Shashank Shekhar

New York, NY, 10033 | 7039302509 | shashankbond1012@gmail.com | LinkedIn

SUMMARY

Data Scientist with 2+ years of experience applying AI and machine learning to real-world problems. Skilled in building chatbots, predictive models, and data-driven solutions that improve performance and decision-making.

TECHNICAL SKILLS

Programming & Development: Python, SQL, Flask, Git
Machine Learning & AI: Supervised & Unsupervised Learning, RAG, Vector Database, S3
Data Analysis & Visualization: EDA, Feature Engineering, Tableau, Storytelling
Databases, Cloud & Infrastructure: PostgreSQL, AWS

EDUCATION

Katz School of Science & Health - Yeshiva University   **Jan 2024 - May 2025**
Master of Science, Data Analytics & Visualization (GPA: 3.6/4)   *New York, NY*
  • **Coursework:** Machine Learning, Data Science, Structured Data Management, Computational Math's & Statistics, Storytelling

IIIT Bangalore   **Apr 2023 - Dec 2023**
Advance Certificate Programme, Data Science (GPA: 3.4/4)   *Bangalore, INDIA*
  • **Coursework:** Machine Learning, Data Toolkit

Guru Nanak Institute of Technology   **May 2018 - May 2022**
Bachelors, Computer Science Engineering (GPA: 3.6/4)   *Kolkata, INDIA*
  • **Coursework:** Machine Learning, Data Structure

WORK EXPERIENCE

People Tech Group | *Data Engineer*   **Sep 2025 – Present | New York, NY**
  • Designed an AI-based document search tool that connects SharePoint files with AWS Kendra for search and Bedrock (Claude 3) for summaries.
  • Used secure login and tagging so users only see allowed files, making document search faster and easier.

Peblink | *Data Scientist Intern*   **May 2025 – Aug 2025 | New Jersey, NJ**
  • Analysed viewer behaviour with AI-assisted data mining and predictive modelling to pinpoint churn triggers and content performance factors across 700K+ users.
  • Generated data-driven personas using unsupervised learning and demographic clustering to support targeted customer focused marketing strategies.
  • Created a Tableau Dashboard to analyse the continuous behaviour of the customer to find out who is our targeted customer.

Yeshiva University, Shevet Glaubach Centre | *Data Scientist Intern*   **Jan 2025 - May 2025 | New York, NY**
  • Developed and deployed a GPT-4o RAG chatbot using Lang Chain and Chroma DB, improving answer relevance by 40% and reducing hallucination by 28%.
  • Built data pipelines to scrape, process, and store academic content into Chroma DB vector databases, enabling faster and more relevant search results.

Harman Connected Services Pvt. Ltd. | *Data Scientist*   **Apr 2022 - Dec 2023 | Pune, INDIA**
  • Collaborated with engineering and product teams to enhance ImedOne using data insights and software improvements, reducing diagnostic errors by 24% and improving overall system reliability.
  • Executed a predictive model on 50,000 mental health records using Python and scikit-learn; SVM achieved 91% precision in identifying key patterns
  • Enhanced diagnostic accuracy for high-risk groups, cutting null error rate from 69.34% to 24%.

PROJECTS

Mobile App Development| *GitHub*   **Aug 2025 - Sep 2025**
  • Integrated AI-like logic into the app to provide smart delivery ETA predictions, order assistant flows, and proactive notifications tailored to user behaviour.
  • Designed and developed a full-featured React Native food delivery app using Expo SDK, supporting user authentication, real-time order tracking, and secure payments.

NYPD Calls for Service Data | *GitHub*   **Sep 2024 - Dec 2024**
  • Researched NYPD calls for service data in Tableau, revealing Brooklyn as the highest-crime borough with peak incidents from 12 am to 3am.
  • Delivered borough-specific crime insights through Tableau dashboards, enabling data-backed decisions by non-technical stakeholders and resource planners.

Disease Data Modelling and Warehousing Project | *GitHub*   **Mar 2024 - May 2024**
  • Architected a scalable data model in PostgreSQL with a dimensional structure and ETL pipelines to optimize disease data storage and streamline data flow.
  • Created ER diagrams using dB Schema and built dashboards for visualizing disease trends, enhancing healthcare decision-making

CERTIFICATIONS

  • Advance Certificate Programme in Data Science