





# Sachin Bahuleyan

 [GitHub](#)  [Linkedin](#)  [sachin270895@gmail.com](mailto:sachin270895@gmail.com)  [+91-9990015516](tel:+91-9990015516)

## EDUCATION

### IIT Mandi

M.Tech Computer Science

Aug 2022 - May 2024

CGPA: 8.25

## SKILLS

Python, Pytorch, Statistics, Machine learning, Deep learning, Data science, NLP, LLMs, RAG, Pyspark, Linux, Langchain, Git/GitHub, Problem Solving

## EXPERIENCE

### Data Scientist - Opinium.ai

May 2025 - August 2025

- Developed an AI engine integrating **BERT fine-tuned** for sentiment analysis, **GNN** and **knowledge graph** to predict the impact of social media content.
- Improved the company's knowledge graph with relevant data for product enhancement by building a **web scraping tool** for RSS news feeds, **leveraging text embedding models** and python tools.

### Data Scientist- D2K Technologies | Statistical Analysis, Predictive Modeling

Aug 2024 - March 2025

- Distance to Default (A Credit Risk Model)**
  - To gain understanding of loan portfolio conducted **Exploratory Data Analysis (EDA)** on Dena Bank's loan data, identifying patterns and insights.
  - Designed and deployed **classification and regression models** using **Databricks** and **MLflow**, achieved a **91% success rate** in predicting defaults.
  - Built an interactive dashboard** in Dash Plotly for monitoring and visualizing insights, to be used by the data science team and banking stakeholders.
- Analyzed client's loan disbursement data** to generate insights and reports using pandas, plotly, and matplotlib.
- Delivered a RAG-based POC** for client - querying financial reports using LLMs (OpenAI GPT, LLaMA, Gemini, HuggingFace).
- Assessed tools such as **MLflow**, **DVC**, **ClickHouse**, and **Azure Databricks**, identifying a good fit for building ML pipelines.

## PROJECTS

### Two-Tower Recommendation System | Pytorch, MLflow, FAISS, SQLite

 [Project Link](#)

- Developed and trained a two-tower model for product search using **contrastive loss**, leveraging the **Amazon ESCI dataset** and **multilingual embeddings**.
- Implemented efficient product retrieval by indexing product embeddings using **FAISS**, enabling **fast similarity search** for user queries passed through the trained query tower.
- Managed the **model lifecycle using MLflow** for logging and serving, including the creation of model serving endpoints for deployment.

### End to End RAG Based Chatbot | Python, LangChain, FAISS, Flask

 [Project Link](#)

- Developed a **RAG-based chatbot** using the **Langchain** framework.
- Sentence embeddings created with MiniLM v6 from Hugging Face for efficient text processing.
- Utilized FAISS for vector database retrieval, enhancing search efficiency within the chatbot.

### Airbnb New User Booking | Python

 [Project Link](#)

- Predicted new users' first booking location to enable Airbnb to deliver personalized content and improve demand forecasting.
- Conducted comprehensive EDA to analyze data distribution, outliers, and feature relationships.
- Trained and evaluated models (Random Forest, Decision Trees, XGBoost), achieving a 0.86 public score on Kaggle.

## POST GRADUATE PROJECT

### Reinforced Bot Behavior | Unity 3D, C#, Logitech G29 Driving Wheel, VS Code

May 2024

- Developed a realistic driving environment** in **Unity 3D** with Logitech G29 integration for advanced agent control.
- Trained agents using reinforcement learning (PPO)** and imitation algorithms in dynamic multi-agent scenarios.

## RELEVANT COURSEWORK

Statistical Methods, Deep Learning, Advanced Data Structures & Algorithms, Linear Algebra, Probability

## ACHIEVEMENTS

**GATE 2022: Achieved 98 percentile**