Janmajay Kumar

Data Analyst/Engineer

Mühlstrasse 20, Tübingen 72074, Germany •janmajay@iitdalumni.com•+ 4917677680078
GitHub- https://github.com/QED137
LinkedIn - www.linkedin.com/in/janmajaykumar

SUMMARY

Junior Data Analyst/Engineer skilled in ETL pipeline development, KPI analysis, and data integration to optimize production efficiency and decision-making. Experienced in building dashboards, tracking key performance indicators (KPIs), and automating data workflows to improve operational insights and business intelligence. Proficient in Python, SQL, machine learning, and cloud platforms, with a passion for using data to drive production optimization and process automation.

SKILLS

- **KPI Analysis & Dashboarding:** Tableau, Plotly, Streamlit, Matplotlib, Seaborn.
- **Data Engineering & ETL Pipelines:** Google Cloud Functions, BigQuery, SQL, APIs, Pandas, NEO4J
- **Production Data Integration & Reporting:** Automating data workflows, visualizing operational data, optimizing production performance, MLFlow
- **Machine Learning & AI Applications:** RAG pipelines, LLMs, OpenAI APIs, TensorFlow, PyTorch, scikit-learn
- **Programming & Development:** Python, SQL, C++, JavaScript, FastAPI, Docker
- **Languages :** English (Fluent), German (B1-pursuing)

PROFESSIONAL EXPERIENCE

WBS Coding School – Data Science Trainee, Berlin, Germany | 08/2024 - 12/2024 *WBS Coding School is a leading international training center, specializing in hands-on, industry-relevant training in data science, machine learning, and cloud technologies.*

- **Database Engineering: Google Cloud ETL Pipeline**
  - Designed and deployed an **ETL pipeline** using **Google Cloud Functions**, **BigQuery**, and **Cloud Scheduler** to automate data ingestion, transformation, and storage.
  - Processed weather and flight data from **OpenWeather API** and **AeroDataBox API**, using **Pandas** to clean and store results in **MySQL** and **BigQuery**
  - Streamlined database workflows, ensuring efficient integration between APIs and database systems for analytics and reporting.

- **Graph-Powered Retrieval-Augmented Generation (RAG)**,
  - Developed a recommendation system combining **Neo4j graph databases** and **LLMs**, delivering intelligent, context-aware movie recommendations.
  - Utilized **semantic embeddings** for similarity matching and retrieval-augmented generation to improve recommendation accuracy.
  - Built an interactive **Streamlit web application** to enhance user engagement and experience.

- **Predictive Machine Learning for House Price Prediction**

  - Implemented machine learning models such as **Gradient Boosting**, **Random Forest**, and **XGBoost**, achieving an **R^2 score of 0.93** and reducing **RMSE to 18,710** on test data.
  - Applied **GridSearchCV** for hyperparameter tuning and developed feature selection pipelines to optimize model performance.
  - Tools and Libraries: **scikit-learn**, **Pandas**, **NumPy**, **Matplotlib**, and **Seaborn**.

University of Tübingen – Research Assistant, Tübingen, Germany | 02/2022 - 06/2023
Worked as part of the LEGEND 1000 international collaboration, a leading experiment dedicated to studying neutrinoless double beta decay to uncover the physical phenomena of the early universe that led to the disappearance of antimatter.

- **Monte Carlo Simulation:** Developed and implemented simulations using **GEANT4 (C++ framework)** to compare neutron capture on water and gadolinium, improving experimental design and efficiency.
- **Data Analysis with ROOT and Python**: Conducted statistical analysis on simulated data using **ROOT** and **Python**, concluding that neutron capture efficiency increased by 90% with 0.2% gadolinium.

Alle-Dinge, Data Integration and Backend Support, Tübingen, Germany, 2/2023 - Present
Alle-Dinge is a startup based in Tübingen focused on developing a multipurpose app for organizing personal data, managing digital containers, and enhancing user productivity.

- Developed a backend PDF reader feature that scans and organizes PDFs within the app, transforming them into structured HTML pages for easier accessibility and interaction.
- Implemented machine learning methods to enhance OCR (Optical Character Recognition) accuracy, improving the conversion of scanned PDFs to HTML format.

EDUCATION

WBS Coding School Berlin, Germany 08/2024- 12/2024 University of Tübingen, Germany,
Master in Astro and Particle Physics, 10/2018 - 10/2021

REFERENCES

- **Prof. Dr. Josef Jochum**
  Eberhard Karls University Tübingen, Germany
  josef.jochum@uni-tuebingen.de

Tübingen, 06.02.2025