Mahmoud TRIGUI « Data Scientist »

Street El Ain Km 6, Environnemental Boulevard
3042 Sfax – TUNISIA

Mobile phone TN : (+216) 97 304 501
Website Mahmoud Trigui — Data Scientist
E-mail trigui.mahmoud7@gmail.com
LinkedIn www.linkedin.com/in/mahmoudtrigui
Zindi https://zindi.africa/users/Mahmoud_Trigui/competitions
Portfolio https://tinyurl.com/MTprtflio
37 years

Profile

Data Scientist with **10+ years** of experience specializing in **structured** data analytics/machine learning, time series forecasting, and NLP solutions. With expertise in both **R** and **Python** ecosystems. I am particularly passionate about feature engineering and its crucial role in machine learning effectiveness.

Core competencies include:

- **Structured Tabular data Analysis & Machine Learning solutions**
- **Time Series Forecasting**
- **Clustering algorithms**
- **Advanced Feature Engineering**
- **NLP & LLM implementations**

Consistently expanding expertise through **competitive** programming and continuous **learning** from leading educational platforms. Combines technical proficiency with strong analytical thinking to deliver impactful **data-driven solutions**.

Professional Experiences

Professional Positions

February 2024 – Present - **Sofrecom Tunisia (Orange Group) (Sfax Hybrid, Tunisia) - Data Scientist**

Image Classification Pipeline with Multi-Stage LLM Analysis

- **Image Triage & Classification System**
  - Implemented intelligent filtering to separate useful imagery from irrelevant content
  - Developed specialized classification for telecom infrastructure (photos vs. cartographic plans)
- **Technical Implementation**
  - Created custom LLM **prompts** for context-specific classification tasks
  - Implemented base64 encoding and image preprocessing for LLM compatibility
- **System Features**

- Multi-stage classification pipeline: utility detection → photo/plan segmentation → infrastructure type identification
- **BigQuery** integration for structured data logging and performance **tracking**
- Automated image **routing** and storage organization based on classification results
- Confidence scoring and detailed justification generation for each classification decision

Time Series Forecasting Framework for Resource Planning

- **Project Scope**
  - Developed 12-month and 52-week **forecasting** system for **multiple independent** time series
  - Created forecasts for diverse targets to support budget, resource, and staffing planning
- **Technical Implementation**
  - Built a semi-**automated** forecasting framework in Python using **MLforecast** library
  - Designed 3-step **optimization** process:
    - Optimal lag selection
    - Feature variable identification
    - Model and hyperparameter tuning
  - Implemented comprehensive **feature engineering** including:
    - Date-based variables
    - Seasonality/ Trend/ Periodicity components
  - Model and hyperparameter tuning
- **Techniques**
  - Applied target **transformation** functions for improved accuracy
  - Developed **ensemble** modeling approach for result **smoothing**
  - Created automated pipeline for handling multiple independent series and differrent **granularities**
- **Client Engagement & Visualization**
  - Presented forecast results directly **interactive dashboards**
  - Facilitated data-driven decision making through **PowerBI** and **LookerStudio**

PDF Audit Information Extraction System with LLM

- **Intelligent Document Processing**
  - Developed automated extraction system for audit reports using **GPT-4**
  - Implemented multi-page **PDF** processing with **image conversion**
  - Built robust validation system for extracted data
- **Technical Implementation**
  - Designed modular Python architecture for scalable processing
  - Created custom data **validation** using Pydantic models
  - Built automated **JSON** output generation
- **System Features**
  - PDF to image conversion with optimized resolution
  - **Base64** encoding for **LLM** compatibility
  - Structured data validation and normalization

FTTH Access Orders Forecast

- **Project Scope**
  - 12-month forecast horizon (weekly/monthly granularity)
  - Multi-dimensional predictions: global, departmental, and offer-specific
  - Limited historical data and feature availability
- **Technical Implementation**
  - Developed comprehensive feature engineering pipeline
  - Leveraged advanced time series **decomposition** techniques
  - Implemented ensemble modeling approach using R's **feasts** and **fable**
  - Achieved ~5% RMSE through optimized model combination
- **Production Deployment**
  - Migrated solution to Dataiku platform
  - Converted R code into production-ready flow zones
  - Validated performance using Dataiku's native models

