Shreyas Durairajalu

San Jose, CA 95112 | +1 (408) 646-0386 | shreyasdurairajalu@gmail.com| [Linkedin](#) |[Github](#)

EDUCATION

San Jose State University   **San Jose, California**
Master of Science, Applied Data Intelligence   **January 2025 - Present**
- **Coursework:** Mathematical methods, Data Warehouse & pipeline, Distributed Systems for Data Engineering, Machine Learning, Business Intelligence and Data Visualization, Big Data Technologies, Deep Learning, Generative AI.

PROFESSIONAL EXPERIENCE

Senior Data Analyst|EXL Health   **January 2022 – December 2024**
- Delivered 89% accuracy in bill-type classification by implementing optimized K-means clustering models with PySpark.
- Performed exploratory data analysis (EDA) to identify key drivers of overpayments in healthcare claims. Developed a predictive model to flag high-risk claims, enabling proactive audit selection and reducing financial leakage; rigorously validated model performance.
- Deployed Hive and PySpark pipelines to run MoM overpayment audits, uncovering $300K in contract-level discrepancies.
- Managed analytics for a $150M+ revenue generating focus group by developing a provider level framework that enhanced clinical finding accuracy and audit precision.
- Automated concept development using Selenium and Python, and built Databricks BI dashboards to monitor audit performance and data spend, saving 10+ hours/week and enabling data-driven claim selection decisions.

Software Intern| Buddi.ai   **June 2021 – December 2021**
- Enhanced Buddi.ai's proprietary Medical Language Processing (MLP) engine structured & unstructured EMR notes, PDFs, and HL7 data into a Clinical Contextual Graph, achieving over 95% accuracy in clinical information extraction.
- Built RESTful APIs with FastAPI for backend integration, improving processing time by 25%.
- Designed an HL7 parser to extract and normalize structured healthcare data, increasing system interoperability.
- Wrote unit test cases to ensure robustness and accuracy of the data extraction pipeline, improving reliability and reducing debugging time during integration.

PROJECTS

RAGQuant: Retrieval-Augmented Insight System for Quant Strategy Debugging
● Engineered a modular, object-oriented RAG framework using LLaMA, FAISS, and Python to analyze strategy logs, configuration changes, and market data for real-time debugging and insight generation.
● Designed scalable API layers and integrated hybrid retrieval components to support semantic traceability, version comparisons, and strategy performance explanations.

Agentic AI System for Web-Scale Healthcare Policy Discovery (Open Source @ EXL Health)
● Developed an agentic AI system using LLaMA, RAG, and LangGraph to autonomously extract, analyze, and track healthcare policy updates from CMS and web-scale sources.
● Integrated web scrapers, FAISS/Pinecone vector stores, and knowledge update modules for real-time monitoring, semantic retrieval, and concept library enrichment.

SKILLS

Languages & Databases:
 Python, SQL, Cypher, Neo4j, HTML, CSS, C++, TypeScript,Java
Libraries, Tools and other skills:
 Apache Spark, PySpark, Hive, Databricks, AWS (S3, Redshift, Lambda), Apache Airflow, Kafka, Docker, Kubernetes, Spark MLlib, MongoDB, Neo4j, Tableau, Power BI, Redis, Node.js, React, Pandas, NumPy, SciPy, SymPy, Matplotlib, Git, GitHub, MySQL, PostgreSQL, Hadoop, ETL, Prefect, Snowflake, ServiceNow, FastAPI, Flask, Streamlit, MicroStrategy, Agile, Azure, REST API, JSON, Microsoft Excel, Alteryx, Linux, Probability, Statistics, Linear Algebra, unit tests.
Certifications: SQL for Data Science by IBM, Introduction to Machine Learning, Stanford, Deep learning.io Specialization Data Structures and algorithm from amazon coursera.