Btool Hamoui

Data Scientist | Data Engineer | Data Analyst | Software Tester

@ batoolmhdmaher@gmail.com   ☎ 0507749208   🐦 @batoolhm90   in www.linkedin.com/in/batool-hamawi-17553269

 https://github.com/BatoolHamawi

Experience

Data Scientist (Internship) | DNA (Design and Analysis)

📅 February 2022 - present   📍 Jeddah

- Retail audit Automation project:

Retail audit invoices information extraction automation:

- Converting the unstructured data in invoices images into semi-structured data (json) using AI-powered OCR API.
- Develop web pages (HTML, CSS, Flask) to manage and coordinates the process of uploading images, assigning users, and loading the extracted information into Postgresql database.

FMCG categorization automation (current work):

- Building classification models by fine-tuning Bert to predict the categories of items (PyTorch).
- Extract the weights and brand of items by named entity recognition NER.

- Google reviews analysis of (Alshiaka stores, Barn's cafe and, Petromin Express) for market research analysis:

- Automating ETL pipeline (web scarping (BeautifulSoup,Selenium) and PostgreSQL).
- Performed correlation analysis, Sentiment analysis and Created Word Clouds to identify common words.

- User Acceptance Testing (UAT) on e-commerce mobile application (current work):

- Worked independently as a Software Tester on executing Test Cases and generated automation scripts using Selenium.

Machine Learning Researcher | Research and Innovation-Deanship of Scientific Research, Umm- Al-Qura University.

📅 April 2021 - November 2021   📍 Remote

Working with team of academic members to conduct research and data science projects.

Training Trainer | Bayan School (Data science and AI)- Contract

📅 January 2021-Present   📍 Remote

Responsible for providing training courses to the students.

Machine Learning Researcher | Center of Innovation and Development in Artificial Intelligence, Umm- Al-Qura University

📅 April 2020 - December 2020

Conduct three research projects (Topic extraction, misinformation detection and correlation investigation) on 6 million COVID-19 tweets dataset.

- Collect Arabic tweets from Twitter and store the collection in Nosql database (using Tweepy API and MongoDB).
- Perform filtration, pre-processing and converting text using feature representation techniques(n-grams, TF-IDF, and building word embeddings (word2vec and Fasttext)
- Deploy Machine Learning models unsupervised (LDA, NMF, and k-means) and supervised (SVM,Logistic regression, NB, RNN)
- Conduct trends investigation, analysis and thoroughly produce visualizations graphics.

Technical Support | Al-Muhagirat Center for women

📅 Sep 2018 - July 2019   📍 Makkah

Summer Trainee | Maternity and Children's Hospital

📅 June - July 2012   📍 Medina

Summary

Over three years of experience in different areas of data science, data ETL, data mining, and software testing. Passionate about working with various data types, deploying machine/deep learning models, and coming up with data-driven insights. I also have extensive experience in the field of NLP, and thoroughly enjoy working with Arabic text and employing methodologies to solve intricate problems.

Education

Master of Science (MS) in Computer Science and Engineering

Umm Al-Qura University

📅 June 2020

Bachelor degree of Information Systems

King Abdulaziz University

📅 June 2013

Software Development Skills

Programming:

Python   R language   C#

DataBase:

MongoDB   My SQL   PostgreSQL

Testing Tools:

Selenium   Jira   Postman

Data Visualization:

Tableau   Dash - Plotly   Power BI

Analysis Skills

- Experience in creating and supporting a data management workflow from data collection, storage, and analysis to training, validation, and data visualizations.
- Fluency with machine learning algorithms, such as CNN, RNN and fine-tune transformer.
- In-depth Knowledge of NLP techniques such as PoS tagging, Named Entity recognition, Topic Modeling, and Word embeddings.
- Working with various data formats such as PDF, JSON, XML and CSV.

Projects

2022 | Plant disease detection
Deploying Convoluitona neural Netowork CNN for classifying Leaf images.

