Ranjeet Nagarkar

+1 (415)-579-6333 | San Francisco, CA 94107 | ranjeet.sn96@gmail.com | linkedIn
Portfolio Website | Medium | GitHub

EDUCATION

University of San Francisco **July 2023 - June 2024**
Master of Science (MS) in Data Science *San Francisco, CA*
Indian Institute of Technology Madras **July 2015 - May 2019**
Bachelor of Technology *Tamil Nadu, India*

SKILLS

Programming/ Visualization:	**Python, SQL,** NoSQL, HTML, Linux, **Tableau, Power BI, Looker**, Excel, PowerPoint, **Airflow, MATLAB**
Machine Learning:	Supervised, Unsupervised, Clustering, **Deep Learning,** NLP, Transformers, **Large Language Models (LLM)**
Big Data/Database:	**PySpark, SparkSQL, Spark,** ETL Data Pipeline, MongoDB, Hadoop, Github, SVN, **Snowflake**, Databricks, Hive
Cloud/ MLOps:	Azure, **AWS (SageMaker, EC2, EMR, S3),** GCP(Pub/Sub, Big Query, Composer), Docker, Flask, CI/CD pipelines
Libraries and framework:	**Pytorch, TensorFlow,** NumPy, Pandas, Plotly, Scikit-Learn, Matplotlib, Seaborn, Spacy, Scipy, Selenium
Web Experimentation:	**Causal Inference, A/B Testing,** Hypothesis testing, Statistical Modeling, Optimization, Experimental Design
Data Science Experience:	**Supply Chain Management**, Transportation Management, Model Deployment, Forecasting

BUSINESS EXPERIENCE

DataKnobs | **Machine Learning Engineer** | San Francisco, USA Oct 2023 – Present

DataKnobs: LLM website builder

- Created a LLM website builder using **langchain** and **llamaindex** library using **Auto-merging Retrieval RAG** using modular approach
- Deployed the website builder on **GCP firebase** with the product projected to surpass **$1 million** in revenue for the company
- Presented the innovative modular approach for html webpage creation at **Forbes CIO** conference generating substantial interest
- Performed **cross-validation** on window size hyperparameter Auto-merging retrieval for **4 LLM metrics** using **TruLens** library

DataKnobs: LLM agents chatbot

- Designed a scalable chatbot framework using DAG architecture and LlamaIndex agents library with **ReAct** prompting
- Estimated cost savings of **$300,000** with enhanced dialogue management for **ReAct** chatbot products
- Designed and Implemented **A/B** tests on chatbot interface using **Google Optimize**, elevating user retention by **8%**

GyanData | **Data Scientist** | Chennai, India June 2021 – March 2023

Indian Navy ILMSAir: Inventory Management System

- Analyzed **~20M** rows of Data on Snowflake using Apache Spark in databricks to extract insights on daily equipment procurement
- Optimized procurement process saving an estimated INR **2 million** and identified potential logistical bottlenecks worth INR **10 million**
- **Employed K-means** hierarchical clustering to group equipment with akin consumption patterns reducing maintenance cost by **~7-8%**
- Deployed **Gradient Boosting** Regressor, engineered 20+ lagged variables, multi-step time series forecasting reducing procurement by **~5%**

Indian Navy CNAMS: Transportation Management System

- Analyzed **~10M** rows of data in **Hive** using **PySpark**, with data streamed and persisted in **Apache Cassandra** for delivery driver insight
- Developed a **Random Forest Regressor** model for driver availability prediction, improving forecast accuracy by **25%** over baseline
- Deployed the predictive model, leveraging **Google Kubernetes Engine** for scalability and real-time model monitoring
- Implemented MLOps with **Docker-deployed** ML models and automated reporting for monitoring concept, data, and model drift.

QBSS insurance: Insurance process expedition

- Performed POC for Named Entity Recognition(**NER**) using BioclinicalBERT for medical reports for insurance claim processing
- **Communicated** with medical experts to incorporate Domain rules into the model improving F1 score from 66% to **85+%**

Indian Institute of Technology Chennai | **Research Data Scientist** | Chennai, India Feb 2020 – April 2021

Robert Bosch RBCDSAI: Research on predictive maintenance

- Built MATLAB Application for **Incremental PCA** early fault detection and process monitoring with savings of **9%** in operating costs
- Implemented fault detection techniques for **error-in-variable** sensor data and real time process monitoring using **Datafeed Toolbox**

PROJECTS

LLM based Twitter complaint resolution project- *End to End production deployment using LLMOps tools Medium Blog | GitHub Repository*

- Used Amazon Kinesis to stream tweets, sentiment analysis using **LSTM** and complaint categorization using tuned **TwHIN-BERT** model
- Built DAG architecture chatbot application for complaint resolution, using **LLM agents** in LlamaIndex library
- Used PGvector for storing vector databases of company data for sentence-window Retrieval RAG response, AWS deployed

Medium recommendation system end-to-end- *Two-tower recommender system using MLOps tools Medium Blog | GitHub Repository*

- Scraped data from medium API, storing data in Google bucket, creating collections in MongoDB using Airflow Composer
- Built a recommender system using Two-tower model to recommend articles to users
- Created a Content based filtering recommender to recommend similar articles to users

MATLAB predictive maintenance Application- *Research MATLAB application using DIPCA algorithm GitHub Repository*

- Created application for analyzing Error-in-Variable time series sensor data from chemical reactors using DIPCA

ACHIEVEMENTS & AWARDS

- Placed 4th out of 300 for Multi-modal LLM **GPT-4 vision** application for Customer complaint resolution on ClarifAI hackathon(link)