DHARMIK DAVE

dharmikdave2902@gmail.com • +49 15238542236 • Koblenz, Germany • LinkedIn • GitHub

PROFESSIONAL SUMMARY

Data Analyst with hands-on experience in Python, SQL, Power BI, and cloud data platforms (Databricks, Azure, Snowflake). Delivered a 35% reduction in enterprise reporting time by optimizing Spark SQL pipelines and built executive KPI dashboards for leadership decision-making. Completing an MSc at the University of Koblenz focused on data engineering and machine learning, including time-series forecasting of CO_2 emissions. Available for Data Analyst / BI roles across Germany and the EU.

PROFESSIONAL EXPERIENCE

Data Analyst Metal Cast Engg. Inc. • Ahmedabad, India • Jan 2021 – Dec 2021

- - Designed and implemented Power BI dashboards (data visualization) monitoring KPIs (impressions, CTR, CPC, conversion metrics), delivering actionable insights and maintaining continuous stakeholder management with leadership for business intelligence and campaign reporting.
- Enhanced query performance in Databricks (Spark SQL), achieving a 35% reduction in reporting time across enterprise-scale datasets.
- Developed and optimized SQL queries to extract, transform, and join data from multiple sources, improving analytics pipeline reliability.
- Performed end-to-end data cleaning, preprocessing, and transformation, ensuring accuracy and trustworthiness across all reporting outputs.

EDUCATION

MSc, E-Government (Data Science & AI focus) University of Koblenz

- Koblenz, Germany Present

Focus: cloud-native data pipelines (Azure, Snowflake, dbt), machine learning, NLP, data governance.

B.E. Mechanical Engineering G.H. Patel College of Engineering & Technology

- India 2021

CERTIFICATIONS

Programming for Everybody (Getting Started with Python) University of Michigan / Coursera

Introduction to Artificial Intelligence (AI), with Honors IBM

Power BI Data Analyst Microsoft / Coursera

Databricks Fundamentals Databricks Academy

CORE SKILLS

Python (Pandas, NumPy), SQL, PostgreSQL, Microsoft Excel, Statistical Analysis, Data Cleaning, Power BI, KPI Reporting, Dashboarding, Data Storytelling, Matplotlib, Seaborn, Databricks (Spark SQL), Azure, dbt, ETL, Data Pipelines, Scikit-learn, Time-Series Forecasting (ARIMA, Random Forest, Gradient Boosting), NLP (basic), Git, Jupyter Notebooks, Agile Collaboration, Stakeholder Communication, Pytorch , Tensorflow

Data Analysis: Python (Pandas, NumPy), SQL, PostgreSQL, Microsoft Excel, Statistical Analysis, Data Cleaning
BI & Visualization: Power BI, KPI Reporting, Dashboarding, Data Storytelling, Matplotlib, Seaborn, Grafana
Data Engineering & Cloud: Databricks (Spark SQL), Azure, Snowflake, dbt, ETL, Data Pipelines
Machine Learning: Scikit-learn, Time-Series Forecasting (ARIMA, Random Forest, Gradient Boosting), NLP (basic)
Tools & Practices: Git, Jupyter Notebooks, Agile Collaboration, Stakeholder Communication

PROJECTS

CO_2 Emission Forecasting for Climate Policy

- Built an end-to-end forecasting pipeline replicating a 2025 peer-reviewed study: benchmarked ARIMA against Random Forest and Gradient Boosting on daily CO_2 emissions (Carbon Monitor) across 4 world regions.
- Engineered lag, rolling, and calendar features and benchmarked ARIMA against Random Forest, Gradient Boosting, and XGBoost on a chronological split; Random Forest achieved $R^2 \approx 0.90$, replicating a 2025 peer-reviewed benchmark, with SHAP for interpretability.
- Delivered an interactive Power BI dashboard and a deployed Streamlit web app; full stack: Python, SQL Server, scikit-learn, statsmodels, Power BI, Streamlit, Docker, and GitHub Actions (CI).

Marketing Analytics Platform

- Investigated declining customer engagement and conversion rates despite increased marketing spend; built analytical reporting to identify root causes and guide budget reallocation.

Incident Similarity Detection Using NLP (In Progress)

- **Building an NLP pipeline to detect duplicate IT incident tickets**, cleaning and normalising free-text descriptions (tokenisation, stop-word removal, lemmatisation) before vectorising them for similarity comparison.
- **Implementing and benchmarking similarity approaches** — TF-IDF with cosine similarity as a baseline against transformer-based sentence embeddings — to measure which surfaces near-duplicate tickets most reliably.
- **Designing the output as a triage aid**, returning a ranked list of the most similar past incidents with their resolutions, so recurring issues can be closed faster instead of re-investigated.

LANGUAGES

English (C1, professional), German (B1, limited working), Greek (A2)