Chandan

Kolkata, West Bengal
7209062108 chandan875792@gmail.com
LinkedIn GitHub

Education

Indian Institute of Technology, Kharagpur Kharagpur, India
Dual Degree (Integrated Bachelor & Master of Technology) *2020 – 2025*
- GPA: 8.07

Kendriya Vidyalaya No.1
Higher Secondary School Certificate *2017 – 2019*
- GPA: 87.6

D.A.V. Public School
Secondary School Certificate *2007 – 2017*
- GPA: 9.2

Experience

AI Developer Kolkata, India
Kalolwala & Associates *May 2025 – Present*
- Developing AI-powered financial analysis solutions using large language models and advanced machine learning techniques
- Leading the design and implementation of specialized LLMs for financial document analysis and investment insights
- Building scalable AI pipelines for real-time financial data processing and sentiment analysis

Data Science Intern
ByteLearn *Apr 2024 – Jun 2024*
- Collaborated with the HR team to develop an automated visualization dashboard to track and infer from employee performance
- Developed a dashboard using various python libraries including Streamlit, Matplotlib and integrated project management tools via APIs
- Enhanced efficiency by 17% and job satisfaction by 21% through formulation of strategies after the analysis of data on employee KPIs
- Streamlined financial data visualization, facilitating clear comprehension of how employee performance relates to financial performance

Data Analyst Intern
Toyadhi, IIT Kharagpur *Dec 2023 – Feb 2024*
- Analyzed hydroponics startups in India and built a predictive model for optimizing crop yields using machine learning
- Researched and collected data on hydroponics startups in India to identify industry trends and effective practices for optimizing crop yields
- Trained XGBoost model on data to predict the yield of 6 hydroponically grown crops including lettuce, cucumber and strawberry
- Reduced MSE by 32% through performing EDA and feature selection, extracting 9 key features using PCA and correlation analysis

Projects

Financial InsightGPT: Specialized Language Model for Earnings Analysis and Investment Insights
Jan 2025 – Present

- Fine-tuned Large Language Model (LLM) for Finance: Specialized in training a state-of-the-art LLM (e.g., Mistral-7B) with multi-source financial datasets including earnings reports, market trends, and financial Q&A, leveraging LoRA and instruction tuning techniques for finance-specific natural language understanding
- Advanced Financial Document Analysis: Developed an AI-powered assistant capable of interpreting SEC filings, extracting key financial metrics, analyzing sentiment from Financial PhraseBank, and performing complex numerical reasoning using datasets like FinanceQA and FinTextQA
- Full AI Pipeline: Data Processing to Deployment: Designed a robust ML pipeline involving data acquisition, cleaning, chunking, instruction-formatting, and model fine-tuning, followed by scalable deployment with API endpoints (FastAPI), high-performance serving (vLLM), and interactive consumer interfaces (Streamlit, Gradio)
- Model Evaluation & Insights Generation: Delivered a high-accuracy financial analysis model, benchmarked with BLEU, Rouge-L, and F1 metrics, providing real-time investment insights, earnings summaries, sentiment classification, and risk assessment for analysts, investors, and finance professionals

Entrepreneurial Insight Engine: Fine-tuning an LLM for Startup Knowledge
Jan 2025 – Mar 2025

- Developed an "Entrepreneurial Insight Engine," a Generative AI model fine-tuned to provide practical, mentor-like advice on startups, product development, fundraising, and monetization
- Curated and preprocessed a specialized dataset of 1000+ instruction-output pairs from expert sources like Y Combinator talks and Paul Graham essays, ensuring high-quality and domain-specific knowledge
- Fine-tuned a DistilGPT-2 (or GPT-2) language model on Google Colab using Hugging Face Transformers, PyTorch, and the datasets library, implementing a robust training and validation pipeline
- Engineered prompts and experimented with generation parameters (temperature, top-k, top-p) to elicit coherent, contextually relevant, and persona-aligned responses from the fine-tuned model
- Conducted qualitative evaluations of the model's outputs, demonstrating its improved ability to answer startup-related queries compared to the base pre-trained model

Twitter Sensitivity Analysis
Sep 2023 – Oct 2023

- Implemented NLTK tokenization, Porter stemming, and TF-IDF extraction to process Twitter data, cleaning text by removing hashtags and URLs
- Applied Naive Bayes and Logistic Regression classifiers, achieving 78% accuracy with Naive Bayes and 82% with Logistic Regression
- Incorporated BERT embeddings and fine-tuned the model for sentiment analysis, attaining 92% accuracy and 93% with RoBERTa

Technical Skills

Programming Languages: Python, SQL, C++

Libraries: Numpy, Pandas, Matplotlib, Scikit-learn, Seaborn, TensorFlow, BeautifulSoup4, Selenium, Bootstrap, OpenCV, Keras, Hugging Face Transformers, FastAPI, Streamlit, Gradio, vLLM

Statistics and Mathematics: Descriptive Statistics, Inferential Statistics, Data Cleaning, Data Transformation, Feature Engineering, Advanced Calculus, Linear Algebra, Numerical and Complex Analysis

Modeling: Logistic Regression, Decision Tree, Random Forests, PCA, CNN, RNN, Transformers, NLP, LLM, LoRA Fine-tuning, Instruction Tuning, Financial Analysis

Online Coursework: Machine Learning Specialization by Andrew Ng, Specialization in Deep Learning by Andrew Ng