Azure Data Factory, Data Lake
Data Analytics: Data Visualization, Tableau, Power BI, MS Excel, Stata, Statistical Analysis, R Studio

EXPERIENCE

Data Engineer | Innova Solutions May 2021 - June 2023
CAPE – Centene Alternate Payment Engine Project

- Engineered a solution using stored procedures, CTE, and partitioning clauses, complimented by microservices, to filter data flowing into **Data Warehouse** (Teradata), ensuring data integrity and boosting overall efficiency by ~23%
- Leveraged Talend optimization functionalities with **parallel processing** to streamline complex **ETL pipelines**, resulting in ~20% decrease in data processing time and enhancing report generation speed
- Enhanced data storage efficiency by ~35% through designing and implementing a **streamlined Talend workflow** that facilitated **NoSQL** (MongoDB) storage in multi-tiered nested JSON files and shortened query response time by ~25%

Provider RTR Project

- Optimized DataStage **ETL** workflows for healthcare data, reducing processing time by ~27% through strategic redesign of transformation processes and implementing **batch processing** methodologies
- Automated processes using PL/SQL procedures to streamline daily tasks such as data backups, indexing, and performance monitoring, resulting in an ~18% improvement in overall operational efficiency within the database

Data Analyst | J. J. Enterprises August 2020 - May 2021

- Collaborated with stakeholders to gather and analyze business requirements**,** developing ad-hoc **SQL** queries and designing workflow management processes in Tableau
- Led **data-driven market research** to inform new product introductions, contributing to ~8% quarterly revenue growth
- Designed and maintained multiple **Tableau dashboards** for real-time monitoring of key performance indicators (**KPIs**), refining customer segmentation and contributing to a ~10% rise in business revenue

Data Science Intern | Medicaps University June 2019 - August 2019

- Achieved precise identification and tracking of individuals and moving objects through the implementation of OpenCV, enabling **real-time monitoring** of suspicious activities at ~78% accuracy
- Orchestrated the development of the Yolo V3 **object detection algorithm** and achieved the best performance with a processing speed of 45 frames per second, surpassing algorithms by a 200% faster rate

PROJECTS

Car Insurance Claim Prediction | R, EDA, Data Mining, Data Cleaning, Oversampling, Undersampling

- Balanced dataset by addressing a 6% (Claim) and 94% (No claim) data imbalance through oversampling techniques
- Employed various predictive modeling techniques, including Logistic Regression, Random Forest, Decision Tree, and Neural Network, for Car Insurance Claim prediction with 82% accuracy and 61% AUC

Tokyo Olympic Azure Data Pipeline | Azure Data Factory, Azure Data Lake, Databricks, Power BI

- Engineered data ingestion using **ADF**, automated extraction of 11k records from GitHub to **Azure Data Lake**
- Transformed and processed large datasets using **Azure Databricks** and **Spark** and developed **Power BI** dashboards, enabling data-driven insights and informed decision-making

HealthCare Management System | PostgreSQL, Entity Relationship Modeling, Database Design

- Integrated hospital operations with ER Modeling, DDL/DML implementation, and advanced queries, Triggers, Procedures, Views, and Functions for efficient healthcare management system reporting