(+91) 8902690755
Kolkata, India
adityadawn98@gmail.com

Aditya Dawn

Data Scientist | AI/ML Enthusiast

GitHub: Adi-ds
LinkedIn: adityadawn-ds
Portfolio: adi-ds.github.io

I'm deeply passionate about Data Science, with a solid foundation in Python and practical industry experience. My expertise lies in exploratory data analysis, data visualization, data pre-processing, machine learning and NLP. 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 machine learning, NLP (Large Language Model-LLM) and time series analysis using Python with a solid two-year track record of hands-on application.
- **Industrial Expertise**: Over 1 year and 8 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 forecasting of commodity prices, using historical price data and historical news data with ML Classification Algorithms and LLM.
- **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

July, 2022 – Present

Cloudcraftz Solutions Pvt Ltd

Kolkata, India

Activities/Responsibilities :

- **Commodity Price Change Prediction**:
  - Successfully crafted a multi-classifier pipeline predicting price fluctuations (in percentage) of commodities within a specific time-frame.
  - Proactively wrote and tested prompts on LLM models (GPT, LLaMa, Gemini) using LangChain to extract news-driven sentiment and relevance scores towards commodity prices.
  - Pioneering a collaboration on integrating textual data with the multi-classifier pipeline for a novel business application.
- **High Frequency Trading Strategy Analysis**:
  - Meticulously evaluated and analyzed diverse trading strategies to assist in optimization for enhanced performance through data visualization and statistical tests.
- **Financial Cash-flow Forecasting**:
  - Collaboratively orchestrated a sophisticated cash-flow forecasting project for an NBFC client.
  - Systematically utilized PgAdmin to cleanse financial data and transformed complex insights into intuitive visualizations. Also, extensively conducted feature engineering to boost predictive accuracy.
  - Ingeniously implemented feature engineering and stacked model approach, achieving a 2.9% RMSE reduction and significant forecasting precision boost.
- **Explainable AI Product Development**:
  - Laboriously developed global and local model explanation techniques for quantifying feature contributions and their impact using Shapley values and other approaches.
  - Masterfully integrated counterfactual-based explanations for enhanced interpretability and user-driven exploration.
  - Actively 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 classification model with approximately 92% accuracy 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.

Internship Experience

Research Intern **March, 2022 — July, 2022**
USAID Project under LISA 2020 *Kolkata, India*
In association with Department of Statistics, University of Calcutta and
National Institute of Wind Energy, Government of India
Project Title : Renewable Energy Modelling

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 **September, 2021 — July, 2022**
A. K. Choudhury School of IT, University of Calcutta *Kolkata, India*
Project Title : Environmental Sound Classification

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

Programming Languages	Python
Data Science & AI/ML	Machine Learning (Regression-&-Classificaton), Deep Learning, Natural Language Processing (NLP), Large Language Model (LLM), Generative AI, Exploratory Data Analysis (EDA), Data Visualization, Statistical Modelling, Predictive Analysis, Time Series (Analysis-&-Forecasting)
Cloud Platforms	GCP, AWS
Python Libraries	NumPy, Pandas, Plotly, Scikit-Learn, TensorFlow, PyTorch, HuggingFace, LangChain
MLOps Tool	Weights & Biases
Operating Systems	Linux, Windows
Microsoft Tools	MS-Excel, MS-PowerPoint, MS-Word
Soft Skills	Time management, Communication, Adaptability, Problem-solving, Teamwork
Communication	Fluent in English and Hindi, Native Bengali Speaker

Education

Master of Science in Data Science **November, 2020 — July, 2022**
University of Kalyani *Kalyani, India*
- Overall CGPA : 9.41

Bachelor of Science (Honours) in Statistics **August, 2017 — October, 2020**
University of Calcutta *Kolkata, India*
- Overall Percentage : 70.12%