

# Soumyadeep Das

+91 6291227288 | [soumyadeepdas@gmail.com](mailto:soumyadeepdas@gmail.com) | Kolkata, West Bengal, India

[Linkedin](#) | [Github](#) | [Kaggle](#)

## SUMMARY

Computer Science undergraduate with strong hands-on experience in Python, data analysis, and machine learning. Built end-to-end data-driven and ML projects involving data preprocessing, model development, evaluation, and deployment using tools like Pandas, Scikit-learn, SQL, and FastAPI. Seeking an entry-level role or internship in Data Analytics, Python Development, or Machine Learning to apply skills to real-world problems.

## SKILLS

- **Programming Languages:** Python
- **Data Science & Analytics:** NumPy, Pandas, Matplotlib, Exploratory Data Analysis (EDA), Feature Engineering, Data Cleaning & Preprocessing
- **Machine Learning:** Scikit-learn (Regression, Classification, Clustering)
- **Backend & APIs:** FastAPI
- **Databases & Storage:** MySQL
- **DevOps & Version Control:** Git, GitHub

## PROJECTS

- **End-to-End Customer Churn Prediction System** Mar 2025 – Apr 2025  
*Tools: Python, Pandas, Scikit-learn, Machine Learning, Streamlit* [\[GitHub\]](#)
  - Built an end-to-end machine learning pipeline on **7,000+ telecom customer records** to predict customer churn in an **imbalanced dataset**
  - Performed feature engineering and applied **SMOTE** for class imbalance, evaluating multiple models including **Logistic Regression, Random Forest, and XGBoost**
  - Achieved **ROC-AUC of 0.84** and **F1-score of 0.62**, and deployed the final model using **Streamlit** for real-time churn prediction
- **E-Commerce Analytics Platform – Customer & Revenue Insights** Jan 2025 – Feb 2025  
*Tools: Python, Pandas, NumPy, SQL (MySQL), SQLAlchemy, Matplotlib* [\[GitHub\]](#)
  - Built an end-to-end analytics pipeline to clean retail transaction data and store it in a normalized MySQL database using Python
  - Performed RFM-based customer segmentation and churn analysis to identify high-value, loyal, and at-risk customers
  - Analyzed revenue trends, seasonality, and customer lifecycle patterns to derive business insights
  - Evaluated product performance using Pareto (80/20) and basket analysis to identify top products and cross-sell opportunities
- **Sales Demand Forecasting using Time-Series Analysis** Apr 2024 – May 2024  
*Tools: Python, Pandas, NumPy, Matplotlib, Statsmodels (ARIMA), Prophet* [\[GitHub\]](#)
  - Performed time-series analysis on **3+ years of retail sales data** to identify **trend, seasonality, and cyclic patterns**
  - Aggregated **daily sales into monthly time-series** and engineered lag-based features for forecasting
  - Built and compared **ARIMA and Prophet models** for multi-step sales forecasting
  - Evaluated forecasting performance using **MAE and RMSE**, and visualized **actual vs. predicted sales trends**
- **Chatgot – RAG-Based AI Tutor for MAKAUT Computer Networks** Mar 2025 – Present  
*Tools: Python, FastAPI, FAISS, Sentence-Transformers* [\[GitHub\]](#)
  - Built an end-to-end **RAG-based AI tutor** covering **100+ syllabus topics** for MAKAUT B.Tech CSE Computer Networks
  - Enabled **fast semantic retrieval (sub-100ms)** using Sentence-BERT embeddings indexed with FAISS
  - Improved system efficiency by **filtering non-academic queries** using a lightweight NLP classifier, exposed via a FastAPI backend

## EDUCATION

- **B.P. Poddar Institute of Management & Technology** 2022 – Present  
*Bachelor of Technology (B.Tech) in Computer Science and Engineering* Kolkata, India
  - **Relevant Courses:** Linear Algebra, Probability and Statistics, Discrete Mathematics, Data Structures and Algorithms, Machine Learning, Artificial Intelligence, Database Management Systems
  - Current CGPA (upto 7th sem): 8.06/10
- **Ballygunge Government High School** May 2022  
*Higher Secondary Education* Kolkata, India
  - Grade: 83.8%