Ayoub KOUSSY
Data Engineering | Looking for First Opportunities

@ ayoubkoussy0@gmail.com   📱 +212 6 22 09 53 36

in LinkedIn : Ayoub Koussy   🌐 Website : koussyayoub.github.io   GitHub : KoussyAyoub

🎓 SCHOLARSHIP

2021-2024	National School of Computer Science and Systems Analysis (ENSIAS), Rabat, Morocco.
2019-2020	Preparatory Classes in the MP (Mathematics and Physics) option, CPGE Réda Slaoui, Agadir, Morocco.
2018-2019	Baccalaureate in Mathematics and Science A, Al Maarifa High School, Ait Melloul, Morocco.

💼 PROFESSIONAL EXPERIENCE

February 2024 – July 2024

Data Engineering Intern, H E N C E F O R T H , Rabat , MOROCCO

During this experience, I worked on a PFE project titled "Big Data Infrastructure Enhancement with Query Expansion". This project was divided into two parts :

> **Designing a Batch Pipeline for Data Processing :** This involved collecting, storing, and processing data from the YouTube API. As part of this phase :
> - Designed the entire pipeline component and conducted benchmarks among all available open-source tools that we could use.
> - Configured the pipeline and addressed all technical aspects.
> - Held weekly meetings and provided daily reports to track progress and address any technical issues during the configuration.
> - Collected and processed approximately 150,000 videos and 134,000 channel pieces of information over a period of 20 days.

> **Benchmarking Methods of Query Expansion :** The second part involved comparing available methods of query expansion :
> - Tested the capabilities of LLM, the Google API, and large graphs for the specific task of query expansion.
> - Integrated both parts of the project to ultimately create a comprehensive research system.

Technologies : | Apache Hadoop | Apache Spark | Shell scripting | Python | Apache Airflow | Docker | Json | LLMS

June 2023 – August 2023

Data Engineering Intern, D A T A P R O T A E C T , Casablanca, Nearshore Park, MOROCCO
- Development of a resilient pipeline for collecting, analyzing, and storing logs.
- Designed and implemented a machine learning model that achieved a notable 93% accuracy rate for analyzing logs to predict cyberattacks.
- Integrated the model into the pipeline for real-time monitoring of logs.

Technologies : | Elasticsearch | Kibana | Logstash | Winlogbeat | Apache Kafka | python | Json | Docker | scikit learn

July 2023 – Septembre 2023

Machine Learning (ML) Intern, T E N D E R P O R T , Munich, Bavaria, Germany
- Development of an AI model for address extraction, with the aim of replacing the use of GPT-3.
- Achieved a notable 96.72% accuracy rate with the developed model, closely rivaling that of GPT-3.
- Deployed the model and integrated APIs, facilitating real-time data acquisition
- Create a robust data warehouse capable of efficiently storing and processing large datasets.

Technologies : | Web-scraping news | NLP classification | Google BigQuery | python | transformers | flair | NLTK | fastApi

June 2022 – August 2022

Virtualization Engineering Intern, O C P G r o u p , SAFI, MOROCCO
- Creating a virtualization architecture for two physical servers using VMware solution.

Technologies : | VMware vSphere Client | VMWare ESXi | Windows server

💻 ACADEMIC PROJECTS

> **Real time social network analysis on tweeter (X) :**
> - Create a Big Data application for real-time processing of massive data, enabling the monitoring of interactions on Twitter (X) related to the Israeli-Palestinian conflict.
>
> Technologies : Apache Kafka· Apache Spark · BD Nosql (Mongo DB) · Mlflow · DVC · Airflow · Power BI

> **Designing an ETL for Big Data Weather analysis :**
> - Developing a pipeline for the extraction, transformation, and analysis of meteorological data from Moroccan stations recorded from 1980 to 2023, with the aim of visualizing climate changes in Morocco.
>
> Technologies : Apache Airflow . Apache Hadoop · Python · Java · MonogDB · PowerBI · Time series processing.

☰ Skills

Languages	Arabic (Native), French (Fluent), English (Fluent)
Programming	Python, R, Java, Scala,Javascript
Libraries / Frameworks	Pandas, GeoPandas, OpenCV, Scikit-Learn, NumPy, Pytorch, Pytorch Lightning, Transformers, TensorFlow, Keras, Fastapi, Flask, Pytest, Torchvision, Selenium, Scrapy, BeautifulSoup,NLTK, spaCy, Gensim
Big Data	Apache Hadoop, Apache Spark, Apache Kafka, Apache Airflow, Pyspark , Data warehouse
Database	MonogDB, Elasticsearch,Kibana ,MYSQL, Neo4j
Software /Tools	Git, Docker, Cuda, Jupyter, LaTeX, Postman, AWS,PowerBI,Jenkins, SonarQube
Deep Learning	CNN, RNN, LSTM, Transformer, Graph Neural Network (GNN), Variationl Autonencoder (VAE) , LLMS