Shravani Hariprasad

+1 (619) 673-2889 hariprasadshravani@gmail.com shravani-h-507280177 shravani-01 Portfolio

SUMMARY

Data Scientist with **around 3 years of experience** delivering ML-driven solutions across fintech and transit domains. Experienced in performing **statistical analysis** and developing **predictive models**, **NLP pipelines**, and **time series forecasting** for large-scale financial datasets, as well as building AI systems for transit agencies worldwide to support asset monitoring, demand forecasting, and real-time decision support. Proficient in **C++, Python, JavaScript, SQL, Scikit-Learn, PyTorch, and cloud platforms (AWS, GCP, Databricks)**, with hands-on expertise in **GenAI** (LLMs, RAG, LangChain). Skilled in **MLOps** and model deployment (SageMaker, FastAPI), alongside strong capabilities in data visualization using Tableau and Power BI. Adept at cross-functional collaboration, **A/B testing**, and translating complex data into actionable business insights.

EXPERIENCE

Hatch **Feb 2024 – Present**

Data Scientist - Transit Agencies (MBTA, MassDOT, CTA, Metro Transit, etc.) *Boston, USA*

- Architected and supported scalable, production-grade information systems on **AWS (S3, ECS Fargate, EventBridge)** integrating SQL-based data stores, **ML services** (**XGBoost**, AUC = 0.87), and downstream analytics, improving Hours of Service compliance monitoring by 20%. Enabled business users through enterprise **BI dashboards** (**Tableau + TabPy**), reducing reporting latency by 40%.
- Designed and deployed real-time, multi-platform monitoring systems integrating edge devices (Raspberry Pi), GPS, video streams, and cloud messaging (Pub/Sub), supporting networked asset health and **predictive maintenance** workflows that reduced asset downtime by 15%.
- Developed and deployed **web-based ML services** and **RESTful APIs** for incident classification and asset condition monitoring, enabling system-to-system integration between ML pipelines and enterprise reporting tools; reduced manual data entry by 80% and accelerated response times by 25%.
- Built and maintained **computer vision** pipelines (YOLOv8 on Vertex AI GPUs) for transit asset inspection, later extending to multimodal AI systems (**LLaVA, Gemini**) for enriched condition assessment (94% accuracy), supporting enterprise asset management decision-making.
- Created **TransitGPT**, a secure **RAG**-based internal web application for engineering document QA using PyMuPDF, SQL-backed metadata, BERT-based reranking, and REST interfaces—improving retrieval precision from 82% → 90% while meeting data privacy and compliance requirements.
- Supported operations, maintenance, and asset planning teams by modeling terminal operations using **discrete-event simulation** (**Salabim**) and delivering system KPIs via web dashboards (**Apache ECharts, Looker Studio**), driving a 15% throughput improvement.
- Led **CI/CD** and data platform support using **dbt + Airflow**, maintaining production data pipelines, technical documentation, and governance policies to ensure reliable access for analytics, reporting, and downstream enterprise systems.

L&T Infotech **Jul 2021 – Jul 2022**

Senior Software Data Engineer - Citi Bank *Mumbai, India*

- Automated natural language commentary variance analysis in financial reporting using **Python**, cutting manual report creation time by 60%.
- Reduced false positives by 20% in **anomaly detection** by implementing **Isolation Forest** for variance breach monitoring, outperforming manual methods.
- Designed deployed large-scale **ETL pipelines** with **Python, PySpark (Databricks, Hadoop)** and **Azure** (**ADF, Data Lake, SQL DW)**, scaling to 500M+ records daily and reducing data prep time by 25%.
- Optimized **SQL** and **SparkSQL** queries through partitioning, caching, and indexing, improving query performance by 30% across financial datasets.
- Leveraged **VLOOKUP** and **advanced Excel** functions (Macros, VBA) to analyze deposit trends by cheque type and other financial indicators.
- Automated repetitive tasks using **CI/CD** pipelines with **Jenkins and Git**, ensuring consistent delivery of updates to financial dashboards and reports.
- Orchestrated **Power BI** dashboard creation, improving agent productivity by 50% and providing real-time insights into key performance indicators (KPIs) using **DAX, Roles, and Row-level security**, resulting in faster, data-driven decision-making.
- Collaborated with enterprise application teams to integrate **Python- and SQL-based analytics services** with **.NET**-based internal applications, supporting **data exchange** with SQL Server and ensuring compatibility with existing ASP.NET enterprise systems.
- Collaborated with cross-functional teams to elicit, analyze, and document business requirements using Agile methodologies such as **Scrum** and **Kanban**.

