Hassan Abdullah Ghauri

>> Data Scientist | ML Engineer

Contact

Phone: +4917646679796

E-Mail: hassanghauri3924@gmail.com

Location: Hamburg, Germany

Github: https://github.com/Hassan3924

LinkedIn: www.linkedin.com/in/hassan-abdullah-ghauri-ba3286125

Education

Master of Business Intelligence and Data Science

International School of Management **(Germany, Hamburg)**

2024 - 2026

Bachelor of Computer Science (Big Data)

University of Wollongong in Dubai (**UAE, Dubai)**

2020 - 2023

Associate in General Studies

Wilbur Wright College **(USA, Chicago)**

2016 - 2019

Skills

Programming: Pandas, NumPy, Matplotlib, Seaborn, **Scikit-Learn, SQL, LangChain**

Machine Learning: Linear Regression, Logic Regression, **Decision Trees, Random Forest,** PCA, Clustering, **Exponential Smoothing**

Other: Github, Data Visualisation (Tableau, Power BI, R), Selenium, Web scraping, **IBM SPSS, Time Series**

Language

- English **(C1, IELTS Score 7.5)**
- Urdu **(Native)**
- German **(A1)**

About Me

I am a dedicated Data Scientist passionate about **leveraging AI and analytics** to solve **complex challenges** and drive strategic innovation. With expertise in statistical analysis, **machine learning**, and **data visualization**, I have practical experience applying data-driven insights to real-world business problems. Proficient in Python, SQL, and AI frameworks, I am skilled in the full data science lifecycle. My **fluency in English (C1) and basic German (A1)** enable me to collaborate effectively in diverse, international teams. I am eager to contribute to impactful projects at the intersection of AI and business strategy.

Current Courses

The Data Science Course: Complete Data Science Bootcamp 2025 — 2025 - Present

- **Statistical Analysis,** NumPy, Pandas, Matplotlib, Seaborn, Linear Regression, **Logistic Regression, Clustering Algorithms**, Factor Analysis

Master deep learning in PyTorch — 2025 - Present

- Deep Learning, **Machine Learning, Neural Networks,** PyTorch, Python, Model Optimization, Regularization Techniques, **Hyperparameter Tuning,** Gradient Descent and Backpropagation, **Linear Algebra**

Experience

Data Science Intern - RANEHub — 2023 - 2024

UAE, Dubai

- Spearheaded a project in **Generative AI,** focusing on the fine-tuning of a Large **Language Model (LLM) specifically trained** on an extensive corpus of medical text data. Helping medical student's lessen their workload with an OpenAI powered medical chatbot.
- Utilized **Gradio,** an interactive **machine learning interface**, to conduct rigorous testing and validation of the model, ensuring its reliability and user-friendliness for medical professionals and stakeholders.
- **Web scraping data** about companies from LinkedIn and performed pre-processing on the data to make it ready for analysis.

Research & Development Intern - The Assembly — 2022 - 2022

UAE, Dubai

- Conducted **data analysis with Excel** and **applied Machine Learning techniques** using Miniconda, enhancing model training and data integrity with Scikit Learn.
- Developed a language translation tool using **Python and API data extraction**, and built a desktop reminder application, showcasing software engineering skills across multiple languages and frameworks like Flutter.
- **Showcased preprocessing techniques** on how to prepare non numerical data ready for training and testing.

Projects

Customer No-Show Classification

- Developed a machine learning model **using AdaBoost** to predict customer no-shows for Hapag-Lloyd, achieving 83% recall and 93% accuracy, **optimizing vessel utilization**.
- Analyzed feature importance and applied **Kendall's tau correlation** for feature selection, enhancing model robustness with key predictors..
- Created visualizations like **ROC curve, learning curve** to communicate model performance, supporting data-driven decision-making **for stakeholders.**
- **Collaborated with Emanuel and Laura** to leverage data understanding expertise, refining dataset insights to improve model accuracy and align with **Hapag-Lloyd's operational needs.**

Body Measurement

- Designed and implemented a mobile app using **React Native** for real-time body measurement capture, utilizing the Expo Camera module for image acquisition.
- Developed and **deployed machine learning models in GCP** using TensorFlow in Python, focusing on accurate body measurement extraction from 2D images.
- Measurements with an **accuracy of over 90%.**

Customer Churn

- Built a **Customer Churn Classifier model** for a Telecommunication industry **using Logistic Regression.**
- Conducted **exploratory data analysis,** and data visualization.
- Performed **hyperparameter tuning** for model optimization.
- Evaluated model performance using metrics such as **confusion matrix, classification, f-test,** etc.
- Improved model efficacy through **feature engineering.**
- Conducted **feature importance analysis.**

Germany Air Passengers Forecast

- Analyzed **real-world passenger** transport data publicly available from Eurostat, focusing on inbound and outbound **passenger traffic in Germany (2002-2025),** to identify trends and seasonality.
- **Developed and validated** an accurate **ARIMA/SARIMAX forecasting model predicting** passenger traffic through 2027, indicating a **strong recovery of Germany's tourism to pre-COVID levels by late 2026.**
- Applied **Exponential Smoothing** techniques like EWMA with varying alpha values to accurately capture patterns in **monthly passenger data.**
- Conducted statistical evaluation using **Ljung-Box Test,** to ensure robust model performance.

Azure Data Lake

- Enabled centralized data storage from multiple sources in **Azure Data Lake.**
- Leveraged the data stored in the **Data Lake for advanced analytics.**
- Enhanced **data security** measures as part of the Data Lake architecture.
- Facilitated **business insights** using additional technologies like Power BI and PySpark.
- Contributed to better **business decision-making** through integrated data analysis.

Titanic Survival

- Developed a **Logistic Regression model** to predict if a person would survive or not
- Made sure to **apply data cleaning** as well as exploratory data analysis
- Tested and **evaluated model performance** using metrics such as confusion matrix, classification, f-test, etc.
- Conducted **statistics techniques** to improve the model even more further and more efficiently.
- Always looking for new ways to make sure to **understand the data** in a deeper sense.