Roudranil Das

Kolkata, West Bengal, India https://roudranil.github.io/
roudranil@cmi.ac.in roudranil-das
Roudranil +91-9874137286

Education

2022 - 2024 (ongoing)
Tamil Nadu India

M.Sc. Data Science
Chennai Mathematical Institute
SGPA (after semester 2): 9.25

2019 - 2022
West Bengal India

B.Sc. Mathematics (Honours)
St. Xavier's College (Autonomous), Kolkata
CGPA: 8.46

Internships

Aug 2023 - Nov 2023

Machine Learning Intern - Hodu Solutions (as industry project at CMI)
Tools used: Python, pandas, yahoo-finance
– Performed fundamental analysis and intrinsic valuation of companies traded on the US markets
– Analysed historical price and trading data of above companies and derived technical indicators
– Analysed performance of common price-action trading strategies in order to rank the strategies and build an ensemble model

May 2023 - July 2023

Data Science Summer Intern - LTIMindtree
Tools used: Python, pandas, spaCy, BERTopic, FLASK-Restful, Streamlit
– Conducted statistical analysis on raw text data and developed a comprehensive data cleaning pipeline tailored for future applications. Implemented data quality checks based on statistical estimates.
– Utilized topic modeling techniques to extract optimum cluster values and identify topic clusters, providing valuable insights for assessing data quality.
– Incorporated the data quality checking pipelne as a RESTful API service, enabling end-to-end capabilities for efficient data processing and access.

Projects

Dec 2023
Link

Finetuning open source LLM's for conversation in Shakespearean English
Tools used: Python, PyTorch, huggingface transformers, PEFT, LoRA, scrapy, pandas
– Scraped and curated a custom dataset for conversational 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.

Dec 2023
Link

Function calling with OpenAI and open-source LLM's
Tools used: Python, openai, instructor, pydantic
– Built a pipeline for structured output from `GPT` models using function calling in OpenAI's api.
– Implemented a custom schema for validating output structure using pydantic.

August 2023 - Nov 2023
Link

Building a n-gram Language Model and word vectors
Tools used: Python, PyTorch, nltk
– 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-occurance Analogue to Lexical Semantics** and computed the closest words and meaning

July 2023 - ongoing
Link

Optical Music Recognition - recognition of components of sheet music
Tools used: Python, openCV, matplotlib
– Collected a repository of sheet music pdf and corressponding embedded metadata.
– Performed segmentation of the staves and staff lines of a printed sheet music pdf file using openCV.

Jan 2023 - Feb 2023
Link

Comparative Analysis of Decision trees, Naive Bayes and Ensemble models
Tools used: Python, scikit-learn, matplotlib, pandas, optuna
– Analysed performances of the algorithms in both classification and regression tasks.
– Performed feature engineering to improve predictive metrics (accuracy and recall) by 3-4 percentage points.
– Carried out hyperparameter tuning for both models and cost complexity pruning for decision tree to improve predictive metrics and generalisability further.

Dec 2022
Link

Audio Classification with CNN-LSTM networks
Tools used: Python, PyTorch, torchaudio, pandas
– Analysed audio samples of the words "zero" to "nine"
– Trained a baseline CNN, a CRNN and a parallel CNN-RNN model with the CNN-RNN model achieving 84% accuracy on test set (+15% over CRNN) and F1-score of 84%.

Oct 2022 - Dec 2022
Link

Deforestation and Forest Conversion
Tools used: R, ggplot, tidyverse, R-Shiny | Chennai Mathematical Institute
– Created visualisations to better understand causes for deforestation. [Report]
– Built a dashboard web app with `R Shiny` for the above. [Dashboard]

Technical Skills

Languages `Python`, `MySQL`, `R`, `C`, LaTeX
ML/AI NumPy, Pandas, Matplotlib, Scikit-Learn, ggplot2, PyTorch, BERTopic, spaCy, NLTK
Misc Minitab, Flask, Streamlit, Linux

Key Courses Taken

Postgraduate Machine Learning and ML theory, Deep Learning and Advanced ML*, Statistics and Visualisation with `R`, Regression techniques*, NLP*, RDBMS and SQL, Python and Data Structures, Design and Analysis of algorithms

* *indicates ongoing courses*

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

Awards and Achievements

2023 **Reliance Foundation Postgraduate Scholarship Awardee**

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

2019 **Jagadish Bose National Science Talent Search Senior Scholarship Awardee**

2017 **Jagadish Bose National Science Talent Search Junior Scholarship Awardee**