

# SAURABH SAMPATKUMAR LOYA

+1 (801)-680-5240 | [s.loya21@gmail.com](mailto:s.loya21@gmail.com) | [linkedin/saurabhloya](https://www.linkedin.com/in/saurabhloya) | [saurabhloya.github.io](https://saurabhloya.github.io)

## SKILLS AND CERTIFICATIONS

**Languages:** Python, SQL, R, Java, JavaScript

**Databases:** MySQL, PostgreSQL, MongoDB, ChromaDB, Pinecone

**ML/Data:** NumPy, Pandas, PyTorch, TensorFlow, Scikit-Learn, Spacy, LangChain

**Frameworks:** Flask, Django, Apache Spark, Apache Kafka, Apache Airflow

**Visualization:** Tableau, Power BI, Matplotlib, D3.js

**Certification:** AWS Certified Cloud Practitioner, Azure Fundamentals

## EDUCATION

**Master in Computer Science**, The University of Utah (GPA: 3.6/4) | Salt Lake City, Utah Aug 2023 – May 2025

**Bachelor in Computer Science and Engineering**, MIT World Peace University (GPA: 3.8/4) | Pune, India Aug 2017 – May 2021

## WORK EXPERIENCE

**Data Scientist**, UHealth, Salt Lake City, Utah Aug 2025 – present

- Built an AI agent to automatically summarize **875+ narrative medical student performance evaluations**, reducing manual review effort for clerkship directors.
- Designed and deployed interactive dashboards with **row-level security**, enabling **125+ students and 50+ faculty and administrators** to securely access personalized performance insights.

**Data Scientist Intern**, BMW Bank of NA, Salt Lake City, Utah Jan 2025 - May 2025

- Prepared a high-quality dataset by performing **data mining, data analysis, and cleaning** efforts using **Python and SQL**, enabling the development of a machine learning-driven loan optimization model impacting \$100M–\$300M in daily originations.
- Engineered **40 features from 688 raw attributes** by applying feature engineering techniques to high-dimensional, unstructured data using **Pandas** and **NumPy**, enhancing model accuracy and data usability.
- Built interactive **Tableau** dashboards to visualize loan performance trends, enabling the Credit and Risk team to make informed decisions on a \$9B+ financial portfolio.

**SDE Intern**, H7 BioCapital, San Francisco, California June 2024 - Sept 2024

- Crafted a full-stack scheduling application using **Django, Python** and **SQL** to streamline investor–startup interactions, reducing email coordination time by ~60%.
- Integrated **Google Calendar & Zoom API** to automate meeting creation, reminders, and calendar between investors and startups.
- Translated business requirements into well-defined technical stories in **Jira**, supporting **agile** development and implementation.

**Data Scientist/Software Engineer**, Volkswagen Group Technology Solution, Pune, India Aug 2021 - July 2023

- Integrated **LLM** and **NLP** into an internal **chatbot**, enhancing human-like responses and expanding its knowledge base by 30%.
- Optimized an **OCR** pipeline using **Azure** Cognitive Services, achieving 8x faster document processing and 15% accuracy gain.
- Engineered a scalable SaaS analytics platform for multi-region survey management using **Python, SQL** and **Tableau**.
- Innovated a **ETL** Data pipeline solution using **Python, Selenium**, and **Apache Kafka** to replace RPA processes, reducing manual efforts, task time and cutting RPA license costs by over **\$15,000 annually**.
- Resolved 100+ production issues, including **data quality and performance** and led the refactoring of a 2 complex projects.
- Contributed to 3 AI/ML proof-of-concept projects with Center of Excellence team, focusing on automation and intelligent systems.

**Data Scientist**, Venkatesh Agencies, Maharashtra, India May 2020 – May 2021

- Analyzed multi-source sales and inventory data for a medical distributor, covering **200K+ SKUs** and 50+ pharmacy partners.
- Built interactive dashboards using **Tableau** to monitor real-time sales, forecast shortages, and visualize seasonal demand.
- Automated data pipelines using **AWS Lambda** and **Snowflake** for scalable, real-time reporting across business units.

## PROJECTS

**CITI Bike Rental Data Analytics and Forecasting:**

- Performed **exploratory data analysis** on 5M+ ride records to uncover seasonal patterns and usage trends across boroughs.
- Built a demand forecasting model using **Apache Spark** and Prophet, improving inventory planning and station rebalancing.

**Android Malware Detection System:**

- Trained a malware detection model from scratch using **Perceptron, Logistic Regression, SVM**, and **Neural Networks**.
- Achieved **0.89 F1 score** in classifying malicious apps based on system call features.

**LLM-Based Medicine Recommendation System**

- Developed a **RAG** system using **Llama 2, LangChain**, and **ChromaDB** to recommend medications from symptoms.
- Fine-tuned embeddings and prompts to achieve **0.87 F1 score** in providing relevant medication suggestions.

## HONORS AND RECOGNITION

- Innovation Award** in recognition of creative thinking in automating training pipeline of chatbot using LLM.
- Achievers Award** for successful project deliverable with Customer Satisfaction Survey rated above 9.