Ujjwal Chowdhury

Data Scientist, Research Analyst

+91 8768686849 u77w41@gmail.com Kolkata, India
in/u77w41 github.com/u77w41 𝕏 UjjwalCh0wdhury

Profile

Data Scientist with a Masterś in Data Science and expertise in NLP, Generative AI, and Stock Forecasting. Skilled in AI solution design, ML model implementation, and cross-functional collaboration to deliver impactful insights. Proficient in process automation, outcome analysis, and effective communication for driving business success.

Areas of Expertise

NLP, Generative AI, Machine Learning Algorithms, Time Series Analysis, Algorithm Optimization, Statistical Analysis, Data Visualization, Model Development, Cross-functional Collaboration

Skills

- **Data Visualization:** Microsoft Power BI, Excel, Tableau, Seaborn, Plotly, Matplotlib
- **Machine Learning and Deep Learning:** Feature Engineering, Model Development, Hyper-parameter Tuning, Neural Networks, Reinforcement Learning, Transfer Learning, Optimization Techniques, MLOps
- **Tools/Frameworks:** Python, R, TensorFlow, PyTorch, Keras, TFLite, MLFlow, PySpark, PostgreSQL, Azure, AWS, LangChain, Streamlit, Docker, Pydantic
- **Natural Language Processing:** Text Generation, Sentiment Analysis, Speech Recognition, Named Entity Recognition, Text Classification, LLM Prompt Engineering
- **Computer Vision:** Image Processing, Object Detection, Image Classification, Image Segmentation, Image Generation
- **Data Analysis and Mining:** Data Mining, Web Scrapping, Statistical Analysis, Time Series Analysis, Anomaly Detection, Predictive Analytics, Survival Analysis
- **Soft Skills:** Problem-Solving, Teamwork, Active Listening, Adaptability, Communication, Analytical Thinking

Professional Experience

Research Executive (AI & NLP) *(Feedsense AI (Formerly Vista Intelligence))* **Kolkata, India** *(Jan 2023 - Present)*

- Led the NLP team, overseeing project developments and team operations.
- Finetuned an RNN-Tranducer driven speech-to-text model to effectively capture Indian accents, decreasing the Word Error Rate from 56.8% to 23.4%.
- Employed a 4-bit quantized Mistral 7b LLM model for summarizing conference call conversations.
- Utilized reinforcement learning models integrating financial data to predict market movements and develop optimized trading strategies.
- Utilized OpenAI API with Langchain to develop a large document summarizer.
- Developed a live audio transcription model for real-time news analysis.
- Developed a trade signal generator model integrating live audio, textual news articles, OHLC data, and quantitative techniques. Achieved over 75% directional accuracy in generating Nifty F&O trading signals, enabling informed trading decisions.
- Created an auto-question generator program to generate questions based on applicant CVs, aiding the hiring team.

Research & Publication

- Investigate How Market Behaves: Toward an Explanatory Multitasking Based Analytical Model for Financial Investments *(IEEE Access, March 2024)* DOI: 10.1109/ACCESS.2024.3369033

Courses & Certifications

- Artificial Intelligence (AI) for Investments *(April 2023)* - NPTEL
- Cloud Computing and Distributed Systems *(March 2023)* - NPTEL
- NISM-Series-XV: Research Analyst *(Feb. 2023)* - National Institute of Securities Markets
- Data Base Management System *(Oct. 2022)* - NPTEL
- Deep Learning for Computer Vision *(Oct. 2022)* - NPTEL
- Data Science Math Skills *(April 2020)* - Duke University, Coursera

Education

MSc Data Science *RKMVERI* **Belur, West Bengal, India** *2021-2023*

Relevant Courses: Probability & Stochastic Process, Data Structures & Algorithms , Statistics, Machine Learning, Deep Learning, Computer Vision, NLP, Optimization Techniques, Linear Algebra, Time Series Analysis, Survival Analysis, Cloud Computing, Multivariate Statistical Analysis, Data Mining, DBMS

BSc Mathematics *Vidyasagar University* **Medinipur, West Bengal, India** *2017-2020*

Relevant Courses: Set Theory, Calculus, Geometry & Differential Equation, Higher Algebra, Real Analysis, Differential Equations & Vector calculus, Group and Ring Theory, Theory of Equation, Graph Theory, PDE, ODE, DBMS, Operation Research, Numerical Methods

Personal Projects

- **Fin-Bot: Advanced Agent based Financial Chatbot**
  (Domain: NLP, LLM, Generative AI, Deep Learning, RAG)
  - Seamlessly integrated web search functionality ensuring comprehensive responses to queries.
  - Implemented a custom vector database for efficient retrieval of financial news articles and concall transcripts.
  - Employed LLM-equipped agent to direct user queries to relevant web or custom database, ensuring up-to-dated, comprehensive insights.

- **SALES FORECASTING AND ANOMALY DETECTION ON WALMART SALES DATASET**
  (Domain: Machine Learning, Time Series Analysis, Deep Learning)
  - Used Factor Analysis for feature extraction.
  - Concepts of time series, machine learning, and deep learning are used to predict future sales.
  - Used unsupervised techniques to detect the anomalies.

- **Deep Bidirectional LSTM Network for Textual Sentiment Analysis**
  (Domain: Deep Learning, Sentiment Analysis, NLP, Web Scrapping)
  - Integrated Twitter API for real-time tweet scraping.
  - Utilized AsyncHTMLSession to scrape news articles from Google News.
  - Leveraged Bi-LSTM architecture to process sequential data and extract meaningful features for sentiment classification.

- **BRAIN TUMOUR CLASSIFICATION**
  (Domain: Computer Vision, Deep Learning, Optimization Techniques)
  - Used Transfer Learning and Fine-tuned several pre-trained models.
  - Explored different optimization algorithms such as Adam, RMSProp, SGD, GD, Adagrad etc.
  - Applied snapshot learning technique to construct an ensemble predictive model.

- **STATISTICAL ANALYSIS OF DIET, EXERCISE AND FITNESS**
  (Domain: EDA, Data Visualization, Data Analysis, Statistical Inference)
  - Collected data using online surveys at different time frames.
  - Employed descriptive statistics to summarize the key characteristics of the dataset.
  - Used Power BI, Tableau, Excel, R, and Python for analysis and visualization.