Sushant Pargaonkar

Data Scientist | Machine Learning Engineer

+918149927271 sushant.pargaonkar97@gmail.com Linkedin Github Portfolio

Education

IIT Guwahati July 2024
M. Tech. in Artificial Intelligence and Robotics (CGPA: 8.42 / 10.00) *Guwahati, Assam*
- **Relevant Coursework:** NLP, Information Retrieval, Machine Learning, Computer Vision, Reinforcement Learning.

College of Engineering Pune May 2019
Bachelor of Technology in Civil Engineering (CGPA: 6.74 / 10.00) *Pune, Maharashtra*
- **Relevant Coursework:** Linear Algebra, Calculus, Numerical Methods for optimization.

Experience

TVS Motors May 2023 – July 2023
Data Science Intern *Bengaluru, Karnataka*
- Implemented **RAG**-based **Natural Language to SQL** generation model using **Langchain** & **Qdrant** vector database with security for prompt and SQL injection.
- Developed **Speaker Diarization** pipeline to convert customer feedback from **speech to dialogue** using **WhisperX, SpeechBrain model** for Indian languages and applied aspect based sentiment analysis.

Air Products August 2019 – October 2021
Associate Structural Engineer *Pune, Maharashtra*
- Lead structural engineering of green H2 Project worth **10000** man hours and collaborated with multiple stakeholders to streamline processes, cutting project timelines by **5x200 hours**.
- Produced **100+** structural design drawings and **20+** design specifications by incorporating Staad-Pro software.

Projects

Master's Thesis — Cold-Start Data Selection for Language Model Fine-tuning | *Python, Pytorch, spacy, C++* Github
- Developed **active learning** algorithm using **Isolation Forest** for initial data selection to fine-tune PLMs such as BERT.
- Evaluated performance over NLP task such as Biomedical NER, Extractive QA & Text Classification.
- Achived average 50% reduction in sampling time compared to state-of-the-art (PATRON) with 2-5% decrease in accuracy.

Optimized GPT-2 with Reinforcement Learning from Human Feedback | *Python, RLHF, PPO, LLM* Github
- Implemented supervised fine-tuning (SFT) and Proximal Policy Optimization (PPO) with LoRA to align GPT-2 model output with human feedback.
- Achieved up to 67% accuracy. SFT model preferred 91% over vanilla model while PPO model preferred 88.9% over the SFT model by ChatGPT.

Accelerating BERT Inference with LoRA and TensorRT for Stack Overflow Question Classification Github
- Optimized PyTorch BERT model for question closure prediction on StackOverflow using ONNX and TensorRT for efficient inference.
- Implemented and compared standard fine-tuning with LoRA (Low-Rank Approximation) fine-tuning, developing custom implementations from scratch.
- Achieved a 4x inference speed-up using NVIDIA TensorRT with INT8 precision, reducing accuracy by only 0.4% compared to PyTorch FP32 on the BERT-LoRA model.

Enhanced Fraud Detection Using Graph Neural Networks | *XgBoost, GNN, RGCN, DGL* Github
- Established a baseline for fraud detection model with exploratory data analysis(EDA) and feature engineering using XgBoost.
- Built a Graph Neural Network-based detection system using Relational-GCN and GTAN.

Technical Skills

Languages: Python, Core Java, SQL
Libraries & Tools: Pytorch, Tensorflow, DGL, Numpy, Pandas, Scikit-learn, OpenCV, Git, Docker
Concepts: Machine Learning, Information Retrieval, Natural Language Processing, Deep Learning, Statistics, Computer Vision, LLM, Neural Networks, Image Processing, Reinforcement Learning, Artificial Intelligence, Data Structures and Algorithms