Sachin Loddiya Karthik

517-243-9938 sachinlkece@gmail.com Sachin GitHub Portfolio Open to relocation

SUMMARY

Data Engineer with 3+ years of experience designing and optimizing scalable data pipelines that improved processing efficiency by up to 40%. Expert in **PySpark, SQL, Azure Databricks, and Synapse**, with proven expertise in transforming complex raw data into actionable business intelligence for enterprise-scale solutions. Also skilled in **applied machine learning, statistical modeling**, and **real-world experimentation** to support data-driven decision making.

TECHNICAL SKILLS

Programming Languages: Python, R, SQL (T-SQL, PL/SQL), Java, C++, HTML, CSS, JavaScript
Machine Learning & AI: Scikit-Learn, XGBoost, TensorFlow, PyTorch, OpenCV, Keras, LangChain
Statistical Analysis: Regression (Linear/Logistic), ANOVA, Hypothesis Testing, Time Series, Clustering (K-Means, DBSCAN), PCA
Data Engineering: Apache Spark (PySpark), Apache Kafka, Apache Airflow, Databricks, Delta Lake, Docker
Databases & Querying: PostgreSQL, MySQL, MS SQL Server, Azure SQL, Oracle DB, NoSQL (MongoDB)
Cloud Platforms: Azure (Data Factory, Synapse, ML Studio, DevOps), AWS (S3, EC2, Glue, Athena, SageMaker), GCP (BigQuery, Dataflow, Pub/Sub)
Visualization & BI Tools: Power BI, Tableau, Seaborn, Matplotlib, Plotly
MLOps & Tools: Git, GitHub, MLflow, Streamlit, FastAPI, VS Code, Jupyter Notebook, RStudio, PyCharm, WinSCP, PuTTY

EXPERIENCE

Data Scientist — WMU, USA **Jan 2025 – Apr 2025**

- Developed an automated class scheduling optimizer using **Google OR-Tools** and **constraint programming**, reducing manual scheduling time by **85%** and optimizing 200+ course assignments across 15 departments.
- Built and deployed a **Streamlit**-based tool to ingest and validate 50+ CSVs in real-time, cutting data entry errors by **90%** and saving over **40 hours/week** of manual effort.
- Applied **operations research** and **combinatorial optimization** techniques to improve academic scheduling efficiency.

Data Scientist — Green Expectations LLC, USA **Jan 2024 – Apr 2024**

- Designed a rule-based **Home Sustainability Scoring model** to process **1,000+ user-level data points**, improving recommendation accuracy by **35%**.
- Enhanced performance of an **NLP-powered AI chatbot** by optimizing real-time data pipelines, reducing response latency by **40%** and boosting user engagement by **25%**.
- Led **feature engineering** and data preparation efforts for intelligent sustainability predictions across user profiles.

Data Engineer — Accenture, India **Jul 2021 – Jul 2023**

- Engineered ETL pipelines using **Azure Data Factory** to reformat supply chain data into **Parquet**, improving processing efficiency by **40%**.
- Processed raw supply chain data in **Azure Databricks** from **Data Lake Gen2**, boosting accuracy by **25%** and halving transformation times.
- Developed external tables in **Azure SQL** and interactive **Power BI dashboards**, improving data access and increasing operational efficiency by **25%**.
- Integrated ADF triggers, Databricks notebooks, and CI/CD pipelines with **Azure DevOps**, accelerating delivery by **30%**.

Data Engineer — Claritrics India Pvt Ltd, India **Nov 2020 – May 2021**

- Architected an ETL pipeline using **Azure Data Factory** and **Databricks** to integrate OCR-extracted data, improving document processing accuracy by **30%**.
- Automated recurring workflows and enhanced reliability, reducing processing time by **40%**.
- Enabled quicker deployments by **30%** with modular coding practices and containerized environments.

EDUCATION

Master of Science in Data Science — Western Michigan University **Aug 2023 - Apr 2025**
GPA: 3.86/4.0 *Kalamazoo, MI*

- **Relevant Coursework:** Machine Learning, RDBMS, Azure Databricks & Spark (PySpark / SQL), Applied Linear Models, Big Data Analysis, Google Cloud Big Data & ML Fundamentals

PROJECTS

Ask My PDF – LLM-Powered Document Q&A System (LangChain, OpenAI, FAISS, PyMuPDF) **Apr 2025 – May 2025**

- Designed a domain-adaptive **Q&A system** leveraging **LLMs and vector search** to extract meaningful insights from complex documents such as government budgets and financial reports.
- Implemented **document chunking, embedding with FAISS**, and semantic search workflows to enable **context-aware natural language querying** and scalable retrieval pipelines.

Enhanced ATS Resume Expert Pro (Streamlit, Gemini Pro, PyMuPDF, scikit-learn, Plotly) **May 2025 – Jun 2025**

- Built a **Streamlit web app** that uses **Google Gemini Pro** to analyze resumes against job descriptions with semantic and keyword-based matching.
- Implemented **PDF/DOCX parsing**, **TF-IDF + cosine similarity**, and **skills gap detection** to deliver actionable ATS optimization insights.
- Created interactive dashboards with **Plotly** and enabled exportable **PDF/JSON reports** and historical tracking using **SQLite**.