

Murad Daryousse

Data Scientist and Engineer

+971561101252 | mdaryousse.ds@gmail.com | Abu Dhabi, UAE | LinkedIn | Kaggle | Github

SUMMARY

Results-driven data scientist and ML/Data engineer with a proven track record in developing innovative solutions that drive business insights and efficiencies. Experienced in building and optimizing machine learning models and data pipelines to drive business insights and automation. Adept at translating complex data into actionable strategies and fostering collaboration across cross-functional teams to achieve organizational goals.

TECHNICAL SKILLS

Programming

- Python: PyTorch, Keras, Tensorflow, OpenCV, Scikit-learn, Pandas, Numpy, NLTK, Matplot, Django, Flask, FastAPI, Scrapy.
- Java: Spring Boot, Apache Jena.
- Scala
- R

Machine Learning and Deep Learning

- Supervised, Semi-Supervised, and Unsupervised Learning: Regression, Classification, Clustering, Anomaly Detection.
- Feature Engineering and Model Selection
- Hyperparameter Tuning: Optuna, GridSearchCV, Bayesian Optimization.
- NLP
- Computer Vision

MLOps

- Model Monitoring and Drift Detection: Evidently AI, WhyLabs.
- Experiment Tracking: MLflow, ClearML
- Data Version Control: DVC

Data Engineering and Analytics

- Big Data: Apache Hadoop, Apache Spark, Apache Hbase, Apache NIFI, Apache Hive, Apache Impala, Presto/Trino.
- Data Base: MongoDB, PostgresDB, Janusgraph, Neo4j, Oracle, MySQL, SQL, PL/SQL
- Message Brokers: RabbitMQ, Apache Kafka.
- Orchestration: Prefect, Apache Airflow.
- Data Analytics: dbt, Dimensional Modeling.
- Data Visualization: PowerBI.
- Logging and Monitoring: Elasticsearch, Kibana.
- Cloud Computing: AWS Batch, AWS Lake Formation, AWS S3, AWS Kinesis, AWS Glue, AWS ECS, AWS ECR.
- Data Architectures: Data Warehouse, Data Lake/Lakhouse, Data Mesh, Data Fabric, Lambda
- ETL/ELT pipelines

DevOps: Git, Github, CircleCI, JFrog, GitLab, CI/CD, Terraform, Docker, Kubernetes, Linux

Semantic Web/Knowledge Graph: OWL, RDFS, RDF, SPARQL, SWRL

PROJECTS

RoboWell

Jul 2024 – Present

AI-powered autonomous system to optimize and stabilize oil well production

AIQ

- Engineered and deployed end-to-end machine learning pipelines to continuously train and evaluate inferential models, driving optimized production for oil wells and improving operational efficiency.
- Designed and implemented Prosper models auto-calibration, which enhanced inferential models accuracy and decision-making processes.
- Tech Stack: Python, Scikit-learn, smt, MLflow, Prefect, DVC, GitLab

Novelty Detection for Data Quality

Jan 2024 – May 2024

Novelty detection models for reports and data marts pipelines quality control

LigaData

- Utilized Local Outlier Factor and Isolation Forest methods as semi-supervised models for detecting abnormal/novel data points
- Achieved 30% enhancement in business reports data quality
- Built a RESTful API for trained models to be used by data pipelines quality checks
- Tech Stack: Python, Scikit-learn, FastAPI

Enterprise Data Lakehouses

Feb 2022 – Jul 2024

Data lakehouses projects tailored for mobile operators using Data Flare Platform

LigaData

- Led data engineering team in implementing data lakehouse and analytics solutions in multiple OpCos: MTN Congo, MTN South Sudan, MTN Zambia, and Zain South Sudan
- Developed and deployed ETL/LET and analytics pipelines
- Built reports, data marts, and business semantic layer
- Modeled data for analysis using dimensional modeling methods
- Tech Stack: Cloudera, Hadoop, Spark, Kafka, Hive, Hbase, Presto, Impala, MongoDB, Airflow, NIFI, Python, Scala, SQL, DSL, dbt, Talend

Flare Time Machine

Jan 2023 – Jul 2024

Big data management, archiving, and protection solution

LigaData

- Led data engineering team to implement data archiving, management, and retention policies for multiple OpCos: MTN Ghana, and MTN Uganda
- Customized and deployed Flare Time machine product for different OpCos
- Modeled data for analysis using dimensional modeling methods
- Tech Stack: Cloudera, Hadoop, Presto, Airflow, Scala, SQL, Shell Scripts.

Meta Coverage Data

Feb 2022 – Dec 2022

ETL pipeline for providing Meta (Facebook) with MTN coverage data

LigaData

- Developed ETL pipelines to provide Meta with coverage data for multiple MTN OpCos: MTN Zambia, MTN Uganda, and MTN Nigeria
- Built data quality monitoring dashboards for coverage data SLA with Meta
- Tech Stack: Cloudera, Hadoop, Hive, Presto, Airflow, Python, SQL, Shell Scripts.

Duwaween Games

Feb 2021 – Feb 2022

Collection of card games

KuwaitNet

- Developed data meshes and event-driven micro-services
- Tech Stack: Python, Django, Spring Boot, Redis, AWS, Docker, Kubernetes, RabbitMQ.

Fake News Detection

Aug 2020 – Dec 2020

Fake new detection platform

Almeta

- Architected social media and news websites data scraping platform
- Utilized NLP methods and tools to clean and preprocess textual data for ML training and inference
- Trained PassiveAggressive classifier as an online ML method to news to fake or real
- Achieved 92% accuracy score on the news dataset
- Tech Stack: Scrappy, Python, Pandas, Scikit-learn, NLTK, Kafka, Redis, CircleCI, Terraform, AWS.