Co-Financing Prediction Model (Data Pipeline Migration & Production Implementation)

- **Architecture Optimization**
  - Converted Python scripts to production-grade **Dataiku** workflows
  - Prioritized **SQL** and native Dataiku components for performance
  - Designed efficient Analytics Base Table (ABT) creation process
- **Model Enhancement**
  - Validated **MLForecast** function effectiveness
  - Tested additional predictive variables
  - Evaluated alternative modeling approaches
  - Confirmed solution robustness through comparative analysis

March 2021 – February 2023 - **Kiota Intelligence (Spain, Remote) - Data Scientist**

Building Data Science solutions for the VC industry. All projects and productions below were built with R Language

- **Predictive Models**
  - Startup survival **prediction**
  - Series A funding probability estimation
  - Pre-valuation modeling
  - Investor-startup **matching** algorithm
- **Interactive Analytics**
  - Built **Shiny dashboard** for funding visualization
  - Implemented dynamic filtering tools
  - Developed multidimensional **outlier** detection system
  - Created temporal **clustering** analysis
- **Automated Reporting**
  - Designed CSS-styled **RMarkdown** reports
  - Implemented automated email distribution system
  - Developed **statistical** testing framework for funding rounds
  - Created interactive filtering system for investor matching

January 2018 – March 2020 - **Tunisia Telecom Group (Tunis, Tunisia) - Data Scientist** at the CVM and Data-Mining department (SQL developer & SAS (Guide/Miner) user)

Customer Segmentation

- **Data Engineering**
  - Constructed **Analytics** Base Table (ABT) from multiple data sources
  - Implemented robust data quality checks and cleaning protocols
  - Executed advanced feature engineering and transformations
- **Advanced Analytics**
  - Performed statistical testing and correlation analysis
  - Developed K-means **clustering** model for behavioral segmentation
  - Created detailed segment profiles for business validation
- **Production Implementation**
  - Designed automated **classification** system for new customer assignment
  - Delivered actionable customer insights for **business strategy**

Multi-SIM User Detection

- Developed and validated **hypotheses** for **dual-SIM** usage patterns
- Engineered comprehensive customer **scoring** algorithm
- Created detailed **user profiles** for target segment

- Implemented scoring model using **SAS Guide** and **Miner**

Family Communities

- Distinguished household-level connections within larger communities from a predefined **communities** result
- Identified extended family **networks** through behavioral patterns

Collaborated with **SAS, KPMG, and Business&Decision experts** in workshop series for different models including :

- **Churn**, **Cross-sell** and Community **Link Analysis** (SAS)

June 2016 – December 2017 - **Tunisia Telecom Group (Tunis, Tunisia) - Data Analyst** at **the CVM and Data-Mining department (SQL developer & SAS (Guide/Miner) user)**

SQL Ad-hoc Request & Dashboard Development

- Customer Value Management (**CVM**) performance
- Network quality **KPIs** & Data service penetration
- Customer segmentation **profiles**
- Sales analytics and **market growth**

Process Automation

- Implemented **VBA** automation for PowerPoint **reportings**
- Established data quality verification protocols

Business Recommendations

- Designed targeted marketing **campaigns**
- Created **Try & Buy** offer frameworks
- Generated data-driven product recommendations

Professional Engagements

February 2021 – July 2025 - **Codementor - Data Science Mentor**

Rated 5 stars by 59 clients for 94 live sessions and projects

February 2022 – December 2022 - **International School of Business (Sfax, Tunisia) - Parttime Statistics Lecturer**

Taught statistics to first and third-year business students

August 2019 – January 2021 - **ZINDI platform for Data Science - Machine Learning Competitor** (R programmer)

- **DigiCow Farmer Training Adoption (Classification), March 2026**
  - Ranked TOP 67% (only 377 succeed to submit among 907 competitors)
  - Predict which farmers will turn training into action
- **agriBORA Commodity Price (Forecasting), December 2025**
  - Ranked TOP 24% (only 354 succeed to submit among 1248 competitors)
  - Forecast weekly market prices for maize in Kenya
- **Zindi New User Engagement (Classification), February 2023**
  - Ranked TOP 70% (only 210 succeed to submit among 1100 competitors)
  - Predict if a new user will be active for 2 months in a row
- **AI4D Predict the Global Spread of COVID-19 (Time Series), April 2020**

