Nilanjan Ghosh

Motivated computer science student specializing in applied machine learning, Python, and Linux, with a focus on crafting data-driven solutions.

India | +917200565892 | csz248005@cse.iitd.ac.in | https://nilanjanghosh-digitalcv.streamlit.app/

Profiles

Nilanjan Ghosh

nil0711

csnil0711

Summary

Aspiring computer scientist with strong skills in Python, Linux, and applied machine learning. Demonstrated ability to create data-driven solutions using statistical modeling and algorithmic optimization techniques. Passionate about innovation and learning new technologies. Seeking a challenging role in a forward-thinking organization where I can contribute my expertise, collaborate on impactful projects, and push the technological boundaries.

Education

Indian Institute of Technology, Delhi — **July 2024- Present**
Computer Science — PhD
https://home.iitd.ac.in/

I am pursuing a PhD in Computer Science at IIT Delhi, focusing on Quantum Machine Learning. My research explores advanced techniques in quantum algorithms and their application to machine learning, aiming to enhance computational efficiency and solve complex problems in data analysis.

Pondicherry University — **December 2022 - June 2024**
Computer Science — M.C.A
9.36
https://www.pondiuni.edu.in/

Completed MCA (Master of Computer Applications) from Pondicherry University with a specialization in ML, AI, Linux, and Python Programming and secured a CGPA of 9.36 Gained hands-on experience in programming, database management, software engineering, and machine learning.

Projects

Evaluating Large Language Model Transfer — **February 2024 - May 2024**
This project demonstrates that transfer learning from Sanskrit to Hindi improves Hindi language model performance over direct Hindi training.
https://github.com/nil0711/LM_Reseach_Project

The project explores the efficacy of transfer learning by training a Sanskrit language model on Hindi data to create a model that performs better on Hindi tasks than a model trained directly on Hindi. This endeavor aims to illustrate that leveraging a pre-trained Sanskrit language model can enhance the performance of Hindi language tasks through transfer learning.

Initial Training Phase: The project begins by independently training two language models using a Transformer-based architecture. The first model is trained on a Sanskrit corpus, and the second model is trained on a Hindi corpus. Each model utilizes a vocabulary derived from their respective training datasets. The architecture comprises essential components such as masked multi-head self-attention mechanisms and feed-forward neural networks within transformer blocks.

Transfer Learning Phase: The core phase involves applying transfer learning to the pre-trained Sanskrit model. The Sanskrit model is further trained on a Hindi dataset. During this phase, the vocabulary evolves from predominantly Sanskrit to a balanced mix of Sanskrit and Hindi, and eventually to a vocabulary saturated with Hindi. This dynamic adjustment allows the model to progressively learn Hindi while retaining its foundational knowledge from Sanskrit.

Vocabulary Management: A critical aspect of this phase is managing the vocabulary. Initially, the Sanskrit vocabulary dominates, but as training progresses, Hindi characters increasingly populate the vocabulary. The vocabulary update strategy includes mechanisms for adding new Hindi characters and potentially forgetting less relevant Sanskrit characters based on their usage during training. This method ensures a seamless transition and prevents abrupt disruptions in the learning process.

Evaluation and Metrics: The effectiveness of the transfer learning approach is evaluated using several metrics: loss, accuracy, precision, recall, F1 score, and BLEU score. These metrics provide a comprehensive view of the model's performance. The transfer-learned model's performance is then compared to the model trained exclusively on Hindi data. The results demonstrate that the transfer-learned model exhibits superior performance, validating the hypothesis that pre-training on a related language (Sanskrit) can enhance learning and performance in a target language (Hindi).

This project underscores the potential of transfer learning in natural language processing, particularly in leveraging linguistic similarities between languages to improve model performance. The findings suggest that starting with a robustly trained model on a related language and fine-tuning it on the target language data can yield significant improvements in performance, showcasing the power and flexibility of transfer learning techniques in multilingual contexts.

