SASHANK PRAYAGA

+4917660600047 ⋄ sashankprayaga93@gmail.com ⋄ Hilden, Germany

PROFESSIONAL SUMMARY

Results-driven Data Engineer with over 4 years of experience in building scalable data pipelines, cloud infrastructure, and machine learning solutions. Skilled in designing and automating end-to-end ETL workflows using Apache Airflow and AWS (Redshift, Lambda, RDS). Proven ability to collaborate cross-functionally with data scientists and process engineers to deliver impactful tools—such as anomaly detection models, recommendation systems, and interactive dashboards using Power BI and QuickSight. Adept at deploying ML models with FastAPI, Streamlit, and MLflow. Passionate about creating reusable infrastructure and enabling data-driven decision-making in production and business environments.

EXPERIENCE

Data Engineer Jun 2022 - Present
Yara International *Berlin, Germany*

- Designed and implemented an AWS data pipeline using Glue ETL, PySpark, Redshift Spectrum, and CI/CD practices, reducing data processing time by 30% for Data Lake and Data Modeling workflows.
- Managed and optimized cloud-based ML infrastructure with AWS CloudFormation, Amazon ECS, and CI/CD pipelines, reducing cloud costs by 20% while ensuring high system uptime and performance.
- Built interactive dashboards using Amazon QuickSight and Power BI, enabling stakeholders to derive actionable insights from cleansed, unified datasets.
- Ensured high data quality and pipeline reliability by implementing validation checks and performance optimizations, resulting in a 20% improvement in processing efficiency.
- Extracted production data from APIs and developed Python-based algorithms to identify optimal production conditions, enhancing throughput and product quality.
- Designed and implemented reusable AWS infrastructure templates (using CloudFormation) to support data science workflows, enabling faster and standardized project deployment.
- Collaborated with data scientists to build a shared cloud-based infrastructure, streamlining model development, testing, and deployment across multiple projects with improved scalability and consistency.

Intern - Data Scientist Jul 2020 - Jan 2021
Bosch Rexroth *Homburg (Saar), Germany*

- Developed and deployed machine learning models using sensor data to detect heat and mass flow anomalies with 95% accuracy, contributing to predictive maintenance efforts.
- Performed exploratory data analysis (EDA) with Plotly and Seaborn, uncovering key operational parameters that impacted system efficiency and performance.
- Improved model performance by 20% through hyperparameter tuning with Optuna, enhancing the reliability and precision of anomaly detection models.

Data Analyst - Last Mile Delivery Sep 2016 - Jun 2017
Amazon Development Centre *Hyderabad, India*

- Optimized last-mile logistics for Amazon North America, improving delivery efficiency by 15%.
- Played a key role in testing a new delivery solution that achieved a 98% success rate.

EDUCATION

M.Sc. in Commercial Vehicle Technology,TU Kaiserslautern — Apr 2022
Kaiserslautern, Germany

B.Tech in Mechanical Engineering, ICFAI Foundation for Higher Education — Jul 2016
Hyderabad, India

PROJECTS

Master Thesis – Defect Detection in Automated Rail-Guided Vehicles Using Machine Learning

- Collected and preprocessed sensor data from unmanned rail-guided vehicles operating in warehouse environments, focusing on detecting 8 distinct mechanical defects.
- Augmented limited defect data by replicating fault conditions, improving model robustness and training data balance.
- Engineered features using sliding windows, TSFEL, and Featuretools, and trained multiple ML algorithms to identify defect patterns, achieving 90% accuracy with the best-performing model.
- Deployed the final model using FastAPI and Streamlit, and implemented MLflow to track experiments, version models, and monitor performance in real-time.

Intelligent Drive Mode Selection Using ML in Mild Hybrid Powertrains

- Collaborated in a team of two to develop a machine learning model for optimizing the controller in a mild hybrid vehicle, intelligently switching between battery-only, engine-only, and hybrid modes based on parameters like speed, RPM, and engine load.
- Designed the solution to maximize fuel efficiency by predicting optimal drive modes using historical driving data and vehicle telemetry.

SKILLS

Programming	Python, SQL, R, C++
Data Engineering	Apache Airflow, Apache Spark, Apache Kafka, ETL, ELT
Cloud	AWS (Redshift, RDS, Lambda, S3, EC2, CloudFormation)
Machine Learning & Deployment	Supervised and Unsupervised Learning, Anomaly Detection, Model Deployment with FastAPI, Streamlit, MLflow, CI/CD with Git
Databases	Datawarehouse(Amazon Redshift, Snowflake), NoSQL(DynamoDB) Relational Databases(Postgres, Amazon RDS), Datalake(Amazon S3)
Visualization	Power BI, Amazon QuickSight, Plotly, Seaborn
Infrastructure	Infrastructure as Code (CloudFormation), Reusable AWS Templates, Containerization (Docker), CI/CD Pipelines
Tools	Jira, Confluence, Microsoft Office Suite (Excel, PowerPoint, Word)
Soft Skills	Cross-functional Collaboration, Problem Solving, Communication, Attention to Detail

PROFESSIONAL DEVELOPMENT

AWS Certified AI Practitioner, Amazon Web Services (AWS)

Apache Airflow 3 Fundamentals, Astronomer

dbt Fundamentals, dbt learn

Google Data Analytics - Professional Certificate, Google (Coursera)

Microsoft Azure Data Fundamentals, Microsoft (Coursera)