KAUST, Saudi Arabia

+966 56 858 8370

abid.med@gmail.com

https://github.com/abid-mohamed

www.linkedin.com/in/med-abid

Mohamed Abid

Data Scientist

PROFILE

Hello, I'm Mohamed Abid, a 44-year-old Tunisian currently residing in Saudi Arabia. I hold an Applied **Msc in Data Science and AI** from the esteemed "**Data ScienceTech Institute**" in France, achieving a First-Class/Summa Cum Laude distinction. Eager to contribute my expertise, I am actively seeking job opportunities in the dynamic fields of Data Science and Artificial Intelligence.

SKILLS

- R
- Python
- SQL
- NetLogo
- MSSQL Server
- JSON
- Neo4j
- MatLab
- Ensemble Model
- Machine Learning
- Time Series Analysis
- Survival Analysis
- Probability / Statistics
- Data Visualization
- Data Cleaning
- Data Manipulation

EDUCATION

10/2023   **Applied Msc in Data Science and AI**

Data ScienceTech Institute, France

Mention: First Class / Summa Cum Laude (Grade 95 | A+)

09/2011   **Bsc in Computer Science Applied to Management**

Faculty of Business and Economics of Sfax, Tunisia

06/2006   **Bsc in Applied Mathematics**

Faculty of Sciences of Sfax, Tunisia

Thesis: "Implementation of multivariate unimodular polynomial matrix completion".

INTERNSHIP

12/2022 - 05/2023 King Abdullah University of Science and Technology (KAUST), Saudi Arabia   **Data Scientist**

GitHub link

Conducted a comprehensive analysis of a raster-based dataset encompassing spatio-temporal information on forest fires in the Amazon from 2001 to 2020. Utilized **R** and key packages such as **terra**, **raster**, and **h2o** for data processing, exploratory data analysis, and ensemble modeling.

Key Contributions:

- **Exploratory Data Analysis:** Conducted exploratory analysis on a high-resolution dataset (500m) with ten variables, capturing factors related to fires, land use, environment, and climate.
- **Class Imbalance and Missing Data:** Addressed class imbalance and missing data challenges, significantly improving model performance.
- **Data Downsampling:** Implemented a down-sampling approach, reducing the dataset to 550 million observations, and dividing it into 11 zones for improved efficiency.
- **Machine Learning Models:** Developed and evaluated machine learning models such as:
    - Distributed Random Forest (**DRF**)
    - Generalized Linear Models (**GLM**)
    - Gradient Boosting Machines (**GBM**)
    - eXtreme Gradient Boosting (**XGB**)
- **Ensemble Modeling:** Created an ensemble model by combining the strengths of individual models within each zone, enhancing predictive accuracy, leveraging **AUC** and **AUCPR** metrics.
- **Visualization:** Visualized the results through dynamic monthly maps and time trend charts, providing insights into fire probabilities in the Amazon over the 20-year period.

Tools: Executed the project using **R** and employed various packages, including **terra**, **raster**, **h2o**, **rsample**, **recipes**, **data.table**, **tidyverse**, **pROC**, **doParallel**, **doSNOW**, **ggplot2**, **tidyterra**.

PROJECTS

Project #1   Electricity Consumption Forecasting   *Time Series Analysis*

GitHub link

Developed predictive models for daily "electricity consumption" in a building based on 47 days of historical data. The dataset included "outdoor air temperature", influencing two distinct approaches:

- **Temperature-Agnostic Model:** Utilized **HoltWinters**, **Auto ARIMA**, **SARIMA**, and **NNET** to forecast electricity consumption, disregarding outdoor temperature. Model selection based on RMSE.
- **Temperature-Informed Model:** Incorporated the impact of "outdoor air temperature" using time series regression models (**SARIMA**, **AutoARIMA**, **NNET**, **VAR**). Selected the best-performing model through RMSE evaluation.

Tools: Implemented in **R** using **forecast**, **keras**, **vars**, **xts**, **ggplot2**, **openxlsx** packages.

Outcome: Applied selected models to forecast "electricity consumption" using the entire dataset, showcasing the impact of temperature information on predictive accuracy.

Project #2   Covariate Impact Analysis and Variable Selection   *Machine Learning*

GitHub link

The objective of this project was to assess the influence of covariate variables on a continuous response variable within two distinct datasets.

Datasets:

- Data1:
    - Comprises 12 observations of the response variable and 2 categorical covariates.
    - Variable Selection Methods: **ANOVA**, **Step-wise Forward**, **Step-wise Backward**, **Lasso**.
    - Applied **lm** function to determine weights for selected variables.
- Data2:
    - Consists of 16 observations of the response variable and 4 continuous covariates.
    - Mitigated small dataset issues using **Bagging** to reduce variance error.
    - Variable Selection Methods: **Adjusted R-squared**, **Step-wise Forward**, **Step-wise Backward**, **Lasso**, Variable Selection Using Random Forests (**VSURF**).
    - Calculated associated errors for each selection method.
    - Identified the **best model by averaging coefficients** from the lowest error procedures.

Tools: Executed the project using **R** and employed various packages, including **glmnet**, **MASS**, **leaps**, **VSURF**, **ggplot2**, **ggpubr**, and **dplyr**.

Outcome: The project resulted in a comprehensive understanding of covariate impacts through rigorous variable selection methods. The utilization of diverse techniques showcased my proficiency in statistical modeling and data analysis using the **R** programming language.

Project #3   Database Synchronization using Python and pyodbc   *Data Wrangling with SQL*

GitHub link

This project involves creating a **Python** script to interact with a database server and retrieve the latest version of a specified table.

The script uses the **pyodbc** package for secure connections to an **MSSQL Server** and communicates with the database by executing queries.

