GAURAV SOOD

Sr. Data Scientist

@ gsood.gaurav@gmail.com   📞 +91 9632714987   📍 Bangalore, India   ⚙ github.com/gsood-gaurav

in linkedin.com/in/gsood-gaurav/

"Senior Data Scientist with total 14 years of experience, and 7 years of experience in Machine Learning, developing open source software in Python, C++, contributing as both individual contributor and team leader, working on Machine Learning, conceptualizing, designing and implementing end to end solutions. Currently working on ideas around Natural Language Processing and Reinforcement Learning to automate Hardware Diagnostics. Have used both traditional machine learning methods like Bayesian Networks, Conditional Random Fields and modern deep learning methods LSTMs, Transformer based models like BERT and GPT"

SKILLS

Python   PyTorch   TensorFlow   JAX   Julia   Flux   NumPy   SciPy   Matplotlib

Large Language Models   Generative AI   Probabilistic Modelling   Reinforcement Learning   Open Source Software

Communication   Leadership

EXPERIENCE

Sr. Data Scientist

HP Inc. June 2021 – Ongoing

🗓 Bangalore, India

- Designed and Developed Large Language Models Based application for web content filtering acheiving a recall of 0.94
- Implemented evaluation scripts to benchmark LLM application's performance on accuracy, cost(token usage) and inference time comparing GPT4/GPT3.5/GPT-4o-mini/GPT-4o
- Mentored three data science interns on Hardware logs analysis using Graph Neural Networks and Large Language Models(GPT-2) achieving precision/recall of 0.85/0.73.
- Fine Tuned 4-bit quantized Small language Chat model on custom data set using QLORA acheiving an accuracy of 0.96
- Developed Telemetery data analysis pipline using Large Language Model embeddings, sentence transformers, UMAP, Variational AutoEncoder and HDBScan

Prompt Engineering   Language Models   FineTuning   Probabilistic modelling   Leadership Skills

Software Engineer Machine Learning

HP Inc. June 2017 – 2021

🗓 Bangalore, India

- Analyzed Sales Time Series Data (Short term/Long term forecasting) using Prophet, ARIMA, Gradient Boosting models.
- Behavior Modelling using Tree Based Models, SHAP, UMAP
- Developed simulator environment for online Reinforcement Learning for automated hardware diagnostics.
- Developed Sequence Tagging Model using LSTMs for discovering insights in call center records acheiving a precision and recall score of 0.91/0.85
- Mentored 4 interns on sequence tagging using Conditional Random Fields to extract insights from call center records.
- Presented and communicated ideas around Probabilistic Modelling and Reinforecement Learning to senior leadership
- Bayesian Networks

Time Series Analysis   CART   Reiforcement Learning   Probabilistic Graphical Models   Data Annotation

Senior Software Engineer

HP Inc. June 2015 – 2017

📅 Bangalore, India

- Designed and Developed a parser for PostScript Language to extract images and meta data which was impacting 50000 hp customers and saving 5 million dollars.
- Designed and developed Telemetery module in python.
- Led team of 7 software developers designing and implementing open source software used by over 1 million customers.
- Worked on understanding PDF specification, design and implemented PDF document processing using PDF libraries like iText and PDFium
- Lead a team of six people which delived Imaging Processing Solutions using Python Imaging Library, TessaractOCT to over 10 PSU banks in india

Algorithms | OpenSource Sofware | Image Processing

Software Engineer

HP Inc. June 2010 – 2015

📅 Bangalore, India

- Designed and Implemented Telemetry module in python using asyncio.
- Led team of 7 software developers designing and implementing open source software used by over 1 million customers.
- Collaborated with customers and stakeholders, gathering requirements and communicating constraints.
- Developed UI, backend, modules in Python for the hp's open source imaging solution (https://developers.hp.com/hp-linux-imaging-and-printing)

Python | Open Source Software

CERTIFICATIONS

Deep Learning Specialization
Coursera Oct-2021 Credential ID G8PG2GY6WFS)

Reinforcement Learning Specialization
Coursera Sept-2021 Credential IDD3LDNU8BRW68

Introduction to Quantum Computing
Coursera May-2021 Credential ID RXQHLEC2WYC8

EDUCATION

Msc Research Speech Recoginition
Indian Institute of Science Bangalore 2006-2009

- 📖 Speech Recognition using Support Vector Machines

BTech Electronics and Communication Engg
GNE Ludhiana 2001-2005