

S SHYAMAL

✉ shyamal.0423@gmail.com ☎ 9390244256 📍 Chittoor, India

PROFILE

Detail-oriented Data Analyst with hands-on experience in SQL, data preprocessing, and exploratory data analysis. Skilled in transforming raw data into actionable insights using Python and data visualization tools. Experienced in building data-driven solutions and communicating insights to stakeholders, with a strong foundation in machine learning and big data technologies.

EDUCATION

The Apollo University, 2022 – present | Chittoor
Btech in CSE (Artificial Intelligence and Data Science)
CGPA: 8.67

SKILLS

Programming: Python
Libraries & Tools: Pandas, NumPy, Matplotlib, Scikit-learn
Data Analysis: Data Cleaning, EDA, Data Visualization, Feature Engineering
Big Data: Apache Spark, Hadoop Ecosystem
AI & Automation: Agentic AI Systems, LLM-based Workflows, Automated ML Pipelines
Other Tools: MLflow, Gradio, FastAPI, Docker

PROFESSIONAL EXPERIENCE

The Apollo Institute of Medical Science and Research, 05/2024 – 06/2024 | Chittoor
Data Analyst Intern

- Extracted and analyzed datasets using SQL to generate key business insights.
- Presented data-driven reports to leadership and stakeholders.
- Collaborated with cross-functional teams to collect and prepare data for analysis.
- Trained team members on data analysis best practices and tools.

PROJECTS

Meta AI – Automated Machine Learning Pipeline Generator | IRJET (IF: 8.315), Mar 2026

- Developed an end-to-end AutoML system for tabular data, automating data preprocessing, feature engineering, and model training.
- Performed hyperparameter optimization using Optuna and evaluated models using performance metrics to ensure reliability.
- Applied SHAP-based explainability and fairness auditing to interpret model predictions and detect bias.
- Built an interactive dashboard for data analysis, experiment tracking (MLflow), and workflow reproducibility.

Recommender System Development using MovieLens Dataset

- Developed and evaluated recommender systems using MovieLens dataset (100K ratings, 9K movies) with Popularity-based, User-User Collaborative Filtering, and Matrix Factorization methods.
- Implemented a modular pipeline using Pandas, NumPy, SciPy, and scikit-learn for data preprocessing, training, and testing.
- Achieved best results with Matrix Factorization, evaluating performance with RMSE, Recall@K, and NDCG@K metrics.

Flight Cancellation Prediction

- Data Cleaning & Preprocessing: Handled missing values, removed outliers, encoded categorical variables, and scaled features.
- Exploratory Data Analysis: Uncovered trends between cancellations and factors like weather, delays, and departure time.
- Model Building: Trained and evaluated Logistic Regression, Decision Tree, Random Forest, and SVM classifiers.

CERTIFICATES

- Python Programming by Sololearn
- Data Science For Engineering by NPTEL
- Acquiring Data by Accenture
- Data Processing and Visualization by Accenture