Social Listening Platform

Feb 2020 – Aug 2020

Collecting, extracting and analyzing data from social media for sentiment analysis and getting insights

HiFi

- Designed and implemented social media data scraping platform
- Used NLP to clean and preprocess textual data for ML training and inference
- Utilized LSTM with SVM to identify the sentiment of texts
- Achieved 94% accuracy score on tweets dataset
- Tech Stack: Python, Scrapy, Scikit-learn, Keras, Tensorflow, NLTK, Hadoop, Hbase, Kafka, Spark, Elasticsearch, Kibana.

Ekhtar

Aug 2018 – Jul 2020

Real-time reporting and crisis detection system

HiFi

- Architected and developed big data platform using lambda architecture
- Used NLP to clean and preprocess textual data for ML training and inference
- Built a knowledge graph framework for crisis detection and tracking
- Tech Stack: Kafka, Hadoop, Spark, Hbase, Elasticsearch, SQL, MySQL, XMPP, Java, Python, Semantic Tech (Knowledge Graph), NLTK, Docker, Linux.

Mobile Regulation Center (MRC)

Jun 2016 – Jul 2019

Mobile devices regulation system to detect and prevent their smuggling

HiFi

- Developed ML model for detecting mobile devices smuggling with 93% accuracy
- Developed Back-end and web APIs
- Tech Stack: Scikit-learn, python, Oracle, PL/SQL, ORDS, PHP, MySQL, Java Spring Boot

EXPERIENCE

ML/Data Consultant

Jul 2024 – Present

AIQ

Abu Dhabi, UAE

- Spearheaded initiatives and implemented proof-of-concept projects leveraging modern MLOps and data orchestration tools. This achieved a 60% improvement in development productivity, ML models traceability, ML/data pipeline reproducibility, and operational observability.
- Designed and built scalable AI and ML solutions to optimize oil production, enhancing the well's efficiency and stability.
- Partnered with cross-functional teams, driving the successful delivery of AI solutions aligned with business objectives and technical requirements.

Data Analytics Team Lead

Dec 2023 – Jul 2024

LigaData

Dubai, UAE

- Managed a team of 5 data analytics engineers, overseeing project timelines and deliverables.
- Spearheaded the implementation of cutting-edge data solutions.
- Supervised data lakehouse projects for multiple telecom OpCos.
- Developed ML models for outliers/novelty detection for data quality control.
- Led Flare Time Machine product implementation and deployment for data archiving, data storage tiering, and data protection for multiple telecom OpCos.
- Conducted performance reviews and provided technical training to the team.

Senior Data Engineer

Feb 2022 – Dec 2023

LigaData

Dubai, UAE

- Developed business semantic reports and data marts.
- Developed ETL/ELT pipelines into data lakehouses.
- Implemented data quality dashboards.
- Developed CI/CD pipelines.
- Led the development of advanced data architecture and integration solutions.
- Mentored junior data engineers, guiding best practices and project execution.
- Conducted data quality assurance to ensure data accuracy and consistency across systems.

Data Science Team Lead

Jul 2018 – Feb 2022

HiFi

Damascus, Syria

- Led data science and analytics team of 6 data scientists and engineers

- Implemented cutting-edge techniques and tools in machine learning, deep learning and artificial intelligence to make data analysis more efficient
- Performed large-scale experimentation to identify hidden relationships between variables in large datasets
- Created advanced machine learning algorithms such as regression, simulation, scenario analysis, modeling, clustering, decision trees and , neural networks
- Prepared and extract data using ETL/ELT best practices
- Coordinated research and analysis activities using unstructured and structured data and use programming to clean and organize data

Big Data Engineer

Jul 2017 – Jul 2018

HiFi

Damascus, Syria

- Designed, built, and maintained big data architectures, systems, and pipelines to handle large datasets.
- Collected, aggregated, cleansed, transformed, and enriched data from various sources (billing transactions, social media, Apps databases...) to prepare it for analysis, often involving ETL processes.
- Stored and managed data in data warehouses, data lakes, and cloud platforms to ensure accessibility and scalability.
- Implemented technical processes and business logic to transform data into meaningful information, ensuring quality, governance, and compliance with organizational standards.
- Developed NLP solutions for text analysis and knowledge graphs
- Implemented big data storage and processing solutions
- Modeled data using different modeling techniques (like dimensional models) for analytical queries.

Software Engineer

Jun 2014 – Jul 2017

HiFi

Damascus, Syria

- Designed and developed softwares and applications, following the software development life cycle (SDLC), which includes creating flowcharts, layouts, and documentation to identify requirements and solutions.
- Documented software designs, functionality, and updates for future reference, including writing reports to communicate with stakeholders and maintain documentation of IT systems.
- Wrote and tested code, ensuring it is well-designed and testable, including debugging and troubleshooting issues to maintain functionality and fix problems.
- Deployed, productionized software and evaluate user feedback to improve features, releasing updates and ensuring compliance with project plans and industry standards.

EDUCATION

Master of Science in Data Science

Damascus University

Damascus, Syria

Jun 2018 – Feb 2020

Bachelor of Science in Software Engineering

Damascus University

Damascus, Syria

June 2010 – Jul 2014

CERTIFICATIONS

- Data science orientation - IBM - coursera.org/verify/GU38YYGX2N2U