GAURAV SOOD

Sr. Data Scientist

@ gsood.gaurav@gmail.com   ☎ +91 9632714987   📍 Bangalore, India   ⊙ github.com/gsood-gaurav
in linkedin.com/in/gsood-gaurav/

Data science professional with 8 years of experience in building machine learning models, translating complex data into actionable insights that directly improve operational efficiency. Possessing 14 years of industry experience, I excel at collaborating with cross-functional teams to pinpoint critical business needs and align data science solutions to achieve optimal results.

SKILLS

GenAI   LLMs   NLP   Time Series Forecasting   Reinforcement Learning

Python   PyTorch   TensorFlow/JAX   Julia   Flux/Lux   NumPy   SciPy   Matplotlib

Communication   Leadership Skills

EXPERIENCE

Sr. Data Scientist

HP Inc. May 2021 – Ongoing

📅 Bangalore, India

- Created Large Language Models(LLMs) Based application for webpage content filtering and topic recommendation, adhering to key NFRs which helped in acheiving **Net Promoter Score of 65** during beta testing of HP's PrintAI program.
- Benchmarked LLM based application's performance on accuracy, cost(token usage) and latency, comparing OpenAI, Llama and Mistral, reducing cost by **30%**.
- Mentored **3** data science interns on automated triaging using Graph Neural Networks and LLMs, completed within an expedited time frame of **2 months**.
- Supervised Small Language Models(SLMs, less than **3B** parameters, **4bit** quantized) exploration and **instruction finetuning** initatives using PEFT acheiving an accuracy of **95%**
- Created Telemetery data analysis pipline processing **0.5 milllion** customer record using LLM embeddings, BERT, dimensionality reduction(UMAP, Variational AutoEncoder) and clustering HDBScan, enhancing operational efficiency by **20%-30%**.

Prompt Engineering   GPT2   TinyLlama   FineTuning   Embedding Models   Graph Nerual Networks

Sr. Software Engineer Machine Learning

HP Inc. July 2017 – May 2021

📅 Bangalore, India

- Analyzed the effects of key macro economic indicators on long and short term forecasts of product sales data which helped in acheiving on-time delivery rate of **90-95%**
- Implemented Hyperparameter Tuning module for Time Series models using Optuna reducing the AutoML framework execution time by **20%-25%**
- Worked on HP Personal Systems User Behavior Modelling using Tree Based Models, SHAP, UMAP, HDBScan for target marketing campaign.
- Developed simulator environment for online Reinforcement Learning approach to automated hardware diagnostics
- Mentored **4** interns on Sequence Tagging Models using BERT, LSTMs, Conditional Random Fields for discovering insights in call center records which lead to **25%-30%** increase in operational efficency.
- Analyzed **0.1 million** call center record using Bayesian Networks for predictive diagnostics as an aid to call center executives and evaluate their decision making process

Time Series Analysis   Regression   Reinforcement Learning   Probabilistic Graphical Models   Data Annotation

PUBLICATIONS

- 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 AND CERTIFICATIONS

Deep Learning Specialization

DeepLearning.ai Oct-2021

Reinforcement Learning Specialization

University of Alberta - Alberta Machine Intelligence Institute Sept 2021

MSc Research Machine Learning

Indian Institute of Science Bangalore 2006-2009

- 📕 Speech Recognition using Support Vector Machines