Ashish Rathore

Data Scientist | asrathore871@gmail.com | +91-9478933503

LinkedIn | GitHub | HuggingFace

Summary

Results-driven Data Scientist with 4 years of experience specializing in Machine Learning and Generative AI, supported by a robust foundation in Computer Science. Adept at integrating innovative Generative AI methodologies with classical ML approaches to address complex, real-world challenges. Recognized for a creative mindset, experimental approach, and commitment to staying abreast of advancements in AI. Seeking impactful positions at the intersection of traditional and cutting-edge artificial intelligence.

Employment History

Data Radian Technologies

Data Scientist | August 2022 – Present

Smart Energy Water

Data Science Engineer | July 2021 – July 2022

Projects

AI-Powered Product Insight Engine

- Developed a scalable AI solution processing over 100,000 customer reviews across 50+ retail products, extracting actionable insights by identifying key themes, product issues, and user patterns.
- Introduced classical NLP and unsupervised techniques (stemming, lemmatization, clustering, embeddings, cosine similarity) to group reviews by common themes (Version 1).
- Transitioned to an LLM-driven pipeline utilizing OpenAI's GPT models, advanced embeddings, and prompt engineering, resulting in insight accuracy improvement from ~65% to ~93% (Version 2).
- Deployed using UnionAI and Flytekit for orchestration; implemented an automated GPT-based QA framework to ensure output reliability.
- Designed an interactive Streamlit dashboard for real-time visualization, enabling product managers to efficiently explore trends.
- Created a conversational insight agent via Microsoft Copilot Studio for Microsoft Teams, facilitating natural language queries and dynamic discovery of insights.

Key Tools: Python, OpenAI LLM (GPT-3.5/4o), Embeddings, Clustering Algorithms, Streamlit, Snowflake, SQL, UnionAI, Flytekit, Docker, Copilot Studio

Claims Intelligence Agent

- Built an AI copilot agent to streamline water damage claims validation, shifting from manual data checks to instant natural language queries.

- Integrated semantic retrieval from Snowflake data, reducing investigation time from 20–30 days to under 2 days.
- Seamlessly deployed within Microsoft Teams using Copilot Studio for enhanced workflow integration.

Key Tools: Python, Copilot Studio, Snowflake, MS Teams

Product Sense – RAG-Based Product Information Assistant

- Designed a GenAI proof-of-concept answering product-related inquiries using Snowflake's Cortex LLM.
- Automated data extraction from 30+ product pages with Crawl4AI for optimized Retrieval-Augmented Generation pipelines.
- Achieved accurate, context-aware responses with hybrid search and reranking.

Key Tools: Python, OpenAI, Crawl4AI, Snowflake, LLM, RAG

Additional Notable Projects:

- **AI Web Search Engine (Perplexity Replica):** Python, LLM, Groq Cloud, Streamlit, Langchain, Brave Search
- **Team GPT:** Collaborative AI chatbot platform for productivity. Python, LLM, Ollama, Streamlit
- **Snowflake SQL Assistant:** Python, LLM, Snowflake, ReAct, Streamlit

Energy Disaggregation

- Developed machine learning solutions to break down household energy consumption into appliance-level usage for optimization.
- Addressed complex data challenges using open-source datasets and domain-specific techniques from academic research.
- Delivered client-facing POCs demonstrating substantial business impact.

Key Tools: Python, EDA, Pandas, Numpy, Sklearn, Matplotlib, Seaborn, Random Forest, XGBoost, KMeans, Time Series Analysis, Azure Cloud, GitHub

AI Faucet Search Engine

- Built an AI-powered tool to verify faucet certifications and assess regulatory compliance via aggregated certification data.
- Enhanced transparency and efficiency in product safety evaluations and competitor analysis.

Key Tools: Python, OpenAI, Crawl4AI, Snowflake, Snowflake Cortex

---

Technical Skills

- **Programming & Tools:** Python, SQL, GitHub, Docker
- **Data Science:** Exploratory Data Analysis, Statistics, Hypothesis Testing, Supervised & Unsupervised Learning (Linear/Logistic Regression, Decision Trees, Random Forest, XGBoost, KMeans, DBSCAN, GMM, PCA, TSNE, UMAP), Recommender Systems, Time Series Forecasting, NLP
- **Generative AI:** LLMs, RAG, Agents, ReAct, Prompt Engineering, Chatbots, Diffusion Models, LoRA, DreamBooth, CrewAI, Knowledge Graphs, Neo4j, Copilot Agents, Tool/Function Calling, Langchain, Langraph

---

Education & Certifications

- **M.Tech. in Data Science & Machine Learning**

  Thapar University, Patiala | Graduated June 2021 | CGPA: 8.57

- **B.Tech. in Computer Science & Engineering**

  Punjab Technical University | Graduated May 2018

- **Certifications:**
  - Generative AI with LLMs, Deeplearning.AI
  - Applied Machine Learning

Applied AIML Case Studies

- Solved 10+ real-world machine learning problems in sectors including healthcare, e-commerce, and EdTech applying techniques such as regression, classification, clustering, and forecasting with end-to-end data science workflows. Developed a replica of a web search engine (Perplexity) using LLMs for enhanced search accuracy and speed.

Other Projects:

- Built a Snowflake SQL assistant using ReAct framework and function calling to combine reasoning and tool use for accurate SQL generation.
- AI-Powered Product Insight Engine (Perplexity replica)
- **Team GPT** — Collaborative chatbot platform for team decision-making.
- **Resume Ranker** – Ranks candidates based on the Job description