

TAREK H

Data Scientist

EXPERIENCES

Data Scientist

Huawei

📅 JUN 2021 — Present 📍 Helsinki

- Performing R&D in computer vision and machine learning for the Huawei Petal search, a global search engine developed by Huawei in the second half of 2020. Petal Search Wins if Award 2022, Signaling Its Dedication to Next-Generation Search Engine
- Working on object detection, segmentation, and recognition in the Fashion Industry.
- Research in optical character recognition, from text detection to recognition using Convolutional Recurrent Neural Networks CRNN and CTC for Latin and other complex languages.
- Models & Prompt Engineering
- Embeddings and Vector Stores/Databases
- Generative AI Agents
- Domain-Specific LLMs
- MLOps for Generative AI (Vertex AI)
- Visual-aware recommendation method to predict CTR of image advertising to rank creatives in display advertising
- Reduce the model's size by applying quantization and pruning.
- We are working on generating synthetic data for text recognition.
- Working with Huawei cloud to train models with multi-GPUs and deploy models as services.
- Develop annotation tools based on deep learning to facilitate annotators' tasks.

Data Scientist

Data Bid Machine

📅 FEB 2020 — JUN 2021 📍 Online

- Data Bid Machine is a Deep Learning AI technology company focused on improving PPC Search Advertising performance. The company's technology amplifies the power of marketers' programs by helping them substantially expand and elevate the volume of relevant, highly performant keyword optimization strategies by 10-50
- Predict the winning bid price of a given auction
- Predict the CTR of new keyword
- Extracting top keywords using NLP models such as BERT, Roberta,... from the product description and competitor's content.
- Classify new keywords according to relevance using classification models.
- Gathering keyword recommendations by using Google auto-suggestion.
- Understand different channels' roles in the customer acquisition
- A/B testing for Ads

EDUCATION

Master's degree in Smart Systems
**National School for Computer Science,
Tunisia**

📅 Sep 2017 – Feb 2020

Bachelor's degree in Computer Science

**Faculty of Sciences and Technology, Sidi
Bouزيد, Tunisia**

📅 2014 – 2017

PROJECTS

A Hybrid Approach for Fake News Detection on Twitter Based on User Features and Graph Embedding

- We tried to find out, the source of fake news, fraud, and scam on online social networks by analyzing the users' profiles and the interactions a user makes in a social network graph (followers, followed by) which gives him a social context.

Food Inspections in Chicago

- Processing and analyzing the Food Inspection's historical data in Chicago to predict the likelihood that a restaurant may earn a critical violation. This can lead to the reduction of thousands of inspection visits that are in charge of annually checking the city, especially when only 15% of these establishments, on average, earn a critical violation.

Predict the value of transactions for potential customers

- Identify the value of transactions for each potential customer in order to personalize services at scale

Home Credit Default Risk

- Predict how capable each applicant is of repaying a loan.

Identify toxicity comments across multiple languages

- The objective is to identify toxicity comments across multiple languages.

Data Scientist Mentor

Kaggle

📅 FEB 2021 — JUN 2021 📍 Online

- Helping grantees build their data science project portfolio.
- Create an exploratory data analysis, shared code/model, participation in a Kaggle competition, a medium post, a YouTube video, etc.
- Define a project (or projects) to tackle, and the grantees will spend 3 months with my mentor working toward producing a tangible deliverable/project that brings them closer to their career aspirations.

Data Scientist Competitor

Kaggle

📅 JAN 2017 — Present 📍 Online

- Competitions Expert. Highest Rank 805 over 150k scientists.
- Kernels Master. Highest Rank 48 over 150k scientists.
- Discussions Expert. Highest Rank 80 over 170k scientists.

Data Scientist Researcher in Social Network

RIADI Laboratory - ENSI

📅 Mar 2019 — Feb 2020 📍 Tunisia

- Creating a hybrid Approach for Fake News Detection in Twitter based on User Features and Graph Embedding.
- Experience with graph analysis, optimizations, and benchmark evaluations.

ACHIEVEMENTS

- My first research paper: "A Hybrid Approach for Fake News Detection in Twitter based on User Features and Graph Embedding" (41 citations).
- Top 6% in TReNDS Neuroimaging competition on kaggle.
- Top 8% in Jigsaw Multilingual Toxic Comment Classification (NLP) competition on kaggle.
- Top 8% in SIIM-ISIC Melanoma Classification competition.
- Top 9% of competitors in the Genetic Engineering Attribution Challenge hosted by DrivenData and altLabs.
- Top 23% in M5 Forecasting - Accuracy competition on kaggle to estimate the unit sales of Walmart retail goods.
- Qualified for African-Arab Collegiate Programming Contest ACPC in Egypt 2019.
- 14th among 200 teams in the Tunisian Collegiate Programming Contest TCPC 2019 and second place in Manouba CPC 2019.

CERTIFICATES

- Machine Learning and Advanced Machine Learning with TensorFlow on Google Cloud Platform Specialization from Google
- Deep Learning and Natural Language Processing Specialization from deeplearning.ai
- Build and Operate Machine Learning Solutions with Azure
- Microsoft Azure for Data Engineering

TECHNICAL SKILLS

- Microsoft Azure: Configuring compute resources, running automated machine learning experiments, and deploying predic-

tive service.

- Python (Matplotlib, Pandas, NumPy, Scikit-Learn)
- ML (Pytorch, Tensorflow, Keras, XGBoost, Random Forest)
- NLP (NLTK, Gensim, spaCy)
- Dash Framework
- SQL, BigQuery