Eyasu Berhane Habte

• Appenzeller Str. 94, 81475 München Germany • GitHub: https://github.com/eyasu11321238a?tab=repositories
• eyasuberhanehabte@gmail.com +491631536455 • LinkedIn: https://www.linkedin.com/in/eyasu-berhane-habte/

SUMMARY

Data Scientist with around 2 years of experience in data analytics, machine learning, and big data systems. Holds a Master's in Digital Engineering (Computer Science) from Bauhaus University with a focus on ML, data Science, and Computer Vision. Skilled in Python, SQL, R, Spark, Kafka, and cloud platforms (Azure). Experienced in building data pipelines, deploying ML models, RAG and developing interactive dashboards for data-driven insights.

TECHNICAL SKILLS

- **Programming Languages:** Python, R, MATLAB, Java, Bash, LaTeX
- **Data Science:** NumPy, Pandas, Matplotlib, Seaborn, SciPy, Scikit-learn, web scraping
- **ML & NLP:** TensorFlow, Keras, PyTorch, NLTK, Gensim, Transformers, Langchain, vector database
- **Databases & Warehousing**: PostgreSQL, MySQL, MongoDB, Snowflake, Redshift
- **Data Visualization:** Power BI, Plotly, ggplot2
- **Big Data:** SQL, Hive, Impala, Hadoop Ecosystem
- **ETL & Pipelines**: Azure, Spark, Airflow DAGs, Kafka
- **Cloud Platforms**: AWS (S3), GCP (BigQuery),

PROFESSIONAL EXPERIENCE

Data Scientist | Deutsche Bundesbank| | Frankfurt, Germany (Intern) **11/2024 – 04/2025**

- Developed web scraping scripts to extract and parse over 5000 product data from the Zalando website.
- Performed ETL processes using Apache Spark in CDSW, loaded the data into the Hadoop ecosystem (HDFS), and managed data using Hive and Impala.
- Implemented machine learning algorithms, achieving 85–90% accuracy on multi-class product classification**.**
- Integrated Zalando's product and price data into the dashboard with Otto and Takko for interactive visuals.

Data Analyst | TEKLE CAR SPARE PART SHOP| | Addis Ababa, Ethiopia **09/2020 – 08/2021**

- Analyzed 10k+ datasets related to engine components, tires, filters, and lubricants to identify trends and patterns, guiding strategic decision-making.
- Used statistical techniques to track profitability, identifying the top 15% high-margin products in revenue.
- Developed inventory strategies that cut seasonal losses by ~20% and optimized stock allocation.

PROJECTS

Retrieval-Augmented Generation (RAG) System for Financial Document QA

- Built a RAG system for question-answering financial report data using FAISS and Sentence Transformers.
- Implemented a multi-format document ingestion pipeline with metadata normalization and chunking.
- Developed a vector store with fast similarity search and incremental updates.
- Integrated Groq's Llama-3.1-8B model to generate accurate, context-aware answers.

Data Explorer: No-Code Data Science Web Application [CLICK]

- Built a web app for no-code data profiling, cleaning, visualization, and ML training.
- Developed advanced data cleaning and feature encoding using JavaScript libraries.
- Created interactive, exportable charts (bar, pie, scatter, heatmap, etc.) for intuitive data exploration.
- Developed the app with React, HTML, CSS, TypeScript, TensorFlow.js, and deployed it with Vercel

Effective Neural Translation Language Model

- Designed a Neural Translation Language Model (NTLM) using Word2Vec and transformer embeddings to improve semantic query-document matching in information retrieval.
- Compared NTLM with traditional models, achieving better performance in ranking tasks
- Explored embedding configurations to optimize retrieval and reranking effectiveness.

Tigrinya Word Connect – Mobile Game Application

- Published Android game featuring 500 levels, developed using React Native, Expo, and TypeScript
- Core mechanics include letter wheel, word validation, bonus words, and coin-based rewards
- Integrated Google Mobile Ads and an ad-based hint system to increase engagement and retention
- Bilingual Tigrinya–English user interface enhanced with animations, sound effects, and haptic feedback

Student Exam Score Predictor

- Built a modular ML pipeline to predict student scores, from data ingestion to model evaluation.
- Trained and compared models (e.g., XGBoost, Random Forest, Linear Regression) with persisted artifacts.
- Deployed the ML model via Flask and Docker for real-time predictions through a web UI.
- Set up CI/CD pipeline with GitHub Actions for automated testing, Docker image builds, and deployment.

Crack Detection using image analysis and object recognition

- Developed a model to detect cracks and quantify their length in building structures using image analysis.
- Preprocessed image data using image binarization, contrast stretching, and morphological operators.
- Implemented SVM and Decision Tree Classifier for accurate crack identification and classification.

Power BI Vehicle Accident Report Project

- Built an interactive Power BI dashboard to analyze vehicle accidents by region, time, and severity.
- Cleaned and transformed Excel data using Power Query and DAX for accurate visualization.
- Identified accident trends and high-risk areas to support traffic safety decisions.

EDUCATION

Master's in Digital Engineering (Computer Science) |Bauhaus University | Germany 10/2022 – 05/2025

- Machine Learning, Computer Vision, Image Analysis and Object Recognition, GIS, Software Engineering Algorithm and Data Structure, OOP, Computer modeling for physical processes, Simulation in Engineering

CERTIFICATION [CLICK]

Certification from Stanford University, DeepLearning.AI, and IBM in:

• Data Science • Machine Learning Specialization • Deep Learning Specialization

• Data Analysis with R • NLP • Managing Big Data • Generative AI with LLM

• ETL and Data Pipelines with Shell, Airflow and Kafka

LANGUAGE

- English (C1), German (B1), Mother Tonge Tigrigna and Amharic