Key Steps:

- **pyodbc Connection:** Establishes a secure connection to the database server using the **pyodbc** package.
- **Query Execution:** Communicates with the database by sending and executing queries to extract the necessary data.
- **Local Comparison:** Compares the database table with the last saved version on the hard drive to detect any changes.
- **Dynamic View Update:** Generates an updated table view only if modifications are identified during the comparison.

This project streamlines the process of keeping a local copy of a specific database table current, enhancing efficiency in data retrieval and utilization.

Project #4   Photon Propagation Simulation   *Agent Base Modeling*

GitHub link | Publication

Developed a simulation model to study **photon behavior in a water tank** and assess the influence of water characteristics on energy propagation.

- **Simulation Features:**
    - Implemented a light source **emitting photons** with random directions, simulating **barrier hits**, **surface reflections**, and **particle collisions.**
    - Explored **four water types** ("Pure-Sea", "Clear-Ocean", "Coastal", and "Turbid-Harbor") with distinct absorption and scattering coefficients.
- **User Interface:**
    - Designed an intuitive interface for water type selection, photon count adjustment, and photo-detector positioning.
    - Real-time visualization provided immediate insights into photon behavior and **total received energy**.
- **Parameter Exploration:** Enabled experimentation through sliders for adjusting parameters like beamwidth and photo-detector position.
- **Tools:** Implemented using **NetLogo** programming language for an interactive and accessible user experience.

PROJECTS

Project #5   **Data Retrieval from MSSQL Server**   *Data Wrangling with SQL*

GitHub link

This project focuses on extracting five types of information from a large **MSSQL Server** database through well-designed SQL queries.

Key Features:

- **Data Extraction Goals:** The project aims to obtain five distinct types of information from the extensive **MSSQL Server** database.
- **Query Techniques:** Standard **SQL** Queries, **Division SQL** Queries, and **Dynamic SQL** Queries with **T-SQL stored procedure**.

Project #6   **Multivariate unimodular polynomial matrix completion**   *Applied Mathematics*

GitHub link

The project focuses on solving a set of multinomial equations utilizing the multivariate unimodular polynomial matrix completion, specifically employing the **Lombardi-Yengui algorithm**. The applications extend to addressing signal processing problems.

- **Algorithmic Approach:** Central to the algorithm is the identification of a "**Gröbner basis**", a crucial step in simplifying the problem and transforming it into a one-variable polynomial solution.
- **Tools:** Implemented the project using **Maple**.

EXPERIENCE

09/2009 - 09/2018 Company "Mhiri Confection", Tunisia   **Accountant**

As an accountant for a workwear clothes manufacturer, I played a pivotal role in optimizing and automating critical processes to enhance efficiency and meet stringent deadlines.

My focus extended beyond traditional accounting responsibilities, as I leveraged programming skills to introduce automation using **MatLab**.

Key Contributions:

- **Invoicing and Payment Management:** Managed invoicing and tracked foreign customer payments efficiently.
- **Automated Document Preparation:**
  - Developed and implemented **MatLab** programs to automate the generation of administrative documents on a weekly basis.
  - Transformed customer information into a structured **MatLab** dataset for seamless data processing.
- **Data Wrangling and Reporting:**
  - Utilized **MatLab** for data wrangling, seamlessly integrating datasets into **Excel** files.
  - Created and updated production planning, providing valuable insights to the management team.
  - Automated the generation of detailed reports for each production order, enhancing communication with the production unit.
- **Export Documentation:** Automated the preparation of invoices and essential financial and customs documents for weekly export operations.

PUBLICATION

- **Mohamed Abid**, Jonatan A. Gonález, Óscar Rodríguez de Rivera, and Paula Moraga "Mapping the Spatio-Temporal Distribution of Fires in the Amazon from 2001 to 2020: An Ensemble Modeling Approach", (submission) Aug, 2023 in **Environmetrics Journal**.

- **Mohamed ABID** (2022, May 29). "ABM for Underwater optical wireless communication in a water tank" (Version 1.0.0). **CoMSES Computational Model Library**. Retrieved from: *https://www.comses.net/codebases/23ce38af-ae87-47bf-b9e3-2523a54fe1a1/releases/1.0.0/*

FORMATION

26/12/2021   Neo4j Certified Professional

Certificate number: 17336311 | **Download link**

03/2019 - 09/2019 Wall Street English of Jeddah, Saudi Arabia

English courses - level 11.

LANGUAGES

Arabic (native), English, French

KAUST, Saudi Arabia

+966 54 565 4898

abid.med@gmail.com

https://github.com/abid-mohamed

www.linkedin.com/in/med-abid

Mohamed Abid

Data Scientist

COVER LETTER

TO

JHONSON HARRY

Director, Web Development Ltd.   June 14, 2019

Dear Sir,

The examples here are as simple and clear as possible.There is an exhaustive list of sample programs provided in the samples folders. The examples here are as simple and clear as possible.There is an exhaustive list of sample programs provided in the samples folders.The examples here are as simple and clear as possible.There is an exhaustive list of sample programs provided in the samples folders.

The examples here are as simple and clear as possible.There is an exhaustive list of sample programs provided in the samples folders. The examples here are as simple and clear as possible.There is an exhaustive list of sample programs provided in the samples folders.The examples here are as simple and clear as possible.There is an exhaustive list of sample programs provided in the samples folders.

The examples here are as simple and clear as possible.There is an exhaustive list of sample programs provided in the samples folders. The examples here are as simple and clear as possible.There is an exhaustive list of sample programs provided in the samples folders.The examples here are as simple and clear as possible.There is an exhaustive list of sample programs provided in the samples folders.

Mohamed ABID

Data Scientist