

# B NIKHIL SHARMA

## Data Scientist

+91-8602784255 , +91-7987203910

<https://www.linkedin.com/in/b-nikhil-sharma>

Bangalore, India 560047

@ b.nikhil.sharma@gmail.com

<https://github.com/NIKHILSHARMA5896>



## EXPERIENCE

### Data Scientist / GenAI Engineer

#### HSBC

01/2024 - Present Bengaluru, India

##### Project: RAG-based Intelligent Banking Assistant

- Developed a scalable PoC using **LlamaIndex** and **ChromaDB** to automate banking FAQs and policy-related queries
- Ingested and indexed **{~500} banking documents** (PDFs, DOCX, web FAQs) using **LlamaIndex**, creating **{~20,000} dense text chunks** for semantic retrieval
- Leveraged **ChromaDB** and **multilingual-e5-large-instruct** embeddings to enable fast, accurate top-5 document retrieval with average latency of **{<200ms}** per query
- Used **Gemini-1.5-flash** to generate high-quality answers from retrieved chunks, achieving **{~85%} answer match rate** against manual ground truth
- Deployed a simple **Streamlit interface** for internal demo use, with real-time query input; handled **{100+} test queries** in trial runs

### Data Scientist / GenAI Engineer

#### HSBC

01/2023 - 12/2023 Bengaluru, India

##### Project : Data Quality Report Generation using Unsupervised Learning & Pretrained Embeddings

- Developed a **dynamic, unsupervised learning system** to generate **monthly data quality reports** for customer attributes (e.g., names, Dobs, document IDs) across **multiple regions and source systems**
- Improved detection of data quality issues by {3x}**, uncovering inconsistencies that were previously missed by static rule-based systems
- Used **multilingual-e5-large-instruct** from **Hugging Face** to generate semantic embeddings for deeper text similarity
- Implemented **dimensionality reduction techniques** to optimize embedding space for efficient and meaningful clustering
- Evaluated clustering performance using **Silhouette Score** and **DBI** to ensure robust group separation and cohesion
- Reduced data quality **issue identification time by {95%}** by replacing static rules with clustering algorithms
- Enabled stakeholders to resolve inconsistencies across **{15} source systems**, improving **data integrity by {35%}**

### Junior Data Scientist

#### HSBC

09/2021 - 12/2022 Bengaluru, India

##### Project: Payments Fraud Detection for the UK Market

- Contributed to building a payments fraud detection machine learning system for the **UK market**
- Reduced **false positive alerts by {30%}**, enabling investigators to focus more on promising cases
- Led efforts in **data preparation**, including **cleaning and standardizing transaction data**, handling **missing values**, removing **duplicates**, and ensuring consistent **timestamp and datatype formatting**
- Explored and trained multiple **classification algorithms** to identify fraudulent patterns
- The model achieved a **precision of {~0.89}**, **recall around {0.77}**, an **F1-score of {0.83}**, and an **ROC AUC of {0.93}**

### Junior Data Scientist - Intern

#### HSBC

03/2021 - 08/2021 Bengaluru, India

- Uncovered **{5} key patterns** from datasets through **exploratory data analysis**, enhancing data-driven decision-making
- Visualized data trends using tools such as **Matplotlib, Seaborn, or Plotly**
- Assisted in **data cleaning** and **preprocessing** for model development
- Developed interactive dashboards leading to **{50%} quicker team insights** using QlikSense

## SUMMARY

I am an accomplished Data Scientist / GenAI Engineer with extensive experience in financial crime detection. My work at HSBC has focused on enhancing data integrity and predictive accuracy through advanced machine learning models. Proficient in Python and Machine Learning, I have developed innovative solutions that streamline operations and improve team development. I am passionate about leveraging technology to drive measurable results

## KEY ACHIEVEMENTS



### RAG Chatbot

Built a RAG-based banking assistant PoC, enabling accurate LLM responses from 500+ internal documents



### Data Quality Enhancement

Built a multilingual, unsupervised learning system using sentence-transformers and clustering to automate monthly data quality reporting



### Enhancing Financial Crime Detection

Developed and tuned machine learning models to reduce false positives in UK payments fraud screening

## SKILLS

Python SQL Machine Learning

Deep Learning Data Analysis

Prompt Engineering NLP Generative AI

LlamaIndex RAG DVC PyTorch

Hugging Face Transformers LLMs AWS

Docker FastAPI Git GitHub MLflow

PySpark Airflow Qlik Sense

## PERSONAL PROJECTS

### Vehicle Insurance Prediction

<https://github.com/NIKHILSHARMA5896/MLOps-Project-Vehicle-Insurance>

Built a machine learning pipeline to predict whether customers will buy insurance, using their demographic and behavioral data

EDUCATION

---

Master of Technology – Industrial Safety Engineering

National Institute of Technology - Tiruchirappalli

📅 08/2019 - 05/2021    📍 Tiruchirappalli, India

---

Bachelor of Engineering: Mechanical Engineering

Shri Shankaracharya Group of Institutions - Bhilai

📅 08/2014 - 05/2018    📍 Bhilai, India