Chat Analysis with Sentiment Analyzer — **Sept 2023- Jan 2024**
A Streamlit application that analyzes WhatsApp chat data using natural language processing techniques and provides interactive data visualizations and a chatbot interface.
https://miniproject-senti.streamlit.app/

This project is a Streamlit application that analyzes WhatsApp chat data using natural language processing techniques and provides interactive data visualizations and a chatbot interface. The main features of the application are:

Data preprocessing: The application parses timestamps, extracts messages, and organizes data into a structured format using Python libraries such as pandas, numpy, and NLTK.

Sentiment analysis and emotion detection: The application uses TextBlob and NLTK to perform sentiment analysis and emotion detection on the chat messages. It also categorizes sentiments into negative, neutral, mixed, and positive, and emotions into joy, anger, neutrality, and sadness.

Data visualization: The application offers a dynamic dashboard, enabling users to explore data visualizations, such as timelines, activity maps, and word clouds. Users can selectively analyze individual or group chat behavior, including message frequency, media sharing, and link sharing.

Sentiment and emotion trend analysis: The application facilitates sentiment and emotion trend analysis over both monthly and daily intervals. It shows how the chat mood changes over time and identifies the most positive and negative days.

The application provides valuable insights into user behavior, sentiment dynamics, and emotional nuances within WhatsApp conversations, making it a powerful tool for users seeking a comprehensive understanding of their chat data.

File manager — **March 2023- May 2023**
A Python file manager GUI using Tkinter and ttkbootstrap, with features such as file manipulation, searching, zipping, opening, properties, and permissions.
https://github.com/nil0711/CODE/blob/main/tkinter_test/test5.py

This project is a Python file manager GUI that allows users to perform various file operations such as copying, moving, renaming, deleting, searching, zipping, unzipping, opening, viewing properties, and changing permissions. The project uses the Tkinter and ttkbootstrap libraries to create a user-friendly and responsive interface. The project demonstrates the use of object-oriented programming, file handling, subprocesses, and error handling in Python.

Skills

Programming
Python
●●●●○
Numpy, Pandas, sklearn, Pytorch, Keras, Streamlit, Seaborn, Mathplotlib, google.generativeai, openai, tensorflow, nltk, Faiss, langchain, Plotly

Machine Learning
Artificial Intelligence
●●●●○
Decision Tree, Neural Networks, Bayesian Network, Markov Model, Hidden Markov Model, Clustering, Classification, Deep Learning, LLM, NLP

Linux
Programming
●●●●●
Linux System Administration, Shell Scripting, Samba, LVM, Git, C/C++/Java/Python coding debugging in terminal

Database
Storing Data
●●●●○
MongoDB, MySQL, ChromaDB

Interests

Coding for fun **Puzzle Solving** **Table Tennis** **Cricket**

Certifications

Computing with Python — **DEC 2023**
NPTEL
https://nptel.ac.in/

Completed Computing with Python course issued by NPTEL under IIT Ropar in December 2023.

Achieved a top score of 88% among 30,000+ learners.

Acquired skills in data analytics, probability, hypothesis testing, regression, clustering, and classification using Python.

Data Analytics with Python — **April 2023**
NPTEL
https://nptel.ac.in/

Completed Data Analytics with Python course issued by NPTEL under IIT Roorkee in April 2023.

Learned how to use Python for data analysis, probability, hypothesis testing, regression, clustering, and classification.

Applied data analytics skills to real-world problems and datasets using Python tools such as Pandas, NumPy, and IPython.

Awards

JRF — **Janurary 2024**
University Grants Commission(UGC)
https://ugcnet.nta.nic.in/

Awarded Junior Research Fellowship (JRF) by University Grants Commission (UGC) National Eligibility Test (NET) in January 2024.

Demonstrated excellence in academic knowledge and research skills in the subject of Computer Science and Applications.

GATE — **February 2024**
IISC Bangalore
https://gate2024.iisc.ac.in/

Qualified GATE 2024 in CS and DA, two competitive and sought-after papers.

Proficient in core topics of data science, machine learning, AI, and general aptitude.

Languages

English
Proficient
●●●●●

Bengali
Mother Tongue
●●●●●

Hindi
Fluent
●●●●○