Siddham Jain

Passionate data scientist with 4 years of hands-on experience, fervently dedicated to building innovative data science solutions and leveraging them for impactful business applications

jainsid07@gmail.com | +91 9829399778 | LinkedIn

EDUCATION

BS, CHEMICAL SCIENCES

INDIAN INSTITUTE OF TECHNOLOGY, BOMBAY
2016 - 2020 | Mumbai, India
Minor: **Entrepreneurship**

SKILLS

PROGRAMMING

Python, PySpark
SQL, R, Javascript
C, C++, HTML

PACKAGES/LIBRARIES

Pandas, Numpy, Sklearn
Keras, Tensorflow
NetworkX, h2O, NLTK, re

TECHNOLOGY

Software Development
Big Data
Data Storytelling
Natural Language Processing
Statistical Modelling
Computer Vision
LLM
Graph Neural Networks

SOFTWARE

Databricks
Amazon Web Services
EMR Studio
Spark and Hadoop
Git
Superset
Airflow

CERTFICATIONS

Deep Learning, Coursera
Generative AI, Coursera
Intermediate Spanish, Duolingo

INTERESTS

Cryptocurrency, Sports
Movies, Music

EXPERIENCE

FREELANCER

May 2024 – Present

- Utilize **Reinforcement Learning from Human Feedback (RLHF)** to help machine learning models learn more efficiently Ensure models adhere to legal and ethical obligations (ie. harmlessness, truthfulness, safety, etc.) Prompt engineering with various parameters and producing assessments of results Quality Assurance for other analysts' work (prompts, feedback, conversations, instruction-following, etc.)
- Designed and developed a **full-stack travel-focused** application, including frontend and backend systems.

SLICE | SENIOR DATA SCIENTIST

March 2022 – May 2024 | Bengaluru, India

- **The Network product**
  - Led a comprehensive network analysis solution using contacts data; built multifaceted tools for pattern observation, community identification & influencers/outliers detection.
  - Applied network analysis in various domains, including:
    * **Fraud Detection:** Identified **outliers** within a community by observing patterns where users were connected unidirectionally, detected **10k** individuals involved in scams & fraudulent activities.
    * **Influencer Analysis:** Identified key influencers at the center of communities and initiated a refer-and-earn campaign. Successfully increased the first transaction rate of the UPI product by **5%**.
- **Long Term Risk Model for Personal Loan Product**
  - Engineered an underwriting model for Personal Loan product using **device specific data** like SMS, location, contacts & app usage.
  - Achieved a substantial increase of **10%** in approval rates while simultaneously reducing the risk of default by **4%**.
- **ID Skip Logic using Location based clusters**
  - Addressed a significant **20%** drop-off rate during the college ID upload step in the product funnel for student users by developing ID skip logic.
  - Leveraged location and network data,creating **density-based user DBscan clusters** to identify the likely college location, resulting in a **26%** increase in ID skipped users and improved user funnel engagement.
- **Competitor Defaulter**
  - Crafted an NLP-based approach to identify defaulters among slice's competitors, using users' SMS data, employing a **sentence transformer model** and pretrained word embeddings for SMS classification.
- **Salary Estimation from SMS**
  - Enhanced the salary estimation methodology by leveraging user SMS data, estimating both **primary and secondary income** of a user.

HDFC BANK | DATA SCIENTIST

Aug 2020 – Feb 2022 | Mumbai, India

- Implemented a robust deep learning model for **crop classification** using **satellite images**, enhancing the precision of agricultural loan assessments.