SHARDUL CHAVAN

shardulc36@gmail.com ⋄ (857)313-5138 ⋄ Union City, CA ⋄ linkedin/shardulchavan36

EDUCATION

Northeastern University, Master of Science in Information Systems, **GPA: 3.72/4.00**   Expected November 2024
Relevant Coursework: Big Data Engineering and Intelligence Analytics, Data Science Engineering

EXPERIENCE

Research Assistant, Northeastern University, Boston, MA   September 2024 - Present
Tech Stack: numpy, Pandas, Scikit-learn, Statsmodels, Data Visualization (Matplotlib, Seaborn, Plotly Dash)

- Led marketing strategy optimization and evaluated campaign effectiveness, addressing business needs and enhancing ROI
- Designed and deployed **ARIMA and SARIMA** models, improving prediction accuracy by 15% and reducing forecast error rates by 10%, enabling precise budget allocation and campaign timing
- Leveraged **Difference-in-Differences (DID)** and **regression analysis** to isolate causal relationships between marketing variables, demonstrating potential 2.8% increase in subscription rates

Data Engineer - Analytics, Skyworks Solutions Inc., Boston, MA   January 2024 - June 2024
Tech Stack: Python, SQL, Docker, ETL Development, Azure SQL Server, Databricks, PowerBI

- Constructed 25+ data pipelines using **Python, SQL, and Airflow** to extract, transform, and load (ETL) data from Azure Data Lake, SQL Server, and flat files into SQL Server Data Warehouse
- Implemented Slowly Changing Dimensions (SCD) and advanced data transformation techniques with **Databricks SQL** and PySpark APIs, improving historical tracking and data quality by 30%
- Collaborated with engineering and business teams to create data models and **PowerBI dashboards** for OPEX, product analysis, and prototype builds, maximizing efficiency across 3+ teams in Business unit
- Engineered Semiconductor product test data analysis application pipeline to extract data from raw test data files into valuable information for RF engineers, by performing data transformation and report generation which optimizes more than 5 hours of manual work on weekly basis

Graduate Teaching Assistant, Northeastern University, Boston, MA   September 2023 - December 2023
Tech Stack: Python, Pandas, scikit-learn, MLlib, HDFS, Predictive Analytics & Validation

- Spearheaded ML pipeline using **PySpark** on **Hadoop cluster** to optimize data loading, perform EDA, and automate feature selection using correlation analysis, and chi-squared tests
- Built and fine-tuned classification models, including KNN, Random Forest, and XGBoost, enhancing predictive accuracy by 25% and strengthening risk assessment and pricing strategies for Prudential Life Insurance
- Researched and benchmarked **GenAI models** (BART, Da-Vinci, Stable Diffusion), applying BLEU, ROUGE, and few-shot learning techniques, elevating performance by 10%

Data Analyst, Accion Labs, India   January 2023 - July 2023
Tech Stack: Python, JavaScript, Snowflake, MySQL, ELT, REST APIs, Natural Language Processing

- Partnered with stakeholders to understand business requirements, identified key data sources and developed analytical solutions aligned with business objectives
- Developed REST APIs using Python and JavaScript, integrated large language model APIs into ServiceNow, boosting virtual agent performance by 20%, and architected vector databases to enhance NLU in knowledge management

SKILLS

Programming & ML:	Python, R, SQL, Scikit-learn, TensorFlow, Keras, PyTorch, XGBoost, NLP(SpaCy, NLTK)
Data Processing & Cloud:	Apache Spark, Hadoop, GCP BigQuery, AWS (S3, SageMaker), Docker, Kubernetes

PROJECTS

AI-Driven News Aggregator (Streamlit, GCP, Docker, JWT, OpenAI, Git CI/CD) ⚙   November 2023 - December 2023

- Developed Airflow pipeline to scrape data, generate vector embeddings, and store text in Azure SQL and **embeddings in Pinecone,** enabling multimodal architecture for precise query responses and real-time news delivery every 10 minutes
- Leveraged **Langchain** and NLP techniques with LLM models to provide precise query responses, optimizing retrieval through similarity search and text summarization, enriching content delivery by 30%
- Deployed AI-driven news application on **GCP** with Docker and FastAPI, leveraging **Git CI/CD pipeline** to ensure scalable, production-ready infrastructure