Mohamed Bilel Hasni

Data Scientist

Country of residence : Luxembourg
Mobile : +352-661702111
Email : hasnimedbilel@gmail.com

Portfolio https://hasnimedbilel.github.io/
Competence areas Data Engineering, Microsoft Azure, Databricks, Data Science, Machine Learning, Generative A
Statistics

Technos

Languages
- Python, Java, SQL,R

Big data Frameworks
- Spark, Pandas, Dask

Cloud
- Microsoft Azure
- Databricks

Machine Learning/NLP
- Pytorch
- HuggingFace
- Scikit-learn

Data Viz
- Tableau, Microsoft PowerBI

OS
- Linux, Windows, MacOS

Certifications

Microsoft Azure
- Microsoft DP-203 : Azure Data Engineer Associate [Link]
- Microsoft AZ-900 : Azure Fundamentals [Link]

Databricks
- Databricks : Certified Data Engineer Associate [Link]

Education

2016–2019 **National Engineering School of Sousse, Tunisia**,
Computer-Science Engineering degree.

2014–2016 **Preparatory Institute of engineering studies el Manar, Tunisia**,
Mathematics & Physics.

Experience

Jan 2023–Ongoing **Data Engineer**, *SFEIR, Luxembourg.*
- Big Data analytics/processing, design and architecture of ETL pipelines in Azure to support data ingestion&integration of operational data sources.
- Big Data Engineering, Azure Synapse Analytics, Databricks, ETL, Spark.

Jan 2020–December 2022 **Data Scientist**, *Putnam PHMR.*
- Application of AI in the healthcare industry: Clustering, unsupervised learning, Natural Language processing, Image processing ...
- Data Engineering, Data Warehousing, ETL, PySpark, SQL, Dask, Pandas.
- Data visualization & Dashboarding.

Jan 2019–Oct 2019 **Data Science Intern**, *SmartTek Services, Dubai.*

- Implemented a research paper on most important factors in uncovering Simbox fraud behaviour.
- Handling large telco datasets: Extraction of CDR (Call Data Records), Create Datawarehouse using Apache Hive.
- Data cleaning and transformation using Pyspark.
- Data loading into Rapidminer Platform.
- Implemented a Machine Learning solution to detect Simbox fraud in a Tunisian Telco operator.
- Implemented a distributed version of the solution using Hortonworks Data Platform (HDP) and Rapidminer, on a cluster of VMs.
- Monitoring of the HDP cluster using Apache Ambari.

keywords: DataWarehousing, ETL, Hive, SQL, Pyspark, Rapidminer, HDP, Apache Ambari

May 2018–Sep 2018 **Data Science Intern**, *Kaoun, Tunisia.*

- Design and creation of a data Warehouse using the ETL integration process.
- Implemented Machine Learning techniques to predict bank customers loan repayment ability.

Projects

InstaPaper **Data Capturing from audit report images using Deep Learning**, *Project funded by the World Bank.*

- Designed and implemented a Deep Learning solution to help the court of audit of Tunisia reduce the manual work of processing audit reports.
- Automatic Data capturing of report images.
- Automatic generation of custom Key Indicators & visualization Charts/graphics.
- Productionize the model as a REST API.

keywords: Data Capturing, Python, Pytorch, Deep Learning, Image segmentation, Detectron, Resnet, Text Detection (CRAFT), Text Recognition (CRNN), FLASK

TAK **Treatment sequences analysis using Kmeans Clustering**.

- Developed a state of the art solution to represent patient's treatment sequences in an intuitive visual format.
- Extraction, transformation and loading of clinical data from different sources to create a DataWarehouse. The objective is to obtain, in a single view, all patients' treatment data. (ETL, SQL, PostgreSQL)
- Used unsupervised learning coupled with image processing techniques to reveal hidden patterns in complex treatment sequences.
- Automatic assessment of the stability of clusters of patients found.
- Productionized the solution into a web application using Flask, bootstrap, Docker.
- Accepted Poster publication for **ISPOR 2022** conference. Link : [tak_abstract]

keywords: ETL, SQL, PostgreSQL, Needleman Wunsch sequences alignment algorithm, Hierachical Clustering, Image processing, Clustering stability evaluation, FLASK, Bootstrap, NGINX

Abstractron **NLP solution to support in literature abstract selection and text classification process**.

- design/implementation of a ETL+Datawarehouse system in PostgreSQL in order to collect abstracts and prepare them for further analysis.
- design/implementation of a multi-modal solution to deal with inputs of different types.
- re-train Transformers using BioBert pretrained embeddings (Masked Language Modeling)
- averaging embeddings of different inputs to produce a final data representation used for abstract classification task.
- define custom loss function to deal with class imbalance.
- Create a Rest API of the final model in order to be consumed in a web App.

keywords: Python, Datawarehouse, ETL, Data preprocessing, Pytorch, Transformers, BioBert embeddings, domain adaptation, Attention models, encoder-decoder architecture, multi-modal learning, FLASK

Publications

2022 TAK (Treatment Sequences Analysis Through K-Clustering), [ispor]

2022 Could Artificial Intelligence Support Prediction of Reimbursement Decisions in Scotland? A Pilot Project, [ispor]

2022 Promises of AI-Assisted Patient Monitoring Methods, [ispor]