

YASH SUD

Data Analyst | Data Science & Analytics

+91 6202958366 | ysud.02@gmail.com | linkedin.com/in/yash-sud-800aaa278 | github.com/YashSud

PROFESSIONAL SUMMARY

Data Analyst and Data Scientist with an M.Sc. in Artificial Intelligence and hands-on experience in data analysis, data analytics, data preprocessing, data cleaning, exploratory data analysis (EDA), SQL, Python, and statistical analysis. Experienced in analyzing large datasets, building automated data pipelines, and deriving actionable, data-driven insights using Pandas, NumPy, and Matplotlib. Skilled at data visualization, report generation, and applying machine learning and data science techniques to support data-driven business decision-making.

TECHNICAL SKILLS

Data Analysis & Data Science: Data Analysis, Data Analytics, Data Science, Pandas, NumPy, Matplotlib, Data Cleaning, Data Wrangling, Exploratory Data Analysis (EDA), Statistical Analysis, Regression Analysis, Model Comparison

Programming & Databases: Python, SQL Querying (MySQL, Oracle SQL), Java, Relational Databases

Visualization & Business Intelligence: Data Visualization, Matplotlib, Microsoft Excel, Data Reporting, Dashboards, KPI Reporting, Trend Analysis

Machine Learning: Scikit-learn, TensorFlow, Keras, Regression, Classification, Model Evaluation, Predictive Analytics

Tools: Git, GitHub, Jupyter Notebook, Apache Hadoop

PROFESSIONAL EXPERIENCE

AI/ML Intern — CSIR–CRRRI (Central Road Research Institute)

Jan 2025 – Jun 2025

- Performed data analysis on 2,000+ traffic images and associated datasets to support AI model development and research reporting.
- Conducted data preprocessing, annotation, validation, and quality checks as part of the exploratory data analysis (EDA) process, improving dataset consistency and reliability for downstream analysis.
- Automated dataset preparation and analytical reporting workflows using Python, reducing manual validation time by 40%.
- Analyzed model performance using statistical evaluation metrics and collaborated with researchers to identify data-driven improvements.

Internship — AD Infocom Systems

Jan 2025 – Jul 2025

- Cleaned and preprocessed 5,000+ image records using Python, Pandas, and NumPy as part of a broader data analytics workflow, improving overall data quality and consistency.
- Built automated data preprocessing pipelines, reducing data preparation time by 25%.
- Performed data analysis on model outputs and performance metrics to identify optimization opportunities and improve processing efficiency.
- Generated analytical reports and data visualizations evaluating preprocessing effectiveness to support project decision-making.

PROJECTS

Regression Analysis — Housing & Health Datasets

- Built a machine learning notebook analyzing California Housing, Boston Housing, and Diabetes datasets; trained, evaluated, and statistically compared Linear and Ridge regression models using Scikit-learn.

AI Research Assistant — RAG Pipeline

- Built a data pipeline that ingests and semantically indexes PDF documents into a ChromaDB vector database, enabling accurate, citation-grounded query responses via GPT-4o.

EDUCATION

M.Sc. in Artificial Intelligence

2023 – 2025

Central University of South Bihar, Gaya | CGPA: 7.58/10.0

Bachelor of Computer Application

2020 – 2023

Patliputra University, Patna | CGPA: 6.63/10.0

CERTIFICATIONS & ACHIEVEMENTS

- Agentic AI Development — Udemy | Python for Data Science & Machine Learning — Udemy
- JavaScript Development — Geekster | C Programming Certification — IIT Bombay
- Event Organizer, Software Freedom Day — Central University of South Bihar (2024)