

Dinesh Arasavalli

Detroit , MI | +1 (313) 403-1534 | Dineshar0820@gmail.com | [Linkedin](#) | [Github](#)

SUMMARY

Data-driven Data Science and Engineering professional with 4 years of experience designing, developing, and deploying advanced analytics, machine learning, and scalable data pipeline solutions. Expertise in supervised and unsupervised learning, predictive modeling, time series forecasting, and Generative AI using scikit-learn, TensorFlow, PyTorch, Lang chain, Hugging Face, OpenAI, and RAG pipelines. Proficient in Python, SQL, data preprocessing, feature engineering, and A/B testing to drive business outcomes. Skilled in MLOps with Docker, Kubernetes, Jenkins, MLflow, and CI/CD pipelines for seamless deployment and monitoring. Experienced in cloud-native data engineering using Snowflake, Databricks, Azure Data Factory, and AWS Redshift for scalable data warehousing. Hands-on expertise with AWS (S3, Lambda, SageMaker, EC2), GCP (BigQuery, Vertex AI), and Azure Machine Learning for end-to-end model lifecycle management. Proficient in Power BI, Tableau, Matplotlib, and Plotly for delivering interactive dashboards and actionable insights.

PROFESSIONAL EXPERIENCE

Data Engineer June 2024 - Present

Goldman Sachs

- Designed **real-time fraud pipelines** using **Apache Airflow, Pub/Sub, and Big Query**, processing **5-8M transactions/day** and enabling sub-second scoring for high-risk activity.
- Engineered **feature processing workflows** in **PySpark and Python**, supporting low-latency inference and improving fraud alert precision across streaming data.
- Developed and deployed **fraud detection models** using **Scikit-learn and XGBoost**, reducing false positives by **~20%** while meeting regulatory precision thresholds.
- Automated **model training, validation, and deployment pipelines** with Vertex AI and MLflow standardizing reproducible workflows and reducing manual release effort for production fraud systems.
- Implemented **explainability frameworks (SHAP, LIME)** for model decisions, supporting audit, compliance, and internal risk governance requirements.
- Designed **monitoring and drift detection pipelines** with cloud monitoring and MLflow tracking to track model performance and pipeline reliability, maintaining **>99% uptime** for real-time scoring infrastructure.
- Built internal APIs and dashboards using FastAPI and Tableau to expose risk metrics, anomaly trends, and model outputs to fraud operations teams for rapid investigation workflows.

Environment: Python, SQL, Airflow, PySpark, Big Query, Vertex AI, MLflow, Docker, Fast API, Tableau

Data Scientist

June 2020 - Jul 2023 | India

Toyota

- Architected **analytics pipelines** using **Apache Airflow, Dataflow, and BigQuery**, processing **8-12M operational records/day** for forecasting and KPI reporting
- Engineered **feature pipelines** with **Python, PySpark, and SQL**, generating **120+ time-series features** for operational decision models.
- Built **forecasting models** using **Prophet, LSTM, and XGBoost**, improving planning accuracy by **~15%** in scheduling workflows.
- Developed **NLP pipelines** with **Hugging Face Transformers (BERT)**, automating classification of **50K+ monthly feedback records** and reducing manual review by **~35%**.
- Implemented **anomaly detection systems** using **Isolation Forest and Autoencoders**, monitoring **20+ operational KPIs** and identifying early irregular patterns.
- Deployed **production ML services** with **Vertex AI, Docker, and MLflow**, ensuring consistent model performance across environments.
- Designed **executive dashboards** using **Tableau and Looker Studio**, translating model outputs into operational insights.

Environment: Python, SQL, PySpark, Airflow, TensorFlow, Scikit-learn, BigQuery, Vertex AI, Docker, MLflow, Tableau

TECHNICAL SKILLS

Programming Languages & Ecosystems: Python (Advanced), R, SQL, Pandas, NumPy, Scikit-learn, PyTorch

Machine Learning & Deep Learning: PyTorch, TensorFlow, Scikit-learn, XGBoost, LightGBM, Hyperparameter Tuning, Time-Series Modeling, Anomaly Detection, Feature Engineering

Generative AI & LLMs: Hugging Face Transformers, Lang Chain, LlamaIndex, RAG Pipelines, Prompt Engineering, LoRA / QLoRA, RLHF, GPT, LLaMA 2/3, BERT, T5, RoBERTa

NLP & Conversational AI: spaCy, NLTK, Named Entity Recognition (NER), Sentiment Analysis, Text Classification, Document Parsing, Tokenizer Optimization

MLOps & CI/CD: MLflow, Airflow, Docker, Kubernetes, Terraform, GitHub Actions, Jenkins, AWS CodePipeline, Experiment Tracking, Auto-scaling APIs, Model Monitoring

Real-Time Data & Pipelines: Apache Spark, PySpark, Kafka (Confluent), ETL Pipelines, Delta Lake, DBT, Feature Stores

Databases & Warehousing: MySQL, PostgreSQL, Oracle, MongoDB, Snowflake, Big Query, Redshift

Cloud Platforms & Infrastructure: AWS (SageMaker, Lambda, S3, ECS, Glue, CloudWatch), GCP (Vertex AI, TPU, AI Platform), Azure ML Studio

Data Quality, Monitoring & Governance: Great Expectations, Prometheus, Grafana, Logging, Alerting, Pipeline Performance Monitoring

Data Visualization & Analytics: Tableau, Power BI (On-Prem Gateway, KPI Dashboards), Matplotlib, Seaborn, Plotly, Business Intelligence Reporting, Error Analysis Visuals

Tools & Environments: VSCode, Jupyter Notebook, Git, GitLab, Containerization Environments

Mathematics & optimization: Linear Algebra, Multivariate Calculus, Optimization Algorithms, PCA, Advanced Analytics

EDUCATION

Master's in data science and business Analytics (Advanced Analytics)

Wayne State University

Bachelor's in computer science and civil engineering (Hybrid Program)

Lovely Professional University