Mehdi EL HAYLALI

Data Scientist

Phone	Email	Location	LinkedIn
+212 708581004	mehdi.el.haylali@gmail.com	Morocco	www.linkedin.com/in/mehdi-el-haylali

Profile

Senior data scientist with 4 years of consulting experience in the Analytics & AI field. My proficiency in Machine Learning, MLOps and Data engineering enables me to develop and deploy scalable solutions that meet business needs. I efficiently collaborate with cross-functional teams to integrate data-driven insights into business decision-making. Example projects include developing a recommendation system for a retail business, predicting customers upsell as well as predicting RFPs' outcomes. In addition to my traditional ML expertise, I earned practical skills in developing LLM-based solutions using langchain framework and frontier models through hands-on trainings.

Key Skills

- Data science, MLOps, Advanced Analytics, Data Engineering, Generative AI (OpenAI, MistralAI, Prompt Engineering)
- Python, R, Spark, SQL, DBT, Git, Docker, scikit-learn, keras, pytorch, CatBoost, LightGBM, MLflow
- GCP, AWS, Azure, Databricks, watsonx.ai
- Analytical thinking, Problem solving, Efficient communication

Work Experience

Data Scientist/Engineer

IBM Consulting, Morocco — April 2021–present

Client: IBM Consulting - internal
Project name: RFP's outcome (Win/Loss) prediction
Objective: Developed, deployed and automated an end-to-end machine learning solution that generates an RFP's win probability and enables Offering Managers to tailor their RFP responses for higher win probability.
Achievements:

- Analyzed IBM signings data using Spark, Python and SQL, processing 600K+ rows of data from a PostgreSQL RDBS.
- Selected and engineered RFPs' features like offering name, customer segmentation, opportunity owner using scikit-learn.
- Built and experimented with many classification models (Logistic Regression, Random Forest, Gradient Boosting algorithms, ANN) using scikit-learn, keras, LightGBM, CatBoost, XGBoost.
- Performed hyperparameters-tuning on CatBoost model using HyperOpt resulting in +80% accuracy.
- Deployed the trained CatBoost model in production as online endpoint for real-time predictions.
- Deployed the trained CatBoost model in production for scheduled batch predictions, scoring 60K+ open deal each week.
- Integrated model's batch predictions into the PostgreSQL RDBS to feed BI tools (tableau).
- Built automated spark jobs for data preparation.
- Built automated python jobs to monitor the model's performance on a weekly-basis and trigger model retraining in case of performance drop.
- Integrated shapley values to explain individual predictions and hence make the model interpretable and trustworthy.
- Developed a creative solution that suggests what changes to make to an RFP's features to increase its win probability.
- Currently working on data migration from Postgres to Db2 and watsonx.data.

Tech stack: IBM Cloud, watsonx.ai, watsonx.data, Python, Spark, SQL, Postgres, Db2, Presto engine, scikit-learn, Keras, CatBoost, HyperOpt, LightGBM, XGBoost, SHAP.

Client: Nespresso USA
Project name: Customer upsell prediction
Objective: Developed a predictive model that estimates customers' upselling likelihood. This allows the business to focus their marketing efforts on customers with high potential.
Achievements:

- Analyzed Nespresso US sales data for capsules, coffee machines and accessories using spark on databricks.
- Labelled customers based on upsell occurrence.
- Engineered features like purchase frequency, total expenditures, coffee preferences and other Nespresso products consumption behaviours.
- Built the end-to-end data preparation spark job.
- Built and experimented with many classification models.
- Enabled machine learning experiments tracking besides seamless models' management and serving using MLflow.
- Performed hyper-parameters tuning for LightGBM model to achieve optimal performance in terms of cumulative gains and lift.
- Deployed the trained model for on-demand batch predictions.
- Generated model explainability charts like Dependence Plots and Force Plots to derive insights on features impacts.

Tech stack: Azure Databricks, Delta Lake, Spark, SQL, MLflow, HyperOpt, scikit-learn, CatBoost LightGBM, Keras, SHAP, CRISP-DM methodology.

Client: Nespresso USA
Project name: Content-based recommendation system
Objective: Built a content-based recommendation system to provide personalised coffee recommendations to Nespresso customers.
Achievements:

- Gathered attributes that describe each Nespresso coffee capsule product like intensity, acidity, roastiness, flavor, aroma..etc.
- Vectorized the collected coffee attributes using scikit-learn.
- Collected and ranked each customer's coffee preferences based on their purchase history.
- For each customer, compute the cosine similarity between its preferred coffees and other coffee items.
- Recommend the top-N similar coffee items for the customer.
- Evaluated model performance using Mean Average Precision & Coverage.
- Enabled experiment tracking using MLflow.

Tech stack: Azure Databricks, Delta Lake, Spark, SQL, MLflow, scikit-learn, CRISP-DM methodology.

Client: DS Smith UK
Project name: Production capacity planning app
Objective: Build a serverless web application to facilitate production capacity planning for supply chain managers.
Achievements:

- Built the end-to-end data ingestion pipeline.
- Designed and implemented the data model for the application in AWS Redshift.
- Translated complex business logic into concrete python code and SQL queries and packaged that as Lambda Functions that get triggered by the user's actions in the frontend.
- Requirements and knowledge gathering from the business stakeholders.

Tech stack: Python, SQL, AWS Glue, AWS Lambda, AWS Redshift & S3.

Client: Newmont Australia
Project name: Advanced analytics for mining operations
Objective: Delivered data products and solutions to support mining operations in Boddington-Australia.
Achievements:

- Batch and real-time data ingestion and parsing from various data sources using Google Cloud Functions, Cloud Run, Python and Docker.
- Designed and implemented advanced SQL queries to answer the business' analytics needs using BigQuery, DBT and jinja.

Tech stack: Python (Pycharm IDE), GCP (GCS, Cloud Functions, Pub/Sub, Cloud Scheduler, Cloud Run, BigQuery), Docker, SQL, DBT, git.

Client: General Motors
Project name: Data science projects migration
Objective: Migrated two data science projects from SAS to Pyspark and SparklyR.
Achievements:

- SAS code migration to PySpark and SparklyR (ETL + time series analysis and forecasting codes).
- Optimization of the Spark jobs' execution time by tuning cluster configuration parameters and implementing caching strategies.
- Productionization of python and R scripts.

Tech stack: Python & R (JupyterLab), Apache Spark & Hive, statsmodels.

Data Science Intern

Saint-Gobain Crystals, France — Mar 2020–Aug 2020

- Created an interactive dashboard to help the customer service department monitor production lead-times and delivery delays.
- Investigated the root cause for material consumption imbalance using data analysis techniques.
- Established a mathematical model to estimate and optimize the production capacity.

Tech stack: Python (Pandas, Bokeh, Holoview, Panel), Data analysis and visualization.

Education

Ecole Centrale Casablanca, Morocco

September 2017–October 2020

Master of engineering — Data science

Courses: Big Data technologies, Statistical Learning, Applications for Deep Learning, High-dimensional Statistics and Massive Data, Optimization.

Preparatory classes, Morocco

September 2015– October 2017

Mathematics, physics & engineering (MPSI-MP*)

Certifications

- Watsonx.ai Technical Sales Advanced – 2024
- IBM Consulting Way Habits - 2024
- Machine Learning Engineering for Production (MLOps) – coursera – 2023
- Databricks Certified Machine Learning Professional – 2023

- Databricks Certified Machine Learning Associate – 2023
- AWS Certified Machine Learning – Specialty – 2022
- Microsoft Certified: Azure Data Scientist Associate – 2021

Languages

English - fluent | French – fluent | Arabic – native