Alharth Alhaj Hussein

Informatics Engineer, Data Scientist, Computer Vision

Profile

- Data Scientist with **10+ end to end Machine Learning projects** (TensorFlow, Scikit-Learn, OpenCV, Hugging Face Transformers, Pandas, NumPy, Matplotlib, Seaborn) and expertise in Python.
- Ranked in Kaggle competitions; solved 150+ programming challenges (Codeforces, UVA).

Practical Experience

Data Scientist, ML Engineer
SHAI for AI, Jordan | March 2024 — August 2024

- Constructed a multi-label toxic comment classifier utilizing Bi-LSTM with Attention and pre-trained GloVe embeddings using TensorFlow, achieving a ROC AUC score of 0.977. Integrated a real-time GUI for interactive toxicity analysis across six categories.
- Trained a deep Convolutional Neural Network (CNN) incorporating data augmentation, Batch Normalization, and other techniques, achieving 97% accuracy in classifying handwritten digit images.
- Developed a Support Vector Machine (SVM) to classify Arabic letters from audio inputs with 63% accuracy on 1,400 samples by employing librosa and scikit-learn. Executed advanced audio feature extraction (MFCC, spectral, rhythm features) and integrated a real-time GUI for recording and analysis.
- Engineered key features and fine-tuned a Random Forest model for diamond price prediction, reducing RMSE by 9.4% on over 50,000 entries.
- Enhanced F1-score by 13% with an ensemble stacking model (RF, SVC, KNN) using scikit-learn on more than 14,000 music tracks for genre classification.
- Conducted exploratory data analysis (EDA) on over 5,000 data science salaries using Power BI and created more than five interactive dashboards.

Data Scientist, ML Engineer
Qader, Iraq | September 2025 — Till Now

• Implemented a full pipeline for cleaning and processing Qader Data and using OpenAI API to apply some complex tasks like correct grammar mistakes and ensuring using language Arabic standard and rewrite in Saudi delict, …

• Planned and Implemented a Statistics Engine for the tests' data on the platform the main goal is to extract information from the behavioral data that we obtain from the tests that the student do.

Contact Details:

Aleppo, Syria
+963 991292874
alharth.alhaj.hussein@gmail.com
Date of Birth: 5 Jun 2001
Nationality: Syrian

Links:

My portfolio
LinkedIn
GitHub
Kaggle
Codeforces
Certificates

Languages:

- Arabic (Native)
- English (B2)

Courses:

- **Machine Learning Specialization** (Coursera, Stanford/DeepLearning.AI)
- **Deep Learning Specialization** (Coursera, Stanford/DeepLearning.AI)
- **Generative AI** (Coursera, Stanford/DeepLearning.AI)

Education

Bachelor's in Informatics Engineering
Aleppo University, Aleppo | September 2020 — September 2025

- **GPA**: 82%
- **Relevant Coursework**: AI, Computer Vision, NLP, Deep Learning, Software Engineering.
- **Academic Project**:
  - **Coursera-Based Recommendation System** (Python, Django, SQL, HTML/CSS):
    - Trained a deep learning model (LSTM and Transformer architectures) to generate personalized suggestions using data scraped from Coursera.
    - Developed a full-stack web application with user-friendly interfaces for the recommendation system.
    - Transformed the scraped data into a relational database and integrated MySQL for seamless interaction with the Python backend.
  - **Human Motion Capture System** (Python, OpenCV, Mediapipe, YOLO, Depth anythingv2, PQt6):
    - Developed a desktop application for real-time 2D and 3D human pose estimation from various camera or video sources.
    - Integrated the "Depth Anything V2" model to accurately estimate 3D depth (Z-axis), enabling realistic motion capture for animation and game development.
    - Implemented functionality to stream motion data in real-time via WebSockets to other applications (in my case Unity) or export it to JSON files.

Technical Skills

- **Languages**: Python, C++, Java, SQL, HTML/CSS
- **Tools**: Git, VS Code, Jupiter Notebook, Kaggle and Google Colab notebook, MySQL Workbench, Microsoft Office.
- **Frameworks**: Django, FastAPI, TensorFlow, Scikit-learn, PyQt6
- **Models**: YOLO, MediaPipe, OpenPose, Depth Anythingv2, BERT, T5, GPT-2
- **Domains**: Machine Learning, Computer Vision, NLP, Data Analysis.
- **Data Preparation:** Data cleaning, feature engineering, feature scaling and normalization, handling imbalanced data, and data augmentation for model training.
- **Data Collection & Reporting:** Data collection processes and tools (Google forms, KOBO Toolbox), Regular data analysis and reporting, Data visualization and dashboards using Power BI, python packages (Pandas, Matplotlib, Seaborn).
- **Information Management**: Archiving and documentation, Technical support in data management and analysis.

Volunteer Work:

More than one year of volunteering in **Blue Bits** Team**,** authored 30+ summarized lecture notes for 100+ collage students, improving study efficiency for exams.