



Akash Chandra Baidya

Email: akashbaidya2@gmail.com

LinkedIn: linkedin.com/in/akashbaidya

GitHub: github.com/AkashBaidya

Mobile: +49 157 58474346

Website: akashbaidya.github.io

Address: Emil-Figge-Straße 25, 44227 Dortmund, Germany

OBJECTIVE

Data Scientist with 3+ years experience in NLP, Gen-AI and machine learning. Specialized in fine-tuning LLMs and anomaly detection for high-dimensional data. Delivered 10+ projects in data science, bioinformatics and research applications, backed by 5 years of software engineering expertise.

EXPERIENCE

- **RWE Supply & Trading GmbH** Essen, Germany
Portfolio and Quantitative Analyst, Part time Jan 2025 - Present
 - Implemented quantitative models and performed portfolio analysis to support trading and risk management decisions.
 - Engineered and optimized data workflows using Databricks, Dremio, SQL, and Python, ensuring seamless data integration and analysis.
- **IMSP - Institut für Marktforschung, Statistik und Prognose** Munich, Germany
Data Scientist, Part time June 2024 - Jan 2025
 - Delivered insights for energy companies with applying different data science models.
 - Created training materials and analytical solutions with various APIs, Power BI and Python libraries and delivered training sessions on data analytics, data engineering, and data science to clients, including Green Bootcamps, enhancing the skillset of over 50 participants.
- **Forschungszentrum Borstel, Leibniz Lungenzentrum** Borstel, Germany
Research Assistant (WHF) on Data Visualization, Part time Sep. 2023 - May 2024
 - Contributed to 2 projects resulting in 2 potential journals on proteomics data in exploratory data analysis:
 - * Analyzed protein cysteine oxidation data to provide insights and visualizations for research purposes.
 - * Visualized "Batch Corrected Proteomics" samples to facilitate analysis and interpretation of experimental results.
- **E.ON RWTH Energy Research Center-Aachen** Aachen, Germany
Research Assistant (WHF), Part time Sep. 2021 - April 2022
 - Developed web applications using Python and Django, focusing on Docker and Ansible for deployment, and built simulation models with probability frameworks to enhance the accuracy and reliability of energy system evaluations.
- **ITM System LLC.** Arizona, USA
Assistant Project Manager, Remote, Full time June 2018 - August 2021
 - Led multiple Django-based project developments, including an Appointment System featuring patient alerts, reminders, and confirmation via voice and text messages, serving different healthcare providers in the USA, Philippines, and Vietnam.
- **SINGER Bangladesh Ltd.** Dhaka, Bangladesh
Software Engineer - Reporting, Full time Nov. 2018 - March 2021
 - Utilized SQL for generating crucial business insights and crafted informative dashboards and reports using Jasper BI for the accounting and marketing department.
 - Engineered a Django-based Rent Management System tailored for precise rent calculation reports and ensured compliance with IFRS 16 standards across organizations.
- **Ubitrix BD Ltd.** Dhaka, Bangladesh
Consultant Software Engineer, Full time Feb. 2017 - May 2018
 - Developed multiple projects utilizing **Python, Android, and PHP**, such as **Speech Therapy Online, Ag Palliative Care, and Natural Family Planning.**

EDUCATION

- **Technische Universität Dortmund** Dortmund, Germany
Master of Science in Data Science (Nov. 2020 - Dec. 2025) — GPA: 2.4/4.00
 - **Courses Completed** Advanced Statistical Learning, Statistical Theory, Information Mining, Deep Learning, Big Data Analytics, Bioinformatics, Machine Learning Paradigms for Complex Data & Industrial Data Science
- **Khulna University of Engineering & Technology** Khulna, Bangladesh
Bachelor of Science in Computer Science & Engineering (Feb. 2012 - July 2017) — GPA: 2.48/4.00
 - **Thesis** Distributed MAC protocols for VANET (Criticalities for RR-Aloha, MS-Aloha, RR-Aloha+ and proposed a new distributed Aloha MAC protocol for VANET)

PROJECTS

- **MALDIX NLP and LLM Fine-tuning Pipeline**(*Python, spaCy, NLTK, PyMongo, IMSP*) Built NLP tools for preprocessing and analysis using Mistral AI, spaCy/NLTK, storing data via PyMongo. Fine-tuned LLMs with maldix.json to improve conversation modeling.
- **Time Series Processing, Anomaly Detection, and Explanation**(*Pytorch, Scikit, Pandas, Internship*) Developed machine learning models, including LSTM, to detect anomalies in energy consumption data from over ten TU Dortmund facilities and these models improved anomaly detection accuracy, optimizing energy management and operational efficiency.
- **AutoML Cashflow Optimization for Insurance Company**(*Python, Automated Machine Learning (AutoML), Power BI, side project.*) Implemented an AutoML solution to optimize cash flow predictions for an insurance company, utilizing Power BI to create dashboards that visualize model performance across multiple metrics, enabling comprehensive insights and informed decision-making.
- **Statistical Analysis of Munich Rent Data**(*R, TU Dortmund*) Conducted non-parametric and parametric tests (Kruskal-Wallis, Wilcoxon, ANOVA) to analyze rental price variations across discrete covariates. Verified assumptions with visualizations (boxplots, QQ-plots), applied Bonferroni correction, and interpreted significant group differences to evaluate spatial and categorical impacts on rent.
- **Credit Card Fraud Detection**(*Python, Power BI, side project.*) Built an end-to-end credit card fraud detection system involving exploratory data analysis (EDA), visualization with Plotly, and dashboards using Power BI. Applied cross-validation and undersampling (SMOTE) techniques to address data imbalance, utilizing Logistic Regression, Random Forrest and Gradient Boosting models to achieve higher accuracy.

SKILLS

- **Programming Languages:** Python (NumPy, Pandas, Scikit-learn, Keras, PyTorch, Pyspark, FastAPI, Requests, Django), R (dplyr, ggplot2), SQL
- **Visualization:** Power BI, Tableau, Plotly, Dash, RShiny, Matplotlib, Seaborn
- **Data Science & Analysis:** Advanced Excel, Statistical Analysis (SciPy, StatsModels, SAS), time series, LLM fine-tuning
- **Databases & Big Data:** SQL (MySQL, PostgreSQL, Oracle), NoSQL (MongoDB), ETL (Apache Airflow), MLflow, Snowflake, Databricks
- **Cloud & DevOps:** Git, Docker, Azure, AWS, Agile, IaaS, SaaS, CI/CD
- **Languages:** Native: Bengali. Fluent: English (C1). Beginner: German (A2).

CERTIFICATIONS

- **Azure Fundamentals**, *Datacamp, 2024*
- **Data Visualization with Power BI**, *Datacamp, 2024*
- **Deep Learning with pytorch**, *Datacamp, 2023*
- **Cluster Analysis in Data Mining**, *Coursera, 2021, Issued by University of Illinois at Urbana-Champaign*
- **Applied Scrum for Project Management**, *Coursera, 2019, Issued by University of Maryland, College Park*