

# Hargurjeet Singh Ganger

gurjeet333@gmail.com | +91 9035828125 | Bangalore, India

[linkedin.com/in/hargurjeet/](https://linkedin.com/in/hargurjeet/) | [github.com/hargurjeet](https://github.com/hargurjeet) | [gurjeet333.medium.com](https://gurjeet333.medium.com)

## Summary

---

Data Science Leader with **15+ years** of experience translating complex **machine learning, generative AI**, and advanced analytics into measurable business outcomes. Deep expertise in LLMs, agentic systems, and scalable ML solutions across AWS and GCP. Known for building high-performing teams and driving impact.

## 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** - Transformers, Vector DB, RAG, Guardrails, AI Agents, LangChain, LangGraph.

**Cloud Platforms and MLOps** - AWS (SageMaker, Lambda, Bedrock, Textract, AWS OpenSearch, Step Function, Cloud Watch), GCP (Vertex AI, BigQuery), MLflow, Airflow, GitLab CI/CD, Docker.

**NLP** - Word Embeddings, POS tagging, Named Entity Recognition(NER), Topic Modelling, BERT/DistilBERT.

**Programming & Tools:** Proficient in **Python** (pandas, NumPy, Scikit-learn, TensorFlow, PyTorch), **SQL**, **PySpark**, and **SparkSQL** for end-to-end data workflows and machine learning pipelines.

Contributor @[stack overflow](https://stackoverflow.com/users/11449400/hargurjeet), @[jovian](https://jovian.ai/@hargurjeet) and @[@PyTorch](https://pytorch.org) communities.

## WORK EXPERIENCE

---

### Senior Data Scientist - [British Telecom \(BT\)](#)

(May 2022 – Present)

- Led end-to-end strategy, architecture, and delivery of an enterprise-grade **conversational AI platform**, leading a team of **4 data scientists** from concept through production deployment.
  - Defined the technical vision to build a data ingestion pipeline leveraging **AWS Textract** and **pdfplumber**, establishing standards to achieve 90%+ extraction accuracy across PDFs, scanned documents at scale.
  - Integrated **AWS OpenSearch** to retrieve multimodal embeddings, handling over **100,000+ docs**.
  - Used **AWS Bedrock** to enable seamless interactions with LLMs for text and image tasks.
  - Improved processing efficiency by implementing **AI agents with LangGraph** to process documents in batches across various independent features, enhancing scalability and modularity.
  - Implemented **guardrails** to reduce hallucinations, ensuring accurate outputs during model inference.
  - Developed an **LLM evaluation framework** to assess key safety and quality metrics such as **prompt alignment, answer relevance, toxicity, bias, and hallucination detection**, leveraging the [Ragas framework](#) for structured evaluation and scoring.
  - Delivered a **production ready AI-powered chatbot** that streamlined document search and retrieval, reducing manual data extraction time by **70%** and accelerating decision-making.
- **BT** aimed to increase revenue from premium products like SD-WAN in a competitive networking market. Led a team of analysts and engineers to develop 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.
- Partnered with senior sales and marketing stakeholders to define requirements and co-own the go-to-market strategy for a Value-Added Services (VAS) recommendation system, leveraging market basket analysis and Apriori algorithms to surface actionable customer insights.
- Orchestrated collaboration across data engineering, product, and sales teams to drive seamless execution, resulting in a 30% increase in VAS sales and measurable expansion of 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.
- Awarded “**Star of the Month**” twice and promoted from Systems Engineer to IT Analyst within 3 years.

## Research Thesis

---

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 using Langraph ([View](#))

- Developed an interactive **Streamlit application** using **LangGraph** and **LangChain** for AI-driven web search, enabling detailed query processing with customizable workflows.
- Designed a pipeline leveraging **TavilySearchResults tools** for real-time data retrieval and processing, with step-by-step output visualization in the UI.

**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