- Ranked 4th position (only 47 succeed to submit among 884 competitors)
- Predict the spread of death rates for the coronavirus globally

- **Uber Movement SANRAL Cape Town, January 2020**
  - Ranked TOP 11%, 13th' position (only 113 succeed to submit among 711 competitors)
  - Predict when and where road incidents will occur next in Cape Town
- **AI HACKATHON Tunisia 2019, September 2019**
  - Ranked 6th position (20 continuous hours, only 54 succeed to submit among 191 competitors)
  - Help Tunisian company STEG **detect fraud**
- **Basic Needs Basic Rights Kenya – Tech4MentalHealth (NLP), July 2020**
  - Ranked TOP 36% (only 492 succeed to submit among 900 competitors)
  - **Classify text** from university students in Kenya towards a mental health chatbot
- **Sendy Logistics, November 2019**
  - Ranked TOP 26% (only 431 succeed to submit among 1143 competitors)
  - Predict the estimated time of arrival (**ETA**) for motorbike deliveries in Nairobi
- **Financial Inclusion in Africa, August 2019**
  - Ranked TOP 13% (only 678 succeed to submit among 1288 competitors)
  - Create a ML model to predict which individuals are most likely to have or use a bank account
- **IEEE Big Data Cup, October 2020**
  - Participated in the Challenge "Predicting Escalations in **Customer Support**" and submitting a solution (rank 39/42 who managed to submit)
- **Uber Nairobi Ambulance Perambulation, January 2021**
  - Ranked TOP 31% (only 331 succeed to submit among 1003 competitors)
  - Use ML to create an **optimized** ambulance deployment strategy in Nairobi
- **Wazihub Soil Moisture Prediction, October 2019**
  - Ranked TOP 37% (only 96 succeed to submit among 700 competitors)
  - Predict soil humidity in 5-minute increments using sensor data from low-cost DIY IOT
- **Female-Headed Households in South Africa, February 2020**
  - Ranked TOP 38% (only 258 succeed to submit among 498 competitors)
  - Predict rate of household inferior to certain wage at each ward in South Africa
- **Akeed Restaurant Recommendation in Oman, August 2020**
  - Ranked TOP 45% (only 242 succeed to submit among 1249 competitors)
  - Build a **recommendation** engine to predict what restaurants customers are most likely to order from
- **UNICEF Arm 2030 Vision #1: Flood Prediction in Malawi, May 2020**
  - Ranked TOP 53% (only 477 succeed to submit among 1633 competitors)
  - Predict **flood** extent caused by storms in southern Malawi
- **South African COVID-19 Vulnerability Map, March 2020**
  - Ranked TOP 52% (only 95 succeed to submit among 259 competitors)
  - Infer important COVID-19 public **health** risk factors from outdated data
- **Sea Turtle Rescue, March 2020**
  - Ranked TOP 59% (only 95 succeed to submit among 259 competitors)
  - **Forecast** the of sea turtles Local Ocean Conservation will rescue each week in Kenya

November 2018 – January 2020 - **Freelance - Data Scientist (R programmer)**

Machine Learning Training Development

- Designed comprehensive ML curriculum and hands-on **labs**
- Created Python notebooks covering:
  - Supervised learning algorithms (Linear/Logistic Regression, SVM, Decision Trees)
  - Unsupervised learning (K-means)
  - Model optimization techniques
  - Best practices in ML implementation

NLP Classification System

- Developed scalable **text classification** model
- Implemented automated category prediction pipeline
- Created framework for handling future data classification

IoT Sensor Analytics

- Built high-precision weight prediction model from millisecond-level sensor data

- Optimized error reduction algorithms

Social Media Analytics Platform

- Developed Facebook **API** integration for metrics collection
- Implemented sentiment analysis on user comments
- Delivered actionable **insights** for social media strategy

Professional internships

June 2015 – December 2015 - **LAAS-CNRS (Toulouse, France) - Research Statistician**

Development of an automatic management of a home security system property (for the Smartfox Project with MyFox)

