Hargurjeet Singh Ganger

gurjeet333@gmail.com | +91 9035828125 | Bangalore, India
linkedin.com/in/hargurjeet/ | github.com/hargurjeet | gurjeet333.medium.com

Summary

Experienced IT professional with **15+ years** in the industry, specializing in **data science, statistical analysis**, and **machine learning and Generative AI**. Expert in LLMs, AI model development. Proficient in **Python**, **SQL**, and cloud platforms like **AWS** and **GCP**, with expertise in building and deploying scalable ML and Agentic solutions.

Skills

Statistics & Machine Learning– Regression analysis, statistical inference, hypothesis testing, supervised and unsupervised learning, ensemble methods, feature engineering, boosting trees, model evaluations.
Model Evaluation & Interpretability: SHAP, LIME, Explainable AI, Precision-Recall, ROC-AUC, RMSE, MSE,, Prompt Alignment, Answer Relevancy, Faithfulness, Context Precision/Recall, Hallucination, Bias, Toxicity.
Generative AI – Vector DB, RAG patterns, Guardrails, PII Filtering, Agents, Fine-tuning, Observability, MCP
LLM Framework and API's – Langchain, Langgraph, CrewAI – tool augmented workflow, multi agent systems.
MLOps – Model deployments, Monitoring and version control, CICD pipelines, Feature store, Data drift, Model Governance, Docker, FastAPI, Gitlab, MLflow, YAML, Alerting and Logging, OpenAI, Anthropic, Gemini API's.
Cloud Platforms – AWS (SageMaker, Lambda, Bedrock, Textract, OpenSearch, Step Function, Cloud Watch)
NLP - Dense/Sparse Embeddings, Named Entity Recognition(NER), Topic Modelling, BERT/DistilBERT.
Programming & Tools: Proficient in **Python** (pandas, NumPy, Scikit-learn, TensorFlow, PyTorch), **SQL**, **PySpark**, and **SparkSQL, Linux** for end-to-end data workflows and machine learning pipelines.
Contributor @stackoverflow, @jovian and @PyTorch communities.

WORK EXPERIENCE

Senior Data Scientist – British Telecom (BT)   (May 2022 – Present)
- Developed and deployed an AI-powered conversational chatbot using LLMs and Retrieval-Augmented Generation (RAG), enabling natural language queries and reducing manual data extraction time by 70%.
    - Built a data pipeline leveraging **AWS Textract** and **pdfplumber**, extracting contractual data with **90%+ accuracy** from PDFs, images, and scanned documents.
    - Designed scalable data pipelines with **AWS Textract**, **pdfplumber**, and **OpenSearch** to process **100K+** multimodal documents with 90%+ accuracy, integrating AWS Bedrock for seamless LLM interactions.
    - Design and implemented multistep agentic workflow using **Crewai**, integrating tool augmented pipelines for batch processing and applying **guardrails** to minimize hallucinations..
    - Developed an **LLM evaluation framework** to assess **prompt alignment, answer relevance, toxicity, bias, and hallucination detection**, leveraging the **Ragas** framework for structured evaluation and scoring.
    - Deployed solutions on AWS with containerisation (Docker), integrated CI/CD pipeline and implemented monitoring using custom logging to track model performance and latency tracking.
- **BT** aimed to increase revenue from premium products like SD-WAN in a competitive networking market.Develop a model to identify and recommend premium products to existing customers.Engineered a **multi-label recommendation model** using **Random Forest** and **XGBoost**, analyzing customer purchase patterns and behaviors.Increased premium product sales by **10%**, directly contributing to BT's strategic objective of growing revenue and strengthening market presence.
- **BT** sought to enhance customer retention and revenue by upselling Value-Added Services (VAS) to its existing customer base. Design a recommendation system to identify and promote relevant VAS offerings.Implemented a **sophisticated recommendation system** using **market basket analysis** and **apriori algorithms** to uncover actionable customer purchasing patterns.Achieved a **30% increase in VAS sales**, driving significant revenue growth and boosting BT's market share.

Data Scientist – [Royal Dutch Shell](#) (Sep 2016 - May 2022)
- Strong experience in **Python** for data wrangling/analysis using **Pandas, Numpy, matplotlib, seaborn**...etc and built **Power BI** dashboard to forecast Materials on-time delivery in **5 geographies** within **1 year saving 10%** of the budget allocations.
- Developed **ML models** (**XGBoost**, **Random forests**, **Decision trees**) for predictive maintenance of types of equipment present at oil refineries resulting in 30% reduction in maintenance cost and a 25% reduction in unplanned downtime impacting workforce efficiency.
- 5+ years of Experience with databases and big data analytics technologies in the industry including **Data Warehousing/ETL** and **relational Databases**.
- Applied statistics techniques such as **classification, clustering, statistical inference**, and understanding of central tendency. Worked with DL libraries scikit-learn, TensorFlow, Keras, PyTorch...etc.
- Experience in working Agile, Scrums, Sprints...etc exposure to version control via GitHub.

IT Analyst – [TCS](#) (Dec 2010 - Aug 2016)
- Performed **System Integration Testing** & **User Acceptance Testing** (UAT) to validate client PoS system.
- Spent one year in the UK guiding offshore teams with the implementation of new PoS software.
- Experience working with card and payment systems, PCI standards and ISO 8583 protocols.

Research Thesis ([View](#))

Investigated the integration of contextual language models (**GPT-3.5, Mixtral, Llama 3.1**) with classical ML models (XGBoost, Random Forest) to enhance predictive accuracy. Developed a methodology to convert tabular data into enriched textual representations and applied PCA for input optimization. Highlighting potential of LLMs in addressing data imbalance.

OPEN SOURCE DATA SCIENCE PROJECTS ([View all](#))

Agentic Search Platform ([View](#))
- Architected an **MCP-driven agentic search service** using FastAPI, enabling structured context propagation, tool invocation, and LLM orchestration for real-time document intelligence workflows.
- Implemented a **deployable backend-first system** with FastAPI APIs, Streamlit UI, async data scraping, and environment-based configuration, demonstrating production-grade API design.

Other Projects: Blog generator using Llama2([View](#))([app link](#)), *Olympics Dataset Analysis*([View](#)), *Traffic Sign Classifier*([View](#)), *Detecting Fake News on Social Media*([View](#)), *Anime Recommendations*([View](#)), House prices Predictions([View](#)), *Bank Note Classifier* ([View](#))

EDUCATION

[LIVERPOOL JOHN MOORES UNIVERSITY](#) 2023 – 2025
M.S. in Machine Learning & Artificial Intelligence

[International Institute of Information Technology Bangalore](#) 2022 – 2023
Executive Post Graduation in Data Science and Artificial Intelligence
Courses: Statistics & Probability, Python, Machine Learning, NLP, Neural Networks, MLOps

[New Horizon College Of Engineering](#), Bangalore, India 2006 – 2010
Bachelor of Engineering, **Electronics and Communication** from **Visvesvaraya Technological University**.

CERTIFICATION ([git repo](#))

I hold multiple certifications in Azure, GCP and Machine learning. Please visit my git repo to view all