March - May 2021 | Arabic Sentiment Analysis
Develop sentiment analysis model using AraBERT. Tools: Python (CAMeL and pyTorch).

2020 | Mining Arabic Twitter during COVID-19 pandemic
Extracting topics and misinformation detection from 6 million Arabic tweets. Tools used: MongoDB, Python.

2017- 2019 | Event Detection from Arabic Twitter.
Deploying machine learning models and NER to detect events from 3 million Arabic tweets. Tools used : MongoDB, Python.

2018 | Arabic Chatbot
Volunteer Work.

2016 | Software Bug Prediction
Detect the severity of a bug report by using machine learning algorithms. Tools used: Python

2013 | School Bus Attendance System Using RFID.
Tools used: C#, SQL Server, Asp.net)

2012 | The Blood Bank Donation System
The Blood Bank Donation System Project for Summer training. (Tools used: MySQL, Java)

Certificates

- Online Course, Software Testing Bootcamp, Tunga Academy - April, 2022
- Online Course, Business Analytics Nanodegree Program, Udacity- January, 2022.
- Online Course, Regulatory and Supervisory Technology (RegTech SupTech), Udemy - October, 2021.
- Online Course, Interactive Python Dashboards with Plotly and Dash, Udemy - 2021.
- Online Course, Information Retrieval, June 2021.
- Online Course, Sentiment Analysis with Deep Learning using BERT, Coursera- March 22, 2021
- Online Course, Supervised Learning with scikit-learn, DataCamp-April,2020.
- Online Course, AI Fundamentals, DataCamp-April,2020.
- Certificate of Attending training course about **the R language to analyze and visualize data, 2020**.
- Certificate of Attending Training of Trainers (TOT) Course, Makkah Chamber, 30th June to 4th July, 2019.
- Certificate of Attending Machine learning by DevRel MENA Study Jams & TOT Saudi Arabia from google, January 2019.

Publications

- Hamoui, B., Mars, M., Almotairi, K. (2020, May). FloDusTA: Saudi tweets dataset for flood, dust storm, and traffic accident events. In Proceedings of the 12th International Conference on Language Resources and Evaluation (LREC 2020).
- Hamoui, B., Alashaikh, A., Alanazi, E. (2020, December). COVID-19: What Are Arabic Tweeters Talking About?. In In Proceedings of the 9th International Conference on Computational Data and Social Networks (pp. 425-436). Springer, Cham.
- Hamoui, B., Alashaikh, A., Sherif, A., Alanazi, E., Nabil, M., Alsmary, W. (2021, December). Google Searches and COVID-19 Cases in Saudi Arabia: A Correlation Study. In 2021 3rd IEEE Middle East and North Africa COMMunications Conference (MENACOMM) (pp. 104-108). IEEE.
- Alqurashi, S., Hamoui, B., Alashaikh, A., Alhindi, A., Alanazi, E. (2021). Eating Garlic Prevents COVID-19 Infection: Detecting Misinformation on the Arabic Content of Twitter. arXiv:2101.05626. Manuscript submitted for publication.

Seminars - Conferences

- **Kaggle Competition**, Arabic Sentiment Analysis, March-May 2021.
- **Homathon Hackathon**, Saudi Federation for cybersecurity, Apr, 2020.
- **Covid-19 Challenge**, Saudi Data Community, March, 2020.
- Guinness World Records for participating in Hajj Hackathon, Aug 2,2018.
- **Hajj Hackathon**, Aug 1-3, 2018.
- First AI Conference on Smart Societies Infrastructure, Technologies and Applications (SCITA 2017), Jeddah, 27-29 Nov 2017.
- Participated with the (SBAS) project in the Information technology Demo Day 8, King Abdulaziz University, 12 Sep 2013.
- Participated with the (SBAS) in The Fifth Annual Undergraduate Research Conference on Applied Computing (URC 2013), Dubai ,12May 2013.
- Cloud computing conference, King Abdulaziz University, 2011.

Languages

Arabic (Native)

English (Professional Working Proficiency)