RAJ GARG

rajgarg021.github.io/blog
rajgarg021@gmail.com
(+91) 75670 25289

SKILLS

Languages: Python, R, SQL, C, LISP, Javascript, Mojo
Frameworks: PyTorch, LangChain, ONNX, Jax, Keras

EXPERIENCE

Personal Sabbatical **Dec'22 - Present**

- Took a break from work to focus on improving my mental health, looking forward to start working again.
- Spent time reading research papers, learning the fundamentals and staying up-to-date with the latest advancements in the AI field.
- Implemented some research papers, experimented with open-source language models and created some cool projects mentioned below.

Monsoon CreditTech (Senior Data Scientist, Gurgaon) **Jul'22 - Nov'22**

- Took full ownership of the project from the start, built and deployed Application Scorecard (Credit underwriting risk model to assess likelihood of default) for loan applications at a leading NBFC.
- Quickly acquired the domain knowledge and improved upon the existing bureau data processing and feature creation scripts which led to an increase in the model performance.
- Automated the feature selection pipeline using Boruta, RFE and Null point importance and the Hyper-parameter tuning pipeline using Optuna.
- Managed and mentored junior data scientists in the team.

MateLabs (Data Scientist, Bangalore) **Jul'20 - Jun'22**

- Built an end-to-end AutoML based SaaS platform, developed the data preprocessing and training modules, which is now being used across industry in various domains.
- Worked on optimising supply chain for some of the biggest FMCG companies using machine learning. Created models for Demand planning and helped them understand the bottlenecks by model interpretability as a result saving up on their revenue.
- Developed a META learning algorithm that selects right model and hyper-parameters for a dataset, which was in future used in the AutoML platform.
- Mentored newly joined data scientists.

MateLabs (Intern, Bangalore) **Mar'20 - Jun'20**

- Research and deployment of forecasting techniques of Statsmodels, ARIMA, SARIMAX, Holt-winter, FB-Prophet.
- Built Entity Embeddings to extract unsupervised features for model training and helped in understanding of Product Similarities.

SELECTED PROJECTS

Llama 3 from scratch

`https://github.com/rajgarg021/llama3-from-scratch`

- Clean implementation of **Llama 3** inference code from scratch with lots of comments explaining every step and every matrix shape change.
- Read research papers for all the building blocks - **Grouped-Query Attention with KV cache, Rotary Position Embeddings, RMSNorm, SwiGLU activation function** to gain a deeper understanding before implementing.

GPT from scratch

`https://github.com/rajgarg021/gpt-from-scratch`

- Implemented a **Decoder-Transformer network (10M parameters)** from scratch.
- Helped gain intuitive understanding of **self-attention**.
- Used optimization techniques like **residual connections** and **layer normaliztion**.

Experimented with various large language models

`https://github.com/rajgarg021/LLM-experiments`

- Experimented with various **open-source models** like **LLama2, Mistral, Mixtral 8x7B, Gemma**, etc to build a **RAG** application for generating answers given a textual context.
- Played around with **LangChain, OpenAI API (GPT-3.5-Turbo), Whisper** and **Pinecone** (to store embeddings) to transcribe a video and generate answers using that as context.

Implemented various quantization techniques for neural networks

`https://github.com/rajgarg021/quantization`

- Implemented **affine and scale quantization** from scratch.
- Trained a simple neural network and implemented **post-training quantization** using PyTorch which resulted in $\sim$75% reduction of the model size.
- Improved the accuracy of the above model by using **quantization aware training**.

Teenytoken

`https://github.com/rajgarg021/teenytoken`

- Implemented **byte-level byte pair encoding** algorithm from scratch to create my own **tokenizer**.
- Working on improving it further to make it equivalent to the **GPT-4 tokenizer**.

Fine tuned BERT for sentiment analysis

`https://github.com/rajgarg021/Fine-tuning-BERT-for-sentiment-analysis`

- Used **HuggingFace's transformers** library to fine tune a **BERT** model for sentiment analysis task.

ACHIEVEMENTS

- Scored **98.89%** in CAT 2018 and converted several **IIM** interviews to admission offers.
- Received **scholarship** from **Bill & Melinda Gates Foundation** during college.