Ranjeet Nagarkar

Email | LinkedIn | GitHub | San Francisco, CA

Data scientist with **over 4 years of experience** in industry and academia.

Education

Master of Science in Data Science ***Aug 2023 – Aug 2024***
University of San Francisco *San Francisco, CA*
Relevant Coursework: Machine Learning, Data Acquisition, Statistics, Data Analysis, Relational Databases (SQL), Data Ethics

Bachelor of Technology in Chemical Engineering ***Aug 2015 – Aug 2019***
Indian Institute of Technology Madras,
Chennai*, India*

Professional Experience

Machine Learning Engineer, Internship Startup ***Oct 2023 – Present***
DataKnobs *Remote*

- Scraped and modified over 1000 website templates, creating a scalable and cost-efficient data pipeline for LLM-based website generation.
- LLM Website generation product and concept expected to generate more than $1million in company revenue
- Estimated cost savings of $300,000 with enhanced dialogue management for improved security and oversight of chatbot products.
- Presented an innovative modular approach for html page creation at **Forbes CIO conference**, generating significant interest and potential for a multi-million dollar business opportunity.
- Designed a scalable chatbot framework using DAG architecture and langchain agents library with chain of thought reasoning.
- Chatbot facilitated client interaction with Data Knobs products, using prompt engineering, and Advanced RAG techniques.

Junior Data Scientist, consultancy ***July 2021 – March 2023***
GyanData *Chennai, India*

- Developed a Forecasting model, Natural Language Processing(NLP) and Image classification(CNN) models
- Data Analysis and modeling for company clients saving an estimated INR 20 million in equipment procurement cost.
- Identified potential problems in the client-vendor selection process as well as possible solutions amounting worth INR 100 million

Data Scientist and Application Developer, Research ***Feb 2020 – Apr, 2021***
Indian Institute of Technology Madras *Chennai, India*

- MATLAB Application development of the XPCA Application project with more than 5000+ lines of code
- The project tackled the business challenge of early fault detection and process monitoring savings of 9% in operating costs.
- Model validation techniques for error-in-variable sensor data, real time data analysis

Select Projects

- **LLM website builder**
  - Spearheaded the development of an LLM (Large Language Model) website builder application using langchain and llamaindex.
  - Implemented Advanced RAG (Auto merging retrieval), prompt engineering, and prompt fine-tuning to optimize LLM response.
- **LLM agent dialogue managed chatbot**
  - Chatbot dialogue management expected to save $300,000 in dialogue management optimization for Dataknobs
  - Orchestrated the creation of an LLM agents managed chatbot, employing DAG graph architecture
  - Conducted end-to-end integration of LLM chatbot functionalities, incorporated Advanced RAG for client vector database retrieval.
- **ARD Process Optimisation project**
  - Streamlined the annual procurement process for the Indian Navy by developing an equipment consumption forecast model.
  - Resulted in an expected savings of INR 20 lakh.
  - Deployed a fine-tuned BERT model to categorize reasons for order rejection comments, achieving an accuracy rate of 77%
  - Deployed a CNN image classifier for payment receipts categorization achieving an accuracy of 81% using image augmentation
- **Chemical Processes modeling Application**
  - The application resulted in an estimated savings of 9% in fault detection and process monitoring
  - Facilitated model creation, validation, enhancing efficiency in fault detection, saving approximately 10000+ man hours
  - Engineers use the application for analyzing Error-in-Variable time series sensor data from chemical reactors using Dynamic Iterative Principal component Analysis

Technical Skills

- Python, SQL (PostgresSQL, Microsoft), Git, GitHub, PyTorch, BeautifulSoup, Selenium, Matplotlib, Langchain, LLamaIndex, PGVector