SOHAM PATEL

Irvine, California | (929) 642-7699 | sohamhp@uci.edu | LinkedIn | GitHub

EDUCATION

University of California, Irvine, **CA** **Sep 2023-Dec 2024**
Master of Data Science, GPA: 3.67/4
Relevant Coursework: Artificial Intelligence, Machine Learning, Statistics, Database Management, Bayesian Modelling, Big Data.

SRM Institute of Science and Technology, Chennai, India **Jun 2019-May 2023**
Bachelor of Technology in Computer Science Engineering, GPA:3.94/4
Relevant Coursework: Machine learning, Statistics, Database Management, Artificial Intelligence, Data Structures.

SKILLS

- **Programming**: Python, C, R, C++, MATLAB, SQL
- **ML & DL**: scikit-learn, Keras, PyTorch, TensorFlow, Neural Networks, Transformers, LLMs, XGBoost, NLP.
- **Data Analytics and Visualization:** Excel, Tableau, PowerBI, ArcGIS, Pandas, NumPy, Matplotlib, Seaborn.
- **Data Engineering**: ETL Pipelines, MLOps, MySQL, PostgreSQL, Apache Spark, Neo4j, MongoDB, Cassandra, Flink.
- **Statistical Analysis:** Distributions, Regression, Bayesian Inference, A/B testing, Time series Analysis, Optimization.
- **Cloud & Tools:** SASS, SPSS, AWS, Azure, GCP, Airflow, OpenCV, Docker, Jenkins, GitHub, CI/CD, Kubernetes, Langchain.

WORK EXPERIENCE

Machine Learning Intern || LiveGood Inc, Irvine, CA **Sep 2024-Dec 2024**
- Developed and implemented **Autoencoder** and **MLP** models to predict health trends, influencing strategy and policy adjustments.
- Enhanced model deployment using **Azure Cloud**, integrating real-time data visualization with **PowerBI**.
- Performed data analysis on health outcomes using **Python** and **SQL**, guiding public health initiatives.
- Automated data preprocessing pipelines in **Python** using **Pandas** and **NumPy**, increasing data quality and model reliability.

Data Science Research Intern || UCI-Health, Orange, CA **Jun 2024-Sept 2024**
- Developed an **AWS-powered ETL pipeline**, enhancing data integration and analytics, using Python for scripting and **AWS Lambda** for real-time data processing.
- Applied **TensorFlo**w and **Keras** to refine **predictive models** for patient diagnosis, improving model accuracy by **40%**.
- Created a **RAG**-based chatbot using finetuned **GPT 3.5** and **Pinecone** trained on comprehensive burn-related medical literature.
- Conducted **A/B testing** on model variants to optimize predictive performance and ensure robustness in clinical environments.

Machine Learning Engineer || Strategic Alliance, India **Dec 2021-May 2023**
- Pioneered **TensorFlow** solutions for **real-time passenger detection** boosted revenue by **30%** and improving operations.
- Developed a **YOLO-based social distancing monitoring tool** using Python and **OpenCV**, achieving **95%** accuracy in real-time compliance tracking and enhancing public safety through **edge computing deployment**.
- Utilized Docker and Kubernetes for efficient model deployment and scaling across multiple platforms.
- Built custom **data pipelines** for data ingestion and preprocessing, leveraging **Apache Spark** for high-volume data handling, which reduced latency by 20%.

PROJECTS

PFAS Analysis using Advanced ML (Group Research Project) || Olivares Lab, UCI **Jun 2024-Present**
- Employed **Apache Spark** for distributed **big data processing**, applying machine learning models such as **SVM**, **Decision Trees**, and **Random Forests** to predict environmental impacts and enhance data-driven decision-making.
- Utilized Python and R for data analysis, developing insights that directly influenced environmental policies.
- Created interactive data visualizations using **Tableau** and Python (**Matplotlib, Seaborn**) to effectively communicate analytical findings, enabling data-driven decision-making and supporting community engagement in policy development.

Healthcare Demand Forecasting System **Jan 2024-Jun 2024**
- Engineered a forecasting model using **time series analysis** in Python, leveraging **ARIMA** to predict healthcare demand with **95% accuracy**, utilizing data preprocessing techniques in Pandas for optimal input quality.
- Integrated model outputs with **AWS** to enhance data scalability and real-time accessibility.
- Presented findings to healthcare managers, influencing staffing and resource allocations to meet forecasted demand peaks.

YOGDAAN (Emergency dispatch system using ML workflows) **Aug 2022-Jul 2023**
- Led the development of a machine learning platform using **Python, R**, and **scikit-learn** to optimize emergency dispatch operations, applying predictive modeling techniques to reduce response times by **20%** and improve resource allocation efficiency.
- Implemented **Apache Spark** to process **real-time emergency data**, enhancing system responsiveness during peak times.
- Integrated **AWS** technologies like **Lambda** and **Elastic Load Balancing** to manage system load effectively.