Harshavardhan Reddy

9292620040 — harsha.rgvr@gmail.com — LinkedIn

SUMMARY

Data Scientist with 4+ years of hands-on experience building, validating, and deploying production-grade machine learning systems across finance, healthcare, and pharmaceutical domains. Specialized in developing practical, scalable solutions for fraud detection, customer analytics, NLP automation, and computer vision workflows. Strong background in statistical modeling, A/B experimentation, and translating open-ended business problems into measurable, data-driven solutions.

TECHNICAL SKILLS

Programming & Data: Python, SQL, PySpark, R
Machine Learning: Regression, Classification, Clustering, Time Series Forecasting, XGBoost, LightGBM, Random Forest, Deep Learning, CNNs, Transformers
NLP: Hugging Face, BERT, SpaCy, NLTK, Text Classification, Named Entity Recognition (NER), Topic Modeling
Computer Vision: OpenCV, OCR, CNN Architectures, Image Classification, Anomaly Detection
MLOps & Production: MLflow, Docker, Kubernetes, CI/CD, Model Monitoring, Drift Detection, REST APIs
Big Data & Databases: Apache Spark, Databricks, Hadoop, PostgreSQL, MySQL, MongoDB
Cloud Platforms: Azure ML, AWS S3, SageMaker
Visualization & BI: Tableau, Power BI
Statistics & Experimentation: Hypothesis Testing, A/B Testing, Experimental Design, Statistical Modeling

EXPERIENCE

Capital One, Albany, NY **Apr 2025 – Present**
AI/ML Data Scientist

- Developed and deployed fraud detection and credit risk models using XGBoost and deep neural networks on large-scale transactional datasets, improving F1-score by 18% and reducing false positives by approximately 25%.
- Built NLP pipelines for customer feedback analysis and automated document classification using Hugging Face Transformers and SpaCy, reducing manual review time by 40%.
- Designed and implemented computer vision workflows for document verification and anomaly detection using CNN-based architectures and OpenCV, cutting manual intervention by 35%.
- Created robust end-to-end ML pipelines covering data ingestion, preprocessing, feature engineering, model training, evaluation, and deployment.
- Implemented experiment tracking, model versioning, CI/CD pipelines, and real-time monitoring using MLflow, Docker, and Kubernetes.
- **Tech Stack:** Python, PyTorch, TensorFlow, Spark, Azure ML, Databricks, Docker, Kubernetes, MLflow

Tata Consultancy Services (TCS) **Dec 2021 – Aug 2023**
Data Scientist

- Built churn prediction models, recommendation engines, and CLV forecasting solutions for enterprise clients, improving customer retention by 15%.
- Designed, maintained scalable ETL pipelines using PySpark / SQL for preprocessing and validating multi-terabyte datasets.
- Conducted feature engineering, statistical analysis, hyperparameter optimization to improve model generalization by 20%.
- Automated model training, evaluation, A/B testing, and deployment workflows using Apache Airflow and MLflow.
- Developed business-facing dashboards and analytical reports using Tableau and Power BI to support decision-making.
- **Tech Stack:** Python, PySpark, Scikit-learn, XGBoost, TensorFlow, Airflow, MLflow, AWS S3, Hadoop

Dr. Reddy's Laboratories **May 2020 – Nov 2021**
Data Scientist

- Developed predictive models for clinical outcome analysis using structured EHR data and unstructured clinical notes.
- Built NLP pipelines to automate literature reviews and extract insights from adverse event reports and physician notes.
- Performed exploratory data analysis, feature selection, and statistical validation to improve model transparency.
- Designed KPI dashboards for regulatory reporting and operational monitoring.
- Collaborated with clinical, regulatory, and data engineering teams to ensure data integrity and compliance.
- **Tech Stack:** Python, R, SQL, Scikit-learn, SpaCy, NLTK, Tableau, Power BI

EDUCATION

Pace University, New York City
Master's in Data Science