

Janmajay Kumar

Data Scientist/Engineer

Mühlstrasse 20, Tübingen 72074, Germany • janmajay@iitdalumni.com • + 4917677680078

[GitHub- https://github.com/QED137](https://github.com/QED137)

[LinkedIn - www.linkedin.com/in/janmajaykumar](https://www.linkedin.com/in/janmajaykumar)

[MyWebPage-www.janmajay.de](http://www.janmajay.de)

SUMMARY

Data Scientist & Data Engineer with a background in physics, specializing in data validation, ETL pipelines on cloud platforms (Google Cloud, with transferable skills for Azure/AWS), and business intelligence. Skilled in Python, SQL, Power BI, and cloud-based data processing (BigQuery, Cloud Functions). Experienced in KPI tracking, trend analysis, and process automation to improve data accuracy and decision-making. Passionate about scalable data architecture, data modeling, and visualization to drive business insights.

SKILLS

- **Machine Learning & AI Applications:** RAG pipelines, LLMs, OpenAI APIs, TensorFlow, PyTorch, scikit-learn.
- **Financial & Business Data Analytics:** KPI tracking, trend analysis, statistical modeling.
- **Data Engineering & ETL Pipelines:** Azure, ETL/ELT on GCP, BigQuery, SQL, APIs, Pandas, NEO4J.
- **Production Data Integration & Reporting:** Automating data workflows, MLOPs, CI/CD
- **Programming & Development:** Python, SQL, MySQL, C++, JavaScript, Docker, bash.
- **Visualization:** Tableau, PowerBI.
- **Languages :** English (Business Fluent), German (B2-actively learning, aiming at C1).

EDUCATION

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.*

- **Numerical Optimization & Predictive Modeling:** Implemented machine learning models (Gradient Boosting, Random Forest, XGBoost) to predict housing prices, achieving an R^2 score of 0.93. Applied GridSearchCV for hyperparameter tuning and model optimization.
- **End-to-End Pipeline on Google Cloud:** Designed an ETL pipeline on Google Cloud (BigQuery, Cloud Functions, Cloud Scheduler) to automate data ingestion and transformation for large-scale analytics.
- **Graph-Powered Retrieval-Augmented Generation (RAG):** Developed a recommendation system combining Neo4j graph databases and LLMs, delivering intelligent, context-aware movie recommendations.

Master's in Astro and Particle Physics (10/2018 - 10/2021)

University of Tübingen, Germany

- **Computational Astrophysics & Numerical Simulations:** Developed Python-based models to analyze complex astrophysical systems, applying advanced numerical techniques.
- **Simulating Hawking Radiation:** Developed a Python-based model to study black hole evaporation dynamics.
- **N-Body Simulation:** Implemented a numerical solver for multi-body gravitational interactions using Python & NumPy.
- **Kepler's Equation & Celestial Mechanics:** Solved Kepler's equation using Fixed-Point Iteration & Newton-Raphson methods for orbital mechanics.
- **Lane-Emden Equation Solver:** Modeled stellar structures & polytropic stars, applying Dahlquist's stability test for numerical solvers.

PROFESSIONAL EXPERIENCE

Data Integration and Backend Support at Alle-Dinge, 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 processing system, enabling automated scanning, organization, and transformation of PDFs into structured HTML pages, enhancing document accessibility and user interaction.
- Implemented machine learning algorithms to optimize OCR (Optical Character Recognition) accuracy, significantly improving text extraction and conversion of scanned PDFs to interactive digital formats.
- Optimized backend performance by integrating efficient data structures and processing pipelines, reducing document retrieval time and enhancing system scalability.

University of Tübingen – Research Assistant, Tübingen, Germany | 04/2022 - 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 Simulations & Mathematical Modeling:** Developed and implemented GEANT4-based Monte Carlo simulations (C++) to analyze neutron capture on water and gadolinium, improving experimental design and efficiency.
 - **Computational Data Analysis:** Conducted large-scale statistical analysis using ROOT (C++/Python) and optimized simulation parameters, concluding a 90% increase in neutron capture efficiency with 0.2% gadolinium.
 - **High-Performance Scientific Computing:** Performed Simulation on computing cluster, i.e MPIK cluster Munich.
-

TRAINING AND CERTIFICATE

- WBS Coding School Berlin, Germany 08/2024- 12/2024
- Google Cloud Database Engineer Specialization 06/2024
- Deep Learning Specialization, Stanford Online 10/2023
- Machine Learning Specialization, Stanford Online 08/2023

CONFERENCE

- Muon Veto of the LEGEND Experiment, University of Vienna, August 2023
<https://indico.cern.ch/event/1199289/contributions/5445863/>

REFERENCES

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

Tübingen, 05.03.2025

