Janmajay Kumar

Data Scientist/Engineer

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

GitHub- https://github.com/QED137

LinkedIn - www.linkedin.com/in/janmajaykumar

MyWebPage-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

Janmajay Kumar

Data Scientist/Engineer

Steffen Göckel
PTV Planung Transport Verkehr GmbH
Haid-und-Neu-Str. 15
76131 Karlsruhe, Germany
Subject: Data Platform Engineer (m/f/d)

Tübingen, 05.03.2025

Dear Mr. Göckel,

I am reaching out to apply for the Data Platform Engineering position at PTV. I first discovered PTV's expertise in traffic planning, simulation, and real-time traffic management during the Karlsruhe job fair, and I was particularly impressed by your work with cloud-based simulation. With a background in C++ simulation, cloud engineering, and data science, I believe my skills align well with the requirements of this role.

During my time at the University of Tübingen, I developed Monte Carlo simulations in GEANT4 (C++) as part of the LEGEND 1000 collaboration, demonstrating that incorporating 0.2% gadolinium enhanced neutron capture efficiency by 90%. This project required running large-scale computations on the Max Planck Munich's computing cluster, further strengthening my expertise in high-performance computing (HPC), large-scale simulations, and complex data modeling. Additionally, it sharpened my C++ proficiency and my ability to work with computationally intensive simulations—skills that are directly applicable to PTV's traffic simulation and data-driven modeling software.

In addition to my academic research, I have acquired practical experience in data science, machine learning, and cloud engineering through the Data Science program at WBS Coding School. My projects include:

- Cloud-Based ETL Pipelines: Developed scalable workflows on Google Cloud (BigQuery, Cloud Functions, SQL), which reduced data processing time by 40% and facilitated real-time analytics.
- Predictive ML Models: Created Gradient Boosting, Random Forest, and XGBoost models using TensorFlow and PyTorch, resulting in a 20% improvement in prediction accuracy for real-world datasets.
- Graph-Based RAG Systems: Integrated Neo4j and LLMs, enhancing recommendation accuracy by 25% through graph-based semantic embeddings.

These experiences have bolstered my capability to develop high-performance AI solutions and cloud-based data platforms, aligning well with PTV's emphasis on scalable data processing, cloud integration, and AI-driven analytics.

Currently, I am employed at Alle-Dinge, a software startup in Tübingen, where we are creating a multipurpose app to organize various data types for both personal and commercial use. My role primarily involves backend development and OCR-powered document automation, contributing to data engineering, integration, workflow automation, and PDF processing. I also utilize Docker to containerize applications, ensuring scalable deployment and enhanced document accessibility.

I am a quick learner with a passion for simulation and data analysis, adept at uncovering patterns from both real-world and simulated data. My experience in software development and data science, combined with my expertise in cloud computing, positions me to contribute effectively to your projects right away.

Thank you for your time and consideration.

Best regards,

Janmajay Kumar