Cognizant Technology Solutions **Aug 2020 – Jun 2021**

Data Engineer Co-op *Chennai, India*

- Developed SQL logic queries for aggregate views using **MySQL Workbench** and **AWS-RDS database**, efficiently managing and extracting insights from a substantial 1TB dataset.
- Detailed data governance requirements, leveraged **Python, Apache Spark and AWS services** (S3, Amazon Glue - ETL processes) to automate tasks and achieve a 5% reduction in data processing time for a 1TB dataset, significantly enhancing workflow efficiency.
- Utilizing **Tableau**, I identified and rectified 7+ gaps in risk analysis, aligning them precisely with client requirements. This streamlined processes by 20%, enhancing decision-making efficiency.

SKILLS

Languages	Python, R, Java, JavaScript, SQL, HTML, CSS
Frameworks	Pandas, scikit-learn, TensorFlow, Keras, PyTorch, NLTK, Django, React.js, .NET, Java Spring Boot
Tools	Excel, Tableau, PowerBI, SPSS, Jira, Hadoop, Spark, Hive, Agile, Git, DevOps, Docker, GitHub, Looker
Database	MySQL, PostgreSQL, MongoDB, Google BigTable
Cloud Technologies	AWS (EC2 Instance, S3 Bucket, Glue, RedShift, SageMaker), Google Cloud (Big Query, DataProc), Azure

EDUCATION

San Diego State University **Aug 2022 – May 2024**
Master of Science in Big Data Analytics(GPA : 4/4) *San Diego, California*

Anna University **Aug 2017 – Jun 2021**
Bachelor of Engineering in Electronics and Instrumentation(GPA : 3.85/4) *Chennai, India*

ACCOMPLISHMENTS

- One of the five recipients of the **2023-2024 SDSU Master's Research Scholarship ($10,000)** for outstanding academic performance and contributions to research.
- Completed Automation Anywhere and Xceptor RPA certifications.
- **Techgium Competition 2019-20** - Certificate of Recognition for innovative concept on Multilingual Text Extraction from Videos using YOLO by Larsen & Toubro Technology Solutions.
- Awarded **Academic Excellence Certificate** for securing 4.0 GPA at SDSU by the Big Data Analytics Department

RESEARCH AND PROJECTS

TailTalk Advisory: AI-Powered Pet Care Chatbot(Team Leader) **Jan 2024 – May 2024**

- Developed a user-friendly Petcare AI app using **React.js**, integrating the state-of-the-art **Llama2** model from **Hugging Face** and pretrained over 100 documents on Petcare.
- Utilized **AWS SageMaker** for scalable and cost-effective model training and deployment, ensuring reliability in managing machine learning workflows within the app's backend infrastructure.
- Implemented **Retreival Augmented Generation (RAG)** for retrieving the embeddings from **Pinecone** vector database for a given user query and generating response using LLMs.

Deep Learning Image Classification with Google Street Views (GSV) - HDMA Lab **Jun 2023 – May 2024**

- Curated a comprehensive dataset from **Google Street View**, and crowdsourcing platforms like **Mapillary** focusing on key downtown streets in San Diego. Extracted significant metadata for sidewalk-based analysis to identify homelessness and their encampments.
- Built a low latency **Computer Vision** model YOLOv8 with high performance, showcasing the potential for real-world applications in addressing homelessness and urban planning.

Content Analysis of HIV Prevention Drug - PrEP(Co-author) **Mar 2023 – Ongoing**

- Co-authoring an **NIH grant-funded** project analyzing public perceptions of PrEP through social media.
- Leveraged Natural Language Processing (**NLP**) techniques, including **topic modeling**, to analyze social media discussions on PrEP, identifying that 40% of conversations centered on concerns about accessibility.
- Visualized and created 5+ dashboards for different categories using **Tableau**, presenting 10+ major characteristics about PrEP.
- Anticipating a 20% increase in awareness, a 30% reduction in stigma, and data-driven recommendations for tailored interventions.

Image Caption Generation of the MSCOCO Dataset (Team Leader) **Feb 2023 – May 2023**

- Built a full-end ML project in **PySpark** on GCP, handling the COCO dataset and optimizing ETL workflows with AirFlow, improving data processing time by 50%. Pre-trained an **Xception-LSTM** model for image captioning, reducing loss to 1.85% and achieving a BLEU score of 0.35.

Optimizing Patient Care: Snowflake Healthcare Analytics on AWS **Dec 2022 – Feb 2023**

- Implemented a patient length of stay (LOS) prediction model using **Amazon SageMaker**, automating status notifications, and integrating real-time predictions into the Snowflake database for streamlined operations and improved resource allocation.

Human Activity Recognition (Team Leader) **Aug 2022 – Dec 2022**

- Examined machine learning techniques for diverse human activity prediction in hospital environments with both raw **time-series** and feature-engineered data.
- Attained an impressive 96.47% accuracy and a minimal 3.3% error rate on feature-engineered data with ML models like **Linear SVC**, outperforming the LSTM **deep learning** model.