

# Sairindhri Bhattacharya

Tel: 408-891-2814 | [sairin@umich.edu](mailto:sairin@umich.edu) | LinkedIn: [www.linkedin.com/in/sairindhri](https://www.linkedin.com/in/sairindhri) | Github: <https://github.com/sairin94>  
**Data Enthusiast in AI/ML with 5 years of Work Experience located in Santa Clara, CA with a valid work permit**

## SKILLS:

**Programming Languages:** Python, SQL, Java, Linux, HTML, CSS, JavaScript, jQuery, Typescript, Linux

**Database And Datawarehouse:** Postgres, MySQL, AWS Redshift, Snowflake, Elastic Search

**Big data Technologies:** Apache Spark, Hadoop, HDFS, Apache Kafka, Docker, Kubernetes

**ETL:** AWS Glue, Apache Airflow, DVC, AWS Matallion

**Cloud Stack and Web Development:** AWS, Bootstrap, Angular 7, Web Service (REST API), Flask, Google API

**Tool:** Jira, Confluence, Git, Tableau

**Machine Learning:** TensorFlow, Pytorch, LLM, Mlops, Sklearn, NumPy, Pandas, Jupyter, NLP, Recommendation System, MLflow

## WORK EXPERIENCE

**University Of Michigan, Ann Arbor**

*Research Assistant*

**Michigan, USA**

**May 2023 – Dec 2023**

**Economic Dynamics of Conflict: Modeling U.S. War-GDP Impact (Capstone)**

**Oct 2023 – Dec 2023**

- Researched 'Impact of America's Involvement in Global Wars on U.S. GDP' for Capstone Project using **LSTM, GRU, 1D CNN** for multivariate time series-based GDP prediction.
- Applied advanced data analysis and modeling, including **sliding window technique** and **hyperparameter tuning**, to provide critical insights into the economic effects of global conflicts on GDP. Link <https://github.com/sairin94/war-economic-prediction>

**Restaurant Radar: Machine Learning Web Application**

**May 2023 – Sep 2023**

- Handled **6M+** Yelp reviews, set up an AWS data pipeline to an S3 Data Lake, and structured data with **Athena** using SQL, while addressing data challenges
- Utilized **BERT** and fine-tuned the **Llama-2-7B model** in **PyTorch** for detailed sentiment analysis, topic modelling from the reviews.
- Designed a web app showcasing restaurants on Google Maps with sentiment visuals, integrated with **Flask**, and deployed on **AWS LightSail**, with **DVC automation**. Link: <http://yelp-analysis.com>

**Tata Consultancy Services**

**India**

*Data Engineer - full-time*

**2017 – 2021**

- Designed and implemented scalable data pipeline around Apache Kafka, Apache Spark, Apache Airflow integrated with AWS Redshift, processing over 10TB of data daily and deployed using Docker and Kubernetes container.
- Managed and optimized relational databases like PostgreSQL and MySQL, ensuring high availability and performance.
- Created PySpark scripts to interact with Hadoop data lake for data transformation, performed ETL to provide specific insights from diverse set of data and store extracted data into data lake.
- Optimized SQL queries and database structures, achieving up to 50% improvement in response time.
- Developed interactive and customized dashboards in Tableau and Altair, presenting key performance indicators (KPIs) such as customer acquisition cost (CAC), customer lifetime value (CLV), and customer churn rate to stakeholders.
- Designed star and snowflake schema data marts, improving query performance by 20%.
- Integrated machine learning models into MLOPs CI/CD pipelines, leading to 20% churn reduction and a 34% increase in customer loyalty.

## EDUCATION

**University of Michigan, School of Information, Ann Arbor, MI**

**Jan 2022 – Dec 2023**

M.S. in Applied Data Science

GPA: 3.95/4.00

**West Bengal University of Technology - Kolkata, India**

**Aug 2012 – Aug 2016**

Bachelor of Technology in Computer Science

GPA: 3.85/4.00

## PROJECTS

• **Urban Safety Insights: Analyzing NYC Traffic Accident Trends and Interventions**

Analyzed 2.63M NY records using **AWS Postgres** and conducted **multivariate regression** to link car types to accident rates and used **geospatial clustering** for pinpointing accident hotspots. Through hypothesis testing and **A/B testing**, devised intervention strategies tailored to NYC street segments.

• **Price Pulse: Scalable House Price Prediction Model**

Designed a scalable pricing model using **PyTorch**, achieved an RMSE of 0.15 and a cross-validation score of 0.91 with robust MLOps pipelines using **MLflow**, covering the entire spectrum from data preprocessing to model training, deployment, and regular retraining. Proactively tackled data drift issues to guarantee model robustness and relevance.

**CERTIFICATION:** Post Graduate Diploma in Data Science and Business Analytics from Texas Austin