

# ROUDRANIL DAS

📍 Kolkata, West Bengal, India

🌐 <https://roudranil.github.io/>

✉️ [dasroudranil@gmail.com](mailto:dasroudranil@gmail.com)

🌐 [roudranil-das](#)

🌐 [Roudranil](#)

📞 +91-9874137286

## WORK EXPERIENCE

---

### Data Scientist

**Kantar** (Data Science and Innovation, Kantar Analytics Practice)

June 2024 - (ongoing)

Chennai, Tamil Nadu, India

- Using multiple techniques for synthetic data generation, and engaging in R&D for techniques to evaluate quality of synthetic data

## EDUCATION

---

**M.Sc. Data Science, Chennai Mathematical Institute**

CGPA: 9.25

2022 Aug - 2024 Apr

Chennai, Tamil Nadu, India

**B.Sc. Mathematics (Hons), St. Xavier's College (Autonomous), Kolkata**

CGPA: 8.46

2019 July - 2022 May

Kolkata, West Bengal, India

## INTERNSHIPS

---

**Data Science Summer Intern, LTIMindtree**

May 2023 - July 2023

Tools used: [Python](#), [pandas](#), [spaCy](#), [BERTopic](#), [FLASK-Restful](#), [Streamlit](#)

- Utilized topic modelling techniques to get topic clusters and implemented end-to-end RESTful API service for the pipeline.

## PROJECTS

---

**Finetuning open source LLM's for conversation in Shakespearean English**

Dec 2023

Tools used: [Python](#), [PyTorch](#), [huggingface transformers](#), [PEFT](#), [LoRA](#), [scrapy](#), [pandas](#)

[🔗 Link](#)

- Scraped and curated a custom dataset for chat models in modern and Shakespearean english.
- Finetuned 3 open source LLM's including Mistral-instruct-7B on this dataset with parameter efficient fine tuning using LoRA and 4bit quantization.

**Building a n-gram Language Model and word vectors**

August 2023 - Nov 2023

Tools used: [Python](#), [PyTorch](#), [nltk](#)

[🔗 Link](#)

- Prepared the corpus using tokenization and data cleaning and built a 4-gram language model for next word prediction and sentence generation.
- Built word vectors on the same corpus using **Co-occurrence Analogue to Lexical Semantics** and computed the closest words and meaning.

**Comparative Analysis of Decision trees, Naive Bayes and Ensemble models**

Jan 2023 - Feb 2023

Tools used: [Python](#), [scikit-learn](#), [matplotlib](#), [pandas](#), [optuna](#)

[🔗 Link](#)

- Performed feature engineering to improve predictive metrics (accuracy and recall) by 3-4%.
- Carried out hyperparameter tuning for both models and cost complexity pruning for decision tree to further improve predictive metrics and generalizability.

## TECHNICAL SKILLS

---

**Languages** Python, MySQL, R, C, L<sup>A</sup>T<sub>E</sub>X

**ML/AI** NumPy, Pandas, Matplotlib, Scikit-Learn, ggplot2, PyTorch, BERTopic, spaCy, NLTK, huggingface transformers

**Misc** Git, Flask, Streamlit, Linux, BeautifulSoup, Scrapy, Docker, Pytest, Unittesting, MIFlow, Databricks

## KEY COURSES TAKEN

---

**Postgraduate:** Machine Learning, Deep Learning and Advanced ML, Statistics and Visualisation with R, Regression techniques, Natural Language Processing, Computer Vision, MLOps, Intro to GenerativeAI, Data privacy, RDBMS and SQL, Python and Data Structures, Design and Analysis of algorithms

**Undergraduate:** Linear algebra, Real analysis, Multivariable Calculus, Differential equations, Mathematical Statistics, R, C

## AWARDS AND ACHIEVEMENTS

---

**2024** Reliance Foundation Postgraduate Scholarship Awardee

**2019** INSPIRE Scholarship for Higher Education, Dept. of Science and Technology, Government of India

**2017, 2019** Jagdish Bose National Science Talent Search Scholarship Awardee