

I'm deeply passionate about statistics and data science, with a solid foundation in Python and practical industry experience. My expertise lies in exploratory data analysis, data visualization, data pre-processing and machine learning. I'm actively cultivating my software engineering skills, aiming to bridge the gap between data science and development. Alongside technical proficiency, my communication and presentation skills allow me to effectively convey complex insights to diverse audiences, ensuring data-driven decisions lead to actionable results.

## PROFILE SUMMARY

- **2 Years of Diverse Data Science Experience:** Proficient in statistical modeling, machine learning and time series analysis using Python with a solid two-year track record of hands-on application.
- **Industrial Expertise:** Over 1 year and 6 months as a Junior Data Scientist, I've developed and deployed machine learning models in cloud environments and cultivated software engineering skills.
- **Current Project:** Currently, I'm engaged with a project related to High Frequency Trading.
- **Academic Excellence:** Holder of a Master of Science in Data Science from the University of Kalyani, combining practical experience with theoretical foundation.

## JOB EXPERIENCE

**Junior Data Scientist**  
Cloudcraftz Solutions Pvt Ltd

**July, 2022 – Present**  
Kolkata, India

### Activities/Responsibilities :

- **Analysing and Developing Strategies for High Frequency Trading:**
  - Developed and deployed a comprehensive framework facilitating the backtesting of high-frequency trading strategies utilizing Options for a proprietary trading firm.
  - Currently evaluating and analyzing the performance of diverse trading strategies to assist in optimizing strategies to enhance overall trading performance.
- **Detecting Suspicious Financial Activities in Transaction Networks:**
  - Developed transaction network visualization solutions for enhanced insights into complex financial data.
  - Identified rogue nodes through comprehensive analyses, considering transaction volume, centrality metrics, and transaction amounts.
- **Financial Cashflow Forecasting:**
  - Collaborated on a sophisticated cashflow forecasting project with an NBFC client.
  - Utilized PgAdmin for preprocessing raw financial data from the client's database.
  - Translated intricate insights into intuitive visualizations.
  - Conducted extensive feature engineering to boost predictive accuracy.
  - Created a robust machine learning model with a 2.9% RMSE reduction.
  - Implemented a stacked model approach for significant forecasting precision improvement.
- **Explainable AI Product Development:**
  - Devised methods using Shapley values for interpreting machine learning models.
  - Developed global and local model explanation techniques for quantifying feature contributions and their impact.
  - Integrated counterfactual-based explanations for enhanced interpretability and user-driven exploration.
  - Collaborated with cross-functional teams to ensure actionable insights for non-technical stakeholders.
- **Sentiment Analysis for Stock Price Prediction:**
  - Collaborated with an international client to build a comprehensive solution.
  - Developed a web scraping and NLP pipeline to efficiently gather and extract textual data from diverse news websites.
  - Leveraged Hugging Face Transformers for fine-tuning a sentiment analysis model, resulting in high-precision sentiment classification and enhancing the accuracy of stock price predictions.
- **Exploratory Data Analysis Platform Development:**
  - Designed and deployed an in-house EDA platform for tabular and time series data.
  - Implemented user-friendly visualization tools for non-technical users.
  - Contributed to the platform's statistical analysis capabilities.
  - Enhanced utility for comprehensive data exploration within the organization.
- **Automating Database Labeling:**
  - Provided NLP-based dataset labeling services for an international client.
  - Streamlined the database labeling process.

## INTERNSHIP EXPERIENCE

---

### Research Intern

USAID Project under LISA 2020

In association with Department of Statistics, University of Calcutta and  
National Institute of Wind Energy, Government of India

**Project Title :** Renewable Energy Modelling

March, 2022 — July, 2022

Kolkata, India

#### Activities/Responsibilities :

- Conducted thorough data exploration with Data Visualization and Exploratory Data Analysis.
- Implemented advanced Regression-based-Time-Series Models for GHI predictive forecasting.
- Achieved an impressive R-squared score of 0.92 in the forecasting model.

### Research Intern

A. K. Choudhury School of IT, University of Calcutta

**Project Title :** Environmental Sound Classification

September, 2021 — July, 2022

Kolkata, India

#### Activities/Responsibilities :

- Hands-on project in Environmental Sound Classification using the ESC-50 dataset.
- Applied audio processing techniques for spectrogram extraction.
- Implemented Convolutional Neural Network (CNN) models for sound classification.
- Demonstrated consistent proficiency with an impressive average accuracy score of 87%.

## PERSONAL PROJECT

---

### Project Title: Crafting Efficient Large Language Models with Precision and Flavor

- Designed and implemented a robust training pipeline for Large Language Models (LLMs) using key Python libraries and optimization techniques.
- Implemented quantization and LORA techniques for model size reduction and optimization, achieving resource-conscious brilliance.
- Orchestrated a comprehensive training pipeline achieving a sophisticated language model capable of complex task while respecting computational constraints.

**Tools Used :** HuggingFace, Transformers, Peft, PyTorch, Weights and Biases

**Project Repo :** [lora-peft-powered-llm-adaptation-with-accelerate-wandb](#)

## SKILLS AND INTERESTS

---

<b>Programming Languages</b>	Python, R
<b>Data Science &amp; AI/ML</b>	Exploratory Data Analysis (EDA), Data Visualization, Statistical Modelling, Predictive Analysis, Time Series, Deep Learning, Natural Language Processing (NLP) , Computer Vision
<b>Python Libraries</b>	NumPy, Pandas, Plotly, Altair, Scikit-Learn, TensorFlow, PyTorch, Hugging Face
<b>MLOps Tool</b>	Weights & Biases
<b>Database</b>	MySQL
<b>Operating Systems</b>	Linux, Windows
<b>Microsoft Tools</b>	MS-Excel, MS-PowerPoint, MS-Word
<b>Soft Skills</b>	Time management, Communication, Adaptability, Problem-solving, Teamwork
<b>Communication</b>	Fluent in English and Hindi, Native Bengali Speaker

## EDUCATION

---

### Master of Science in Data Science

University of Kalyani

- Overall CGPA : 9.41

November, 2020 — July, 2022

Kalyani, India

### Bachelor of Science (Honours) in Statistics

University of Calcutta

- Overall Percentage : 70.12%

August, 2017 — October, 2020

Kolkata, India