

Sreedithya Dulloor

sreedithyadulloor@gmail.com | (774) 271-4784 | [LinkedIn](#)

SUMMARY

Data Science graduate with a strong foundation in data analytics, statistical modeling, and business intelligence. Proficient in Python, R, SQL, Tableau, and Power BI, with hands-on experience analyzing large datasets and developing data-driven insights. Skilled in data visualization, exploratory data analysis (EDA), and predictive modeling to support strategic decision-making. Experienced in using Azure Databricks for big data processing and leveraging machine learning techniques to uncover trends. Passionate about transforming raw data into actionable insights that drive business efficiency and growth.

SKILLS

Languages & Tools: Python, R, SQL, Go, C++, Java, MATLAB, MySQL, Azure Databricks, AWS, Snowflake

Machine Learning & Big Data: PySpark, TensorFlow, Keras, Scikit-learn, Spark ML, Hadoop, MPI

Frameworks & Platforms: MLflow, Spark, AWS SageMaker, Tableau, Power BI, Jupyter, RStudio

Core Strengths: Predictive Modeling, Deep Learning, A/B Testing, Debugging & Testing (Unit, Canary, Smoke), Object-Oriented Design, System Architecture, Model Deployment, MLOps Integration

ACADEMIC PROJECTS

Predictive Analysis of US Domestic Flight Delays and Cancellations

- Processed over 1.5 million flight records from Kaggle using Python, PySpark, and Azure Databricks, optimizing big data processing for machine learning models.
- Developed predictive models using Linear Regression, Random Forest, and Gradient Boosting, achieving a lowest RMSE of 36.37 with Gradient Boosting, improving prediction accuracy for flight delays.
- Identified key delay predictors, including TAXI_OUT time and DISTANCE, providing actionable insights for airline scheduling improvements.
- Created interactive Tableau dashboards visualizing flight delay trends, assisting in optimizing airline operations and resource allocation.

Movie Data Analysis & Recommendation System

- Analyzed a large-scale movie dataset with thousands of records using Azure Databricks and PySpark to uncover trends in viewer preferences.
- Built a recommendation system using the Alternating Least Squares (ALS) algorithm, achieving high user preference prediction accuracy.
- Identified top-rated genres such as Drama, Comedy, and Action, contributing to strategic content planning in entertainment.
- Conducted user engagement analysis, revealing that top 20% of users contributed over 50% of ratings, emphasizing the role of power users in movie recommendations.

Parallelizing Music Recommendation System Using MPI

- Optimized a music recommendation system using Message Passing Interface (MPI), reducing computation time by over 50% on large datasets.
- Implemented machine learning models including k-NN, SVM, Decision Trees, and Random Forests, maintaining comparable accuracy to serial implementations while improving scalability.
- Parallel execution on 4 and 8 processors significantly decreased processing time, demonstrating the feasibility of parallel computing for real-time recommendations.
- Findings applicable to streaming services, enhancing user experience by reducing recommendation latency and improving scalability.

EDUCATION

UNIVERSITY OF MASSACHUSETTS DARTMOUTH

Master of Science in Data Science, GPA: 3.83/4

North Dartmouth, MA

May 2024

OSMANIA UNIVERSITY

Bachelor of Science in Mathematics, Statistics and Computer Science, GPA: 3.65/4

Hyderabad, India

September 2021

LEADERSHIP

UNIVERSITY OF MASSACHUSETTS DARTMOUTH

September 2022 – May 2024

- Big Data Club – Member. Actively participating in meetings and working on interesting projects for campus departments, outside companies or just to add to my personal portfolio.
- Digital Scholarship Hub (DiSH) – Member. Working on engaging with and creating knowledge outside of a traditional class setting. We develop and present projects on data and digital content.