Hemanth D

Data Scientist

dhemanthsai3@gmail.com +1 (857) 385-6963 Jersey City, USA LinkedIn

Summary

Curious and impact-driven Data Scientist with 4.5+ years of experience designing and deploying scalable ML solutions, real-time analytics systems, and cloud-native data pipelines. Proven track record in predictive modeling, time series forecasting, A/B testing, and dashboard development across industries including cloud infrastructure, higher education, and customer operations. Adept at translating business needs into robust data strategies using Python, PySpark, SQL, Tableau, and AWS. Collaborative and detail- oriented, with a strong foundation in both technical execution and stakeholder communication.Master's in Computer Science from UMass Boston.

Key Achievements

- Improved server capacity planning through predictive modeling, leading to approximately one point two million dollars in annual savings.
- Reduced incident response time by nearly forty percent with real-time triage alerting systems at MoneyGram.
- Developed and deployed a sub-second fraud detection system using Kafka and Spark Streaming for real-time anomaly alerts.
- Enhanced dashboard performance for over two hundred users at UMass Boston by optimizing SQL queries and data processing workflows.
- Automated machine learning model retraining using Airflow and SageMaker, cutting manual intervention by more than eighty percent.

Skills

Programming Languages: Python, SQL (PostgreSQL, Snowflake, BigQuery), R, Bash
Machine Learning Libraries: Pandas, NumPy, Seaborn, Plotly, Scikit-learn,XGBoost, statsmodels, TensorFlow, PyTorch, Ultralytics
Analytics & Visualization: Tableau, Power BI, Looker, Streamlit, Matplotlib, GeoPandas
Machine Learning Algorithms: Regression (Linear, Logistic, Ridge), Classification, Clustering, Random Forest, Feature Selection, Cross- Validation, Hyperparameter tuning
Statistical Methods & Experimentation: Time Series (ARIMA, Prophet), A/B Testing, Hypothesis Testing
ETL & Orchestration: PySpark, Apache Airflow, dbt, Kafka, Fivetran, Informatica, Talend
Cloud & MLOps: AWS (SageMaker, Lambda, Redshift, EC2), Azure (Synapse, Databricks, Data Factory), GCP (BigQuery, Dataflow), Docker, Git, Flask, Kubernetes
Data Tools & Methodologies: Alation, Collibra, Excel (Advanced Pivot, Power Query), JIRA, Agile/Scrum

Professional Experience

Data Scientist
Arthrex
03/2024 – present | Massachusetts, USA

- Applied ARIMA, SARIMA, Prophet, XGBoost, and LightGBM for time series forecasting of server demand, using Python, improving pipeline performance and demand prediction processes.
- Built automated ML pipelines using Apache Airflow, AWS SageMaker, and Git, integrating weekly model training, testing, and deployment cycles.
- Developed interactive, geospatial dashboards in Tableau using GeoPandas, visualizing server load and demand patterns for Clara AI and NRF initiatives.
- Performed segmentation and clustering analysis using Snowflake, Dask, and Pandas, identifying user behavior patterns for regional strategy alignment.
- Conducted A/B testing and hypothesis testing with Seaborn, Plotly, and statsmodels to evaluate feature changes and performance metrics.
- Collaborated with DevOps teams to implement model monitoring and retraining triggers using custom Python scripts and integrated them with AWS CloudWatch.
- Optimized data ingestion and processing using PySpark, standardizing schema parsing from Redshift logs to reduce data prep overhead.
- Worked in Agile teams using JIRA and followed Scrum methodologies, aligning data science KPIs with business goals via cross- functional collaboration.

Data Analyst
University of Massachusetts Boston
01/2022 – 12/2023 | Massachusetts, USA

- Designed and implemented PySpark pipelines that processed over 500GB of student engagement and performance data daily, reducing data load and transformation time by 35% and enabling near real-time analytics.

- Tuned complex PostgreSQL and MySQL queries, resulting in a 40% improvement in dashboard response time for academic performance reports viewed by 200+ stakeholders, including deans and department heads.
- Developed Power BI dashboards integrated with Azure Synapse to track enrollment trends, dropout rates, and funding distribution across 50+ programs, improving strategic academic planning.
- Implemented real-time anomaly detection in financial aid and student billing systems using Kafka and Spark Streaming, enabling alerts for suspicious transactions with sub-second latency.
- Oversaw AWS-based infrastructure (EC2, Lambda, Redshift) to ensure cost-effective, scalable, and secure compute for long-running analytics jobs and ad hoc query support.

Associate Data Analyst
TCS

08/2020 – 12/2021 |
Bengaluru, India

- Conducted detailed analysis of over 8 million customer support logs using Python and SQL, identifying root causes of delays and reducing average ticket resolution time from 3.2 to 2.4 days (25% improvement).
- Built Looker dashboards and dbt models to monitor SLA compliance and ticket escalation trends across 15+ support datasets, helping support leadership identify underperforming regions.
- Mapped and visualized 30,000+ geocoded tickets using GeoPandas, revealing technical issue clusters in high-volume corridors and guiding infrastructure enhancements in Asia-Pacific markets.
- Applied graph analysis with NetworkX to identify frequent sequences in 5K+ chatbot queries, leading to a 30% boost in intelligent routing accuracy and resolution rate.
- Created Kafka-driven AWS Lambda functions to automate alerts for operational triage, reducing response lag to high-severity events by 40%.

Projects

Forecasting Server Demand with ARIMA & XGBoost

- Built a hybrid time series forecasting model using ARIMA and XGBoost to predict server demand based on infrastructure usage logs from Redshift. Integrated the solution into a production-grade Airflow pipeline for weekly retraining and used AWS SageMaker for scalable deployment. The model enabled cloud operations teams to proactively manage server allocation, reducing overprovisioning by 10% and improving forecast accuracy by 18%. Visual dashboards were created in Tableau for real-time monitoring and reporting to leadership.

Real-Time Anomaly Detection in Financial Aid Systems

- Developed a real-time fraud detection system using Kafka, Spark Streaming, and PySpark to monitor financial aid transactions and flag suspicious behavior. The solution ingested live event streams, applied anomaly detection logic, and generated alerts with sub- second latency. Detected anomalies were stored in Azure Synapse and visualized through Power BI dashboards for compliance teams. This system significantly improved fraud detection speed and reduced manual investigation workload by 30%, strengthening institutional financial controls.

Education

M.S. in Computer Science
University of Massachusetts
Boston

01/2022 – 12/2023 | USA

B.S. in Computer Science Engineering
REVA University

05/2017 – 08/2021 | India

Certificates

Data Analysis with Python