- **Intelligent Access Control System**
  - Developed behavioral **pattern** recognition model for homeowner identification
  - Created real-time intrusion detection algorithm
  - Implemented **automated** alarm management system
- **Technical Implementation**
  - Designed real-time data processing pipeline
  - Built **statistical** learning models for behavior analysis
  - Developed automated decision-making system
- **Publication**
  - Authored research book "Gestion Automatique d'un Système de Sécurisation des Biens à Domicile"
  - Published by European University Editions (**EUE**)

Scientific Training

IBM Watson AI Lab Achievement

- **Competition Winner**
  - Ranked 1st among 100+ participants in IBM **Watson** Services competition
  - Demonstrated expertise in Watson AI platform implementation
- **Instructor Role**
  - Selected to lead subsequent AI labs
  - Trained new participants in Watson Services
  - Facilitated hands-on **AI implementation** workshops

Education

2016 MIT xPRO Data Science: **Data to Insights**

2011-2015 ESSAI Higher School of Statistics and Data Analysis, **Engineering School**

2008-2011 IPEIS Preparatory institute for engineering studies (**Mathematics**- **Physics**), *University-level courses required in preparation for competitive exams and entrance into top graduate and engineering schools*

2008 High School degree in **"Computer Sciences" (First Class Honors)**, 15 November *High school*

Computer Skills

Statistical and Modeling **R**, **Python**, **SAS**-Guide, SAS E-Miner

Databases Management Oracle DB (**Toad**), **MySQL**, MS **SQL** Server

Microsoft Office Good proficiency of the **office** pack (Excel, Word and PowerPoint)

Assessments ::

- **LinkedIn**
  - Machine Learning, August 2020
  - **Tansact-SQL**, February 2020
  - **R**, November 2019

- **Pluralsight**
  - Expert for Bringing Data Science into the Business , February 2021
  - Expert for Feature Engineering , February 2021
  - Expert for Programming R for **Data Analysis** , February 2021
  - Expert for Cleaning data with R , February 2021
  - Expert for **Data Wrangling** with R , February 2021
  - Expert for Microsoft Excel Data Insights , February 2021
  - Expert for Data Science , October 2020
  - Expert for R as a Language , October 2020
  - Expert for **Interpreting** Data with R , October 2020
  - Expert for Machine Learning , September 2020
  - Expert for Data Analytics , April 2020
- **DataCamp**
  - Advanced for Understanding and **Interpreting** Data, September 2020

Certificates & Badges

MIT Professional X

- Data Science : Data to **Insights**, November 2016

Stanford Online

- **Statistical** Learning (using R), March 2020
- Machine Learning by **Andrew Ng**, August 2019
- **Mining** Massive Datasets (**Big-Data Algorithms**), October 2018
- **SQL**, December 2016

DataQuest

- Data Analyst in R **Path**, July 2020

Udacity

- Data Analyst **Track**, April 2019

Coursera

- **Advanced** R Programming (from The Johns Hopkins University), September 2024
- Creating Features for **Time Series** Data (from SAS), August 2024
- Practical Time Series Analysis (from The State University of New York), March 2024
- Data Science Orientation (from IBM), October 2018
- Inferential and Predictive Statistics for Business (from Illinois University), March 2017
- **Text Mining** and Analytics (from Illinois University), March 2017
- **Cluster** Analysis in Data Mining (from Illinois University), February 2017
- Managing Big Data with **MySQL** and TERADATA (from Duke University), May 2016

edX

- Analyzing and Visualizing Data with Excel, May 2017
- Querying with Transact-SQL (from Microsoft), May 2016
- Introduction to Python for Data Science (from Microsoft), April 2016

IBM Badges

- Applied Data Science with R – Level 2, August 2019
- Build your own **Chatbot** – Level 1, November 2018
- Node-Red : Basics to **Bots**, November 2018

SAS Badges

- Predictive Modeling and Text Mining, July 2019
- **Exploratory** Data Analysis, June 2019
- **SAS Programming** 1 : Essentials, November 2018
- Introduction to **Anova**, Regression and Logistic Regression, October 2018

DataCamp Courses

- Forecasting Product **Demand** in R, April 20020
- Time Series with R **Track** (6 courses), March 2020
- **Statistics** with R : Correlation and Linear Regression, May 2016

Languages

Arabic Mother tongue

French Intermediate

English **EFSET 75 C2 Proficient** (February 2023)