**NAME:**

**Email:**

**Phone:**

**Data Scientist | AI & ML Engineer | Cloud MLOps Specialist**

## **PROFESSIONAL SUMMARY**

* Data Scientist and AI/ML Engineer with **10 years of experience** designing, developing, and deploying large-scale machine learning models using Python, TensorFlow, and PyTorch across enterprise-level data ecosystems on AWS, Azure, and GCP.
* Architected automated MLOps workflows using Kubeflow, MLflow, and Airflow with CI/CD orchestration through Jenkins and GitHub Actions.
* Built cloud-native data ingestion, preprocessing, and feature engineering pipelines using PySpark, Pandas, and Snowflake integrated with distributed computing frameworks.
* Implemented containerized model deployment with Docker and Kubernetes, ensuring high scalability and seamless version control across environments.
* Developed and fine-tuned generative AI and NLP models using LLM architectures integrated with LangChain and Hugging Face Transformers.
* Optimized ETL and ELT processes with Databricks and AWS Glue, enabling reliable data availability for real-time machine learning systems.
* Automated retraining, evaluation, and drift detection workflows with Prometheus, Grafana, and model monitoring frameworks.
* Integrated model explainability frameworks such as SHAP and LIME to ensure interpretability and compliance within enterprise pipelines.
* Deployed RESTful inference services through FastAPI and Flask for batch and streaming prediction endpoints using Kafka.
* Collaborated with data engineering and DevOps teams to implement infrastructure as code with Terraform and CloudFormation.
* Designed scalable data warehouses with Snowflake, BigQuery, and Redshift, supporting analytical and ML workloads.
* Implemented secure access and data encryption policies aligning with enterprise security and compliance frameworks.
* Led automation of feature stores for reusability across multiple machine learning pipelines using Feast and Snowflake.
* Built custom AutoML pipelines for model selection, cross-validation, and automated hyperparameter tuning.
* Mentored teams on data governance, versioning, and lifecycle management practices following modern MLOps standards.

## **Technical Skills**

| **Category** | **Skills** |
| --- | --- |
| **Programming** | Python, PySpark, SQL, Scala, R |
| **Cloud Platforms** | AWS (SageMaker, Glue, ECS, Redshift, S3), GCP (Vertex AI, BigQuery, GKE), Azure (Azure ML, AKS, Data Factory, Azure DevOps), Databricks |
| **ML/Deep Learning** | TensorFlow, PyTorch, Scikit-learn, Spark MLlib, Generative AI, LLMs, NLP, Hugging Face Transformers, LangChain |
| **MLOps & Orchestration** | MLflow, Kubeflow, Airflow, Jenkins, GitHub Actions, ArgoCD, Terraform, Docker, Kubernetes |
| **Data & Databases** | Snowflake (Data Warehouse, Feature Store), BigQuery, Redshift, Kafka, Feast, Pandas, ETL/ELT, Hadoop, Hive |
| **Deployment & Serving** | Docker, Kubernetes, FastAPI, Flask, RESTful APIs, AWS ECS, GCP GKE, Azure AKS |
| **Monitoring & Explainability** | Prometheus, Grafana, SHAP, LIME, Great Expectations, OpenMetadata |

## **PROFESSIONAL EXPERIENCE**

### **Client: J.P. Morgan Chase, Tampa, Florida Duration: Oct 2024 - Present**

**Role: Senior Data Scientist | AI/ML Engineer**  
**Environment:** Python, PyTorch, TensorFlow, AWS SageMaker, Airflow, MLflow, Snowflake, Terraform, Databricks, Docker, Kubernetes, Kafka, GitHub Actions

**Responsibilities:**

* Designed and deployed scalable machine learning pipelines in AWS SageMaker, integrating automated model retraining and evaluation with MLflow tracking.
* Built a centralized feature store on Snowflake with data versioning, improving model reproducibility and experimentation.
* Developed distributed data preprocessing workflows using Databricks and PySpark integrated with Airflow orchestration.
* Implemented real-time inference endpoints in FastAPI for high-throughput scoring workloads on AWS ECS.
* Integrated drift monitoring and alerting dashboards in Grafana for continuous model performance validation.
* Automated infrastructure provisioning with Terraform templates for production ML environments.
* Optimized data ingestion from multiple structured and semi-structured sources through AWS Glue and Lambda.
* Fine-tuned transformer-based language models using Hugging Face for text analytics and classification tasks.
* Created reusable ML components within Kubeflow pipelines to standardize experimentation across teams.
* Deployed models with CI/CD workflows in GitHub Actions integrated with MLflow model registry.
* Implemented lineage tracking and metadata management using OpenMetadata integrated into the pipeline.
* Developed custom evaluation metrics scripts to measure accuracy, drift, and consistency during retraining cycles.
* Collaborated with data engineers to enhance schema validation, version control, and ETL consistency across data pipelines.
* Designed a monitoring framework for tracking inference latency, data quality, and model health metrics.

