Dheeraj Rahul Reddy Piduru

Dallas, Texas, 75252 | +1 (945-308-0889) | pidurudheeraj52@gmail.com | LinkedIn

EDUCATION

The University of Texas at Dallas **Dec 2025**
Master of Science, Business Analytics GPA: 3.72
Relevant Coursework: Advanced Statistics for Data Science, Applied Machine Learning, Applied Deep Learning, Predictive Analytics for Data Science, Prescriptive Analytics, Big Data Analytics, Applied Econometrics

Mahindra University, Hyderabad **Aug 2023**
Bachelor of Technology, Computer Science & Engineering GPA: 3.00

TECHNICAL SKILLS

Languages: Python, R, SQL
Data Science Libraries: Pandas, NumPy, Scikit-Learn, TensorFlow, Keras, PyTorch, XGBoost
Data Analysis & Visualization: Matplotlib, Seaborn, Plotly, Tableau, Power BI, Excel
Databases: MySQL, PostgreSQL, MongoDB, Snowflake
Machine Learning Ops & Tools: MLflow, Databricks, Git, GitHub Actions, Jenkins, Airflow
Cloud Platforms: AWS (EC2, S3, RDS, Lambda, Redshift), Azure, Docker, Kubernetes

WORK EXPERIENCE

Business Analyst Intern | *Skiaverse Private Limited, Hyderabad, India* **Jan 2023 – Aug 2023**

- Utilized **Python (Pandas, NumPy, Scikit-learn)** and **SQL** to preprocess, engineer features, and analyze over **50,000 rows** of real-time market data, uncovering behavior trends that drove a region-specific Go-To-Market (GTM) strategy.
- Developed and deployed a predictive model using **XGBoost and logistic regression** to simulate market responses, contributing to a **10% increase in customer engagement** across diverse product segments within the first quarter.
- Applied **unsupervised clustering** (K-means, PCA) to identify high-conversion customer segments and collaborated with marketing to optimize messaging strategies, improving **marketing ROI by 12%** across targeted campaigns.
- Designed a dynamic **Tableau dashboard** integrated with **AWS Redshift** and **MLflow** to automate reporting on 7 KPIs, improving operational visibility by **15%** and reducing manual reporting time by 8+ hours weekly.
- Built reproducible ML pipelines on **Databricks** and orchestrated batch ETL workflows with **Apache Airflow**, ensuring scalable analytics delivery from raw ingestion to model evaluation and dashboard integration.

PROJECTS

Cloud Data Integration Pipeline | *PostgreSQL, Snowflake, dbt, EC2*

- Created a scalable end-to-end ETL framework in Python and SQL to integrate PostgreSQL with Snowflake, reducing data processing time by **35%** through dbt incremental model execution and schema optimization.
- Prepared and implemented schema-aware transformations using modular dbt SQL logic, improving data consistency and reducing redundant query executions by over **40%** across daily batch workflows.
- Built anomaly detection dashboards with Pandas and automated EC2-based health checks with Slack alerting, ensuring real-time monitoring and maintaining **99%** pipeline uptime in production environments.

Real-time Social Media Sentiment Analyzer | *NLP, Apache Spark, Kafka, Apache Flink*

- Built end-to-end streaming pipelines using Kafka and Apache Flink to classify over **1K+ social media posts per minute** from Twitter and Reddit, reducing insight latency by **30%** and enabling faster campaign adjustments.
- Integrated pretrained NLP models in PySpark with TextBlob for real-time sentiment classification, improving language processing accuracy by over **25%** and enabling immediate categorization of diverse content streams.
- Applied keyword clustering and topic modeling using Scikit-learn to track brand sentiment shifts in real-time, allowing marketing teams to react with **40% faster response** to trending consumer feedback and news cycles.
- Developed a live dashboard using Apache Spark Structured Streaming and Plotly Dash to visualize emerging sentiment trends, reducing manual tracking effort by over **50%** and enabling data-driven marketing interventions.

Dynamic Data Lakehouse for Customer Insights | *Apache Iceberg, Parquet, Spark, Airflow*

- Designed a scalable Iceberg-based data lakehouse architecture to support time-travel queries on historical snapshots, reducing exploratory data analysis (EDA) runtime by **2x** across large-scale datasets.
- Tuned complex Spark SQL transformation jobs with optimized Parquet partitioning and predicate pushdown, accelerating query performance by **45%** while minimizing compute resource utilization.
- Automated ingestion of behavioral and transactional data from **15+ microservices** using Apache Airflow DAGs, improving workflow reliability and reducing data ingestion failures by over **60%**.
- Performed cohort segmentation on processed datasets using Python and Pandas, enriching customer journey analytics and increasing report clarity and business insight depth by **30%**.