

Yashwanth Reddy

Data Scientist

✉ yashwanthreddy335@gmail.com ☎ 2018843230 📍 Jersey city, NJ 🌐 LinkedIn

PROFILE

Data Scientist with 4+ years of experience delivering scalable ML and analytics solutions across cloud and media domains. Skilled in predictive modeling, real-time inference, and experimentation to drive user engagement, retention, and operational efficiency. Proficient in building end-to-end data pipelines, deploying containerized ML services, and enabling decision-making through dashboards and KPI reporting. Strong collaborator with cross-functional teams, combining technical expertise with business acumen to translate complex data into actionable insights that improve customer experience and product performance.

SKILLS

- **Data Science & Machine Learning:** Random Forest, Logistic Regression, XGBoost, LightGBM, Anomaly Detection, Topic Modeling, Sentiment Analysis, A/B Testing, Uplift Modeling, Propensity Scoring, Concept Drift Detection, Performance Decay Analysis
- **Programming & Scripting:** Python (Pandas, NumPy, Scikit-learn, MLflow, Optuna), PySpark, SQL.
- **Data Engineering & Pipelines:** Apache Airflow, AWS Glue, PySpark, SQL.
- **Cloud & Infrastructure:** AWS (S3, Glue), Kubernetes, Docker, Jenkins, Flask, Grafana, Prometheus.
- **MLOps & Model Deployment:** MLflow, Databricks, Jenkins, Docker, Kubernetes, Flask, Log Data, Performance Metrics, Support Interactions.
- **Visualization & Business Intelligence:** Tableau, Power BI, KPI Tracking, Executive Dashboards
- **Collaboration & Leadership:** Cross-functional collaboration, Mentoring, Agile, Peer Reviews, Technical Documentation.

EXPERIENCE

VMware, *Web Resource Data Scientist*

07/2024 – Present | Remote, USA

- Partnered with product and engineering teams to identify and frame 10+ high-impact business problems across SaaS product telemetry and cloud infrastructure, leading to a 15% improvement in customer retention via targeted insights.
- Designed and automated data pipelines using PySpark and SQL to ingest and process ~3 TB of telemetry and log data daily from VMware's vSphere and NSX platforms, reducing data availability latency from 12 hours to under 2 hours.
- Built and deployed machine learning models (e.g., random forest, logistic regression, anomaly detection) to predict VM resource exhaustion and detect anomalous user behavior, increasing proactive alert accuracy by 28%.
- Engineered 100+ features from product usage logs, cloud performance metrics, and customer support data, and conducted hyperparameter tuning using MLflow and Optuna to improve F1-score by 22% across models.
- Delivered containerized ML inference services (Flask + Docker) deployed via Jenkins and Kubernetes, reducing model deployment cycles from 5 days to under 1 day and enabling real-time scoring for 50K+ daily API calls.
- Monitored deployed models for concept drift and performance decay using Grafana and Prometheus dashboards; retrained models monthly using scheduled Airflow workflows, maintaining <5% prediction error across quarters.
- Created interactive dashboards and reports in Tableau and Power BI to visualize model predictions and business KPIs, which were used weekly by senior leadership to guide roadmap and operational decisions.

Accenture, Sr. Data Scientist

10/2021 – 08/2023 | Hyderabad, India

- Led a cross-functional initiative to design and launch a recommendation engine for personalized content across HBO Max, resulting in a 19% increase in average user session duration and a 12% boost in monthly active users (MAU) within the first quarter of deployment.
- Architected and deployed a real-time churn prediction system using ensemble models (LightGBM, XGBoost) on a Spark-based pipeline, enabling targeted retention campaigns that reduced churn by 8.5% YoY in key demographic segments.
- Directed the end-to-end experimentation pipeline, including A/B testing frameworks and causal inference techniques (e.g., uplift modeling, propensity scoring), to evaluate content previews and marketing placements across digital platforms — increasing click-through rates by 2%.
- Managed large-scale data acquisition and enrichment pipelines using Airflow and AWS Glue to process over 5 TB of daily user interaction logs, integrating data from third-party ad platforms, streaming analytics, and CRM tools for unified audience profiling.
- Developed NLP-based models (topic modeling, sentiment analysis) on viewer feedback and closed-caption text to inform editorial decisions and improve trailer targeting, contributing to a 30% lift in trailer-to-watch conversion for new releases.
- Collaborated with product, engineering, and data governance teams to define data standards and deploy modular, reusable ML components using Databricks and MLflow, cutting model delivery timelines by 30% across business units.
- Mentored a team of 3 junior data scientists and analysts, conducting regular peer code reviews, technical deep-dives, and knowledge-sharing sessions to elevate team productivity and ensure reproducibility and scalability in deployed solutions.

EDUCATION

New Jersey Institute of Technology, Master's Degree

09/2023 – 05/2025 | Newark, USA

Data Science

Jawaharlal Nehru Technological University, Bachelor's Degree

08/2017 – 09/2021 | Hyderabad, India

Electrical and Electronics Engineering