

GAURAV SOOD

Sr. Data Scientist

@ gsood.gaurav@gmail.com

+91 9632714987

Bangalore, India

github.com/gsood-gaurav

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 Natural Language Processing and Reinforcement Learning. 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 Skills

EXPERIENCE

Sr. Data Scientist

HP Inc. June 2021 – Ongoing

Bangalore, India

- Designed and Developed Large Language Models Based application for web content filtering achieving 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 achieving an accuracy of 0.96
- Developed Telemetry data analysis pipeline using Large Language Model embeddings, sentence transformers, UMAP, Variational AutoEncoder and HDBScan

Prompt Engineering

GPT2

FineTuning

Probabilistic Modelling

Leadership Skills

Graph Neural Networks

Embedding Models

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 achieving 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 Reinforcement 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. 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 Telemetry 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 delivered Imaging Processing Solutions using Python Imaging Library, TesseractOCT to over 10 PSU banks in india

Algorithms

OpenSource Software

Image Processing

Leadership Skills

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

Asynchronous Programming

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 Recognition

Indian Institute of Science Bangalore 2006-2009

- 📖 Speech Recognition using Support Vector Machines
-

BTech Electronics and Communication Engg

GNE Ludhiana 2001-2005