Chandan

Data Science Intern

chandan875792@gmail.com | https://www.linkedin.com/in/chandan-kumar-3a0396200/
M.TECH | B.Tech | M.Tech
7209062108

Education

Indian Institute of Technology, Kharagpur
Dec 2020 - Apr 2025

Dual Degree (Integrated Bachelor & Master of Technology)

GPA: 8.07

Kendriya Vidyalaya No.1
Apr 2017 - Mar 2019

Higher Secondary school Certificate

GPA: 87.6

D.A.V. Public School
Apr 2007 - Mar 2017

Secondary school Certificate

GPA: 9.2

Projects

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.

Multi-Class Image Classification | Computer Vision
Jan 2024 - Apr 2024

- Developed a robust CNN model for accurate 4-class image classification, employing techniques like dropout and softmax in the final layer.
- Designed the model with an input image size of 224×224, using categorical cross-entropy loss and optimized with Adam optimizer.
- Achieved 90% accuracy after 100-epoch training (batch size: 32); saved the model in h5 format.

Twitter Sensitivity Analysis | DeepLearning.ai | Self Project
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.

Work Experience

Data Science Intern
Apr 2024 - Jun 2024

ByteLearn

- Collaborate 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
Dec 2023 - Feb 2024

Toyadhi, IIT Kharagpur

- to analyze hydroponics startups in India and build 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 our 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

Core Skills

Programing language : Python, SQL, C++

Libraries: Numpy, Pandas, Matplotlib, Scikit-learn, Seaborn, TensorFlow, BeautifulSoup4, Selenium, Bootstrap, OpenCV, Keras

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

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