Mustapha Bouhsen

Data scientist

Data scientist with a strong background in **mathematics** and **computer science**. Competent in the design and maintenance of **pipelines** and analytical systems. Experienced in working with **massive data** sets, applying **statistical** and **machine learning** techniques.

bouhsen.m@gmail.com

Montréal

github.com/mus514

514-603-0115

in linkedin.com/in/mustapha-bouhsen

EDUCATION

Master in Data Science

HEC Montréal

09/2022 - 12/2023

Bachelor in actuarial science

École d'actuariat - Université Laval

09/2029 - 04/2022

EXPERIANCE

Data engineer - intern

Bombardier

09/2023 - Present

Achievements/Tasks

- Automatically migrate (ETL pipeline) files in parquet format to Microsoft Azure using Pyspark, Python and SQL.
- Optimize ETLs that initially took at least 48 hours to reduce it to just 8 minutes, improving business efficiency and productivity.
- Developed ML algorithms predicting plant part prices, employing data analysis, feature engineering, and model optimization.
- Develop Python scripts for extracting data from web service APIs and loading them into databases.
- Developed a robust general-purpose library using Apache Spark to streamline and accelerate the creation of data processing pipelines.
- Implemented pipelines and SQL scripts for tracking and communicating automated migration progress via email.

Data scientist - research

Laval University

04/2022 - 09/2022

Achievements/Tasks

- Data cleaning and replacement of missing values and outelyers using Pandas and Scikit-learn.
- Implementation of ML algorithms such as Radom Forest KNN, Gradient Boosting in Python and R environment to predict the impact of different chemical components on the condition of golf courses.
- Conducting statistical inference to understand the impact of variables on the condition of golf courses such as: Regression Analysis and Hypothesis Testing.

Data scientist - research

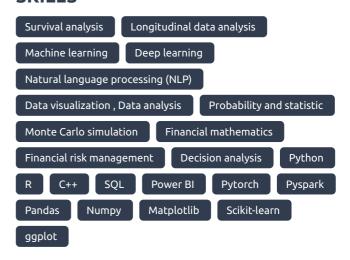
Laval University

04/2021 - 09/2021

Achievements/Tasks

- Analyze data from various optimization projects using R and Python.
- Address missing data and perform data cleaning utilizing Pandas and Scikit-learn.
- Run optimization algorithms on Compute Canada using C++.
- Automate tasks using R and Python and reduce processing time that takes hours in just one click.

SKILLS



PROJECTS

Prediction of the frequency of road accidents in Montreal (05/2023 - 07/2023)

- Provide the city of Montreal with a classification of intersections following intersections of infrastructure based on SAAQ data.
- Design predictive models using several techniques: GLM Poisson, Negative Binomial, Lasso, Gradient Boosting, Ridge and identify relevant variables.

Sentiment analysis from text - using PyTorch-Bert (01/2023 - 04/2023)

- Test different models (LSTM, RNN, Transformer, etc.) and combinations of hyperparameters.
- Test different ways of representing words (Bag of words, TF_IDF, etc.)
- Determine the best performing model using several measures (F-score, accuracy, etc.).

Market risk analysis to manage the exposure of a portfolio (01/2023 - 04/2023)

- Use the Black-Scholes formula to evaluate options.
- Calculate portfolio return using Brownian motion with different volatility, VaR and CVaR using Monte Carlo simulation.
- Risk estimates using the univariate, bivariate Gaussian, student distribution and Gaussian and Student copula.
- Filter the volatility of the logarithmic returns of the underlying using a GARCH model.

Development of a banking services risk analysis model (01/2022 - 04/2022)

- Develop a model to identify customers likely to leave the company.
- Clean data and design models using several techniques: MLP, GLM, Naive Bayes, Gradient Boosting, XGBoost and identify relevant variables.
- Reduce dimension using PCA and hierarchical clustering and K-means.
- Determine the most efficient model (best lift) to facilitate decision-making.

RTC Network Optimization Algorithm (09/2021 - 12/2021)

- Develop an algorithm to find the shortest path to reduce the travel time of RTC lines in Quebec City using C++.
- Simulate trips to calculate average trip times.
- Compare the results obtained with those of Google maps.

LANGUAGES

French - English - Arabic Full Professional Proficiency