SAGAR DHIMAN

Gen AI | Data Scientist | Machine Learning | Data Engineer

Mohali, India | +91-8437003676 | sagardhiman033@gmail.com

PROFESSIONAL SUMMARY

Senior Data Scientist with **7+ years of experience in Machine Learning and NLP**, specializing in building scalable **GenAI & LLM-powered systems**. Proven expertise in end-to-end AI development, from designing robust **Data Engineering pipelines** to deploying production-grade models. Strong background in large-scale information retrieval, cloud architecture, and technical leadership, bridging the gap between complex data infrastructure and advanced AI solutions.

TECHNICAL EXPERTISE

AI, ML & LLM Systems: Large Language Models (LLMs), Graph-Augmented RAG, NLP Pipelines, Semantic Search, Recommendation Systems, Anomaly Detection.
Data Engineering: Data Pipelines, ELT/ETL, Clickstream & Time-Series Processing, Scalable Architectures.
Databases & Cloud: Neo4j, ElasticSearch, Snowflake, Amazon Redshift, MongoDB, PostgreSQL, AWS, Azure.
Programming & Tools: Python, SQL, PySpark, Docker, Apache Airflow.

WORK EXPERIENCE

Tatras Data   May 2018 – Present
Senior Data Scientist   Mohali, India

- *Client Engagement: Haven Safety (Backed by Andrew Ng's FUND.ai)*
    - **AI Leadership:** Serving as Lead AI & Data Engineer (since Aug 2025), contributing to the core AI copilot focused on safety intelligence and investigation workflows.
    - **Graph-Based AI Reasoning:** Designed graph-augmented LLM systems combining industry knowledge graphs, unstructured data, and AI reasoning to support incident analysis.
    - **LLM-Driven Pipelines:** Built AI-centric pipelines where LLMs extract, structure, and enrich investigation data into graph models, enabling explainable, context-aware intelligence.

- *Additional Responsibilities at Tatras Data*
    - **Chatbot Development:** Designed and deployed multi-RAG chatbots using graph-based and semantic similarity retrieval for complex, multi-source enterprise datasets.
    - **Sensor Data Analytics:** Developed semantic fingerprinting methods for sensor historical data, enabling advanced semantic search and time series forecasting.
    - **Project Leadership:** Led a team to scrape and analyze global cuisine data. Developed ML models to classify cuisine types and standardize item names across vendors.

Associate Data Scientist
- **Data Engineering:** Built and optimized ELT pipelines using AWS S3, Glue, and Snowflake, resulting in significant cost savings.
- **Information Retrieval:** Implemented Transformer-based models for large-scale textual data retrieval and abstractive/extractive summarization.
- **Recommendation Systems:** Developed scalable systems using Word2Vec, LDA, and Knowledge Graphs for keyphrase and company suggestions.
- **ETL & API Development:** Automated ETL pipelines for MongoDB/PostgreSQL; deployed REST APIs (Flask, FastAPI) via Docker on AWS.

Junior Data Scientist
- **Clickstream Engineering:** Engineered pipelines with Druid and Kafka, enabling real-time analytics of high-volume user behavior data.
- **Text Recommendation:** Built scalable text recommendation systems with TF-IDF, LDA, and Word2Vec, integrated with Elasticsearch.
- **Keyphrase Extraction:** Developed and deployed solutions using unsupervised and supervised (custom NER) methods on large text datasets.

Trainee Data Scientist
- **Image Processing:** Designed CNN-based models for pneumonia detection and YOLO-based license plate recognition systems.
- **Duplicate Bug Detection:** Conducted large-scale duplicate detection using TF-IDF and topic modeling, optimizing for MAP and Recall.

    – **Web Scraping:** Automated price comparison scrapers handling high data volumes across multiple platforms.

Sabudh Foundation   Jan 2019 – Present
Mentor
- Mentor students in recommender systems, unsupervised learning, NLP, and AI in healthcare.
- Guide students on end-to-end project pipelines and organize monthly assessments.

EDUCATION

B.Tech, Computer Science and Engineering   2014 – 2018
Guru Nanak Dev Engineering College, Ludhiana
12th CBSE   2013 – 2014
Kendriya Vidyalaya

CERTIFICATIONS

- AWS Certified Cloud Practitioner
- AWS SageMaker Practical – Udemy
- Dataiku ML Practitioner, Core Designer, Advanced Designer, MLOps Practitioner
- Spark & Python for Big Data (PySpark)

LEADERSHIP & RESPONSIBILITIES

- Teaching Assistant – Indian School of Business (ISB), Mohali
- Project Lead for multiple enterprise AI initiatives
- Regular mentor for junior data scientists and interns