Rahul Ghoshal

Panchla, Howrah, India
imrahulghoshal@gmail.com — +91 6289856154 — Portfolio — GitHub

Professional Summary

Data Science postgraduate with strong foundations in machine learning, statistics, and data analytics. Hands-on experience in predictive modeling, NLP-based sentiment analysis, and data visualization using Python, SQL, R, and Power BI, with internship experience at ISI Kolkata.

Education

Master of Science in Data Science *2024 – 2026*
JIS University, Agarpara, Kolkata
CGPA: **9.57 / 10.0**[First Year]

Bachelor of Science in Computer Science (Honours) *2021 – 2024*
Vivekananda College, Madhyamgram, Kolkata
CGPA: **8.97 / 10.0**

Technical Skills

- **Programming Languages:** Python, R, SQL.
- **Libraries & Frameworks:** Pandas, NumPy, Scikit-learn, NLTK.
- **Databases:** MySQL.
- **Data Visualization:** Power BI, Matplotlib, Seaborn.
- **Machine Learning:** Regression, Classification, Feature Engineering, Exploratory Data Analysis.

Internship Experience

Data Science Intern *May 2025 – July 2025*
IDEAS TIH, Indian Statistical Institute (ISI), Kolkata

- Collected, cleaned, and analyzed multi-platform customer reviews (Website, Instagram, Facebook) from FY 2022–2025 to build an end-to-end sentiment analysis pipeline.
- Implemented VADER-based NLP models with custom compound score logic to classify feedback into five sentiment categories.
- Detected a **27.2%** drop in "Highly Positive" sentiment (**81.6%** → **54.4%**) in FY2023–24.
- Performed year-wise and platform-wise trend analysis to identify key customer satisfaction patterns.
- Developed data visualizations and insight dashboards using Matplotlib and Seaborn to support stakeholder decision-making.

Projects

Hypertension Prediction System (GitHub Link)

- Built a Random Forest–based hypertension prediction system on 1,980+ patient records across 3 classes (Normal, Prehypertension, Hypertension), achieving **97.4% classification accuracy**.
- Processed and engineered clinical data by performing cleaning, feature encoding, EDA, and correlation analysis using Pandas and NumPy.
- Developed an interactive Streamlit web interface enabling real-time risk prediction from multiple clinical parameters.

Car Price Prediction (GitHub Link)

- Built a car resale price prediction model using Random Forest Regressor on 300+ vehicle records, achieving **R-squared score of 0.96**.
- Performed data cleaning, categorical encoding, and feature engineering including vehicle age.
- Analyzed and modeled non-linear relationships between price and key attributes such as fuel type, transmission, and ownership history to improve prediction accuracy.

Publication

Banerjee, S., **Ghoshal, R.** (2025). *Natural Language Processing in Revolutionizing Clinical Documentation.*