### **Client:** EMC Insurance (Iowa) **Duration: Jul 2022 – Sep 2024**

**Role: Senior Data Scientist | ML Platform Engineer**  
**Environment:** GCP Vertex AI, BigQuery, TensorFlow, Scikit-learn, Kubeflow, Jenkins, ArgoCD, Cloud Storage, PySpark

**Responsibilities:**

* Developed production-grade ML pipelines using GCP Vertex AI and Kubeflow for training, evaluation, and batch prediction tasks.
* Engineered automated ETL pipelines in PySpark for feature generation and transformation with BigQuery integration.
* Deployed containerized ML models on GKE clusters with Jenkins-based CI/CD orchestration for continuous delivery.
* Implemented metadata tracking and version control through MLflow and GCP Model Registry for all active experiments.
* Built internal libraries for data validation and preprocessing ensuring consistent data structures across multiple projects.
* Integrated real-time model serving APIs using Flask and monitored latency via Stackdriver metrics.
* Collaborated with platform engineers to optimize GPU/TPU workloads on GCP for large model training sessions.
* Developed reusable templates for hyperparameter tuning using Vertex Vizier and automated job scheduling.
* Established automated rollback mechanisms and model promotion workflows through ArgoCD pipelines.
* Designed Spark streaming pipelines to capture incremental data updates for real-time predictions.
* Configured continuous model validation tests using Great Expectations for data quality enforcement.
* Built dashboards in BigQuery and Looker to visualize model metrics, feature importance, and drift patterns.
* Collaborated on cloud cost-optimization strategies by monitoring GPU usage and compute scaling thresholds.
* Standardized model deployment pipelines for reproducible and compliant ML lifecycle management.

### **Client: UnitedHealth Group, Minneapolis, MN Duration: Apr 2019 – Jun 2022**

**Role: Data Scientist | ML Engineer**  
**Environment:** Azure ML, SQL Server, Databricks, Power BI, TensorFlow, Flask, Azure Data Factory, PySpark

**Responsibilities:**

* Built large-scale machine learning models using Azure ML integrated with automated retraining pipelines and data versioning.
* Designed data ingestion and transformation pipelines with Azure Data Factory and Databricks for end-to-end automation.
* Developed PySpark workflows for feature extraction and aggregation, improving data reliability across multiple systems.
* Deployed deep learning models in Azure ML endpoints for scalable inference integrated with Flask microservices.
* Implemented model tracking, lineage, and artifact management using MLflow within Databricks.
* Created parameterized ETL templates for modular data pipelines supporting model experiments.
* Built Power BI dashboards to visualize prediction outcomes, data health, and performance insights.
* Automated CI/CD workflows with Azure DevOps pipelines for model integration and release.
* Created scheduled model retraining processes using Azure Functions triggered by new data arrivals.
* Implemented Python-based data validation scripts integrated into the data factory for schema integrity checks.
* Developed notebooks for distributed model evaluation leveraging Databricks runtime clusters.
* Built alerting systems integrated with Azure Monitor to notify on data anomalies and drift events.
* Collaborated with DevOps teams to deploy models using containerized Azure Kubernetes Service environments.
* Documented all modeling steps, dependencies, and retraining workflows for audit and reproducibility.

### **Client: Couth Infotech Pvt. Ltd, Hyderabad, India** **Duration: Oct 2017 – Jan 2019**

**Role: Machine Learning Engineer | Platform Developer**  
**Environment:** AWS EC2, S3, Redshift, Spark, Python, Scala, Jenkins, TensorFlow, Tableau, Docker

**Responsibilities:**

* Built distributed ML models on Spark MLlib and TensorFlow for large-scale structured and unstructured datasets.
* Designed AWS-based ETL architecture integrating S3, Lambda, and Step Functions for automated data processing.
* Developed and deployed containerized microservices using Docker on AWS ECS for model serving.
* Implemented real-time feature extraction and aggregation using Spark streaming integrated with Kafka.
* Automated CI/CD workflows with Jenkins for model packaging, testing, and deployment.
* Created schema validation scripts using PySpark ensuring consistent ingestion pipelines across regions.
* Collaborated with DevOps engineers to optimize Redshift queries for data storage and feature retrieval.
* Developed and validated ensemble learning models with cross-validation and feature selection scripts.
* Configured CloudWatch and Prometheus dashboards to monitor latency and resource consumption.
* Migrated legacy on-prem batch jobs into scalable AWS-based distributed pipelines.
* Designed custom Python modules for feature normalization and encoding consistency across projects.
* Integrated API endpoints into web dashboards providing live inference capabilities.
* Established automated version control for datasets, models, and artifacts via Git and DVC.
* Participated in agile sprints delivering high-quality, tested, and production-ready AI components.

### **Client: Brio Technologies Private Limited Hyd India** **Duration: May 2016 – Sep 2017**

**Role: Junior Data Scientist | Analyst Programmer**  
**Environment:** Python, R, SQL, Hadoop, Hive, Tableau, Scikit-learn, Git

**Responsibilities:**

* Built predictive models using Scikit-learn and R for structured data with end-to-end pipeline automation.
* Developed Hive queries and MapReduce jobs for distributed processing of large datasets.
* Created Python-based feature extraction scripts with integrated validation and reporting modules.
* Designed exploratory analysis workflows using pandas and visualization frameworks for data insights.
* Implemented version control workflows for datasets and scripts via Git repositories.
* Supported deployment of ML models through simple REST APIs for internal applications.
* Created reusable Python libraries for preprocessing, encoding, and transformation tasks.
* Automated daily ETL scripts for incremental data refreshes using Linux scheduling and shell scripting.
* Documented modeling processes, dependencies, and testing procedures in project repositories.
* Collaborated in agile development cycles contributing to code reviews and sprint planning.
* Built Tableau dashboards displaying aggregated model predictions and feature distributions.
* Optimized SQL queries for data aggregation improving overall processing efficiency.
* Participated in prototype implementation of ML workflows integrated with Hadoop ecosystem.
* Assisted senior developers in creating model validation frameworks and reporting scripts.