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 7 years of experience in Data Science and total experience of 14 years in software industry, contributing as and individual contributor and team leader, building machine learning models from conception to completion to improve operational efficiency. Worked on Natural Language Processing, Time Series Forecasting, Tabular Datasets, Graph Neural Nets and Reinforcement Learning. Have experience in traditional machine learning methods (Probabilistic Graphical Models), deep learning models (LSTMs, CNNs) and Transformer based models (BERT, GPT and LLMs)"

SKILLS

Python   PyTorch   TensorFlow   JAX   Julia   Flux/Lux   NumPy   SciPy   Matplotlib

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

Communication   Leadership Skills

EXPERIENCE

Sr. Data Scientist

HP Inc. May 2021 – Ongoing

📅 Bangalore, India

- Designed and Developed Large Language Models(LLMs) Based application for web content filtering and topic recommendation (Currently in beta phase.)
- Implemented evaluation scripts to benchmark LLM based application's performance on accuracy, cost(token usage) and inference time reducing cost by 30%.
- Mentored 3 data science interns on automating triaging with machine learning on hardware logs using Graph Neural Networks and Generative Pretrained Transformer(GPT2)
- Led the Small Language Models(SLMs) exploration and finetuning initatives using PEFT techniques as an alternative to Large Language Models.
- Developed Telemetery data analysis pipline processing 0.5 milllion customer record using Large Language Model embeddings, BERT, dimensionality reduction(UMAP, Variational AutoEncoder) and clustering HDBScan enhancing operational efficiency by 20%.

Prompt Engineering   GPT2   TinyLlama   FineTuning   Embedding Models   Graph Nerual Networks

Sr. Software Engineer Machine Learning

HP Inc. July 2017 – May 2021

📅 Bangalore, India

- Analyzed Sales Time Series Data (Short term/Long term demand forecasting to improve supply chain processes) using statistical models, Tree based models and Generalized Additive Models.
- Analyzed the effects of macro economic indicators on Demand Forecasts to optimize inventory management.
- Implemented Hyperparameter Tuning module for Time Series models using Optuna reducing the model execution time by 25%.
- Worked on HP Personal Systems User Behavior Modelling using Tree Based Models, SHAP, UMAP, HDBScan for targeted market campaigning.
- Developed simulator environment for online Reinforcement Learning for automated hardware diagnostics.
- Developed Sequence Tagging Model using BERT, LSTMs for discovering insights in call center records which lead to 30% increase in operational efficency.
- Mentored 4 interns on sequence tagging using Conditional Random Fields to extract insights from call center records.

- Presented and communicated ideas about Probabilistic Modelling and Reinforecement Learning to senior leadership
- Analyzed Call center record using Bayesian Networks for predictive diagnostics as an aid to call center executives.

Time Series Analysis | CART | Reinforcement Learning | Probabilistic Graphical Models | Data Annotation

Senior Software Engineer
HP Inc. March 2015 – June 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 0.5 million dollars.
- Worked on understanding PDF specification, design and implemented PDF document processing using PDF libraries like iText and PDFium
- Lead a team of 4 software engineers which delivered imaging processing solutions using Python Imaging Library, Tessaract OCR to over 10 PSU banks in india

Algorithms | OpenSource Sofware | Image Processing | Leadership Skills

Software Engineer
HP Inc. May 2010 – Feb 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 | Asynchronous Programming | Leadership Skills

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

PREVENTIVE DISCLOSURES

- Gaurav Sood, Ranjith Kumar Gopa, Nandan Kumar Nidhi(May 2024) "Representing Tabular Data using Large Lanugage Models" https://www.tdcommons.org/dpubs_series/6998/
- Gaurav Sood (Feb 2021) "A Reinforcement Learning Approach to Printer Diagnostics" https://www.tdcommons.org/dpubs_series/4244/
- Gaurav Sood, Pavleen Kaur Brar, Raghu Anantharangachar (April 2019) "Sequence Prediction for Print/Scan Issues using Bayesian Networks" https://www.tdcommons.org/dpubs_series/2117/

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