**Summary**

* Efficient, team oriented Data Scientist with around 5 years of experience in Data Governance, Data Mining, Machine Learning, and Data Analytics.
* Domain knowledge and experience in Financial, and Digital Processing industries.
* Extensive experience with the entire life cycle of a Data Science project: Data acquisition, cleansing, processing, feature engineering, modeling, evaluation, optimization, testing, and deployment.
* Expertise in Data Governance processes to ensure data availability, usability, quality, consistency, and integrity.
* Highly proficient in ML models and Optimization algorithms such as: Regression, Clustering, Decision Trees, Gradient Boosted Trees, Similarity Matching, SVM, K-NN, Outliers, Neural Networks, Deep Learning, Lasso Regression, Gradient Regression, k-Fold Cross Validation, Error Analysis.
* Strong skills on languages and tools such as: Python, Matlab, R, Jupyter, Excel.
* Extensive experience in Hadoop and Spark ecosystems (Hadoop, Pig, Hive, PySpark).
* Knowledge and experience with the Agile methodology (Scrum, JIRA, Github).
* Familiarity with taking responsibility and ownership for an entire project while effectively communicating all aspects of such process with teammates, managers, and stake holders.

**Experience**

**Data Scientist / Analyst – Deutsche Bank**

November 2018 – PRESENT. Research, build and implement different data governance methods and Machine Learning strategies for next generation investing compliance.

* Currently working on proprietary high-level ML processes to optimize alert detection (Logistic Regression, Similarity Matching, Outliers, Decision Trees and Neural Networks). Specifically, improve the accuracy on True Positive prediction for FX and FI alerts raised by different data sources (In-house, Actimize, etc).
* Research, develop and deploy ML processes that assign and predict a score for all FX and FI alerts (Similarity Matching, Regression, K-NN means).
* Implement different data governance methods to ensure data quality throughout the entire ML process (model development, production modeling, and production storing).

Technology Stack: PySpark, SQL, Python, RapidMiner, JIRA, Jupyter, Excel.

**Data Scientist - Consultant at HoliSym (start-up)**

May 2017 – November 2018 HoliSym is a FinTech all-inclusive asset management platform. It is a whole experience that takes users from learning to engaging, exploring, understanding, simulating, and then trading. Responsibilities: Advise about and produce all the different data governance schemes that have been put in place for the platform. Strategize and implement data collection system for user generated data.

* Implement an algorithm that ensures all incoming financial information’s quality before it’s pushed to simulation/trading models.
* Implement a system that receives all user generated data, validates it, and stores it.
* Study, research, and plan the development of a recommendation feature that allows users to receive information on another users’ performance, tendencies, and interests.

Technology Stack: Python, Hadoop (Pig Latin), Jupyter, Excel, Weka.

**Data Engineer/Scientist at Hi-fi**

March 2015 – March 2017 (2 years) Responsibilities: Study different approaches to improve recommendations system based on Acoustic Analysis on audio content (Data Mining, Machine Learning, Digital Signal Processing). Propose different solutions and present to Hifi team.

* Implement different data governance methods to ensure data quality throughout the entire ML process (model development, production modeling, and production storing).
* Develop a music genre recognition algorithm based on Auditory Sparse Coding method that utilizes an unsupervised clustering based ML model to classify a dataset of over a million songs.
* Device a tempo detection system based on Spectral Autocorrelation algorithm for each song on the entire Hi-Fi catalog of over a million songs. This SA algorithm works on the audio frequency domain and finds ‘energy’ events that form patterns and allow forming an idea of the tempo of the song

Technology Stack: Python, Hadoop, AWS, Excel, Weka, Matlab.

**Data Engineer/Scientist at BRICCON SAS**

August 2012 – December 2014 (2+ year) Responsibilities: Design, build and validate all master databases for the different departments at BRICCON SAS. Design and implement the automation schemes to collect information, pollinate such master databases, and maintain them.

* Validate information contained in all databases. Maintain, supervise and improve data and data quality related processes.
* Extract relevant information from digital copies (.pdf file) of all contracts, which is then organized in a database containing information such as client name, type of contract, monetary value of the contract, etc. Such database is then used as a master database to provide information relevant for other databases in the different departments of the company (finance, budget, quality control, etc).
* Oversee data governance practices in all of the company’s departments (e.g make sure control forms/reports are being created in a way that allows the automated system to collect all information in properly manner).

Technology Stack: Python, Hadoop, Excel.

**Skills**

**Languages:** Python**,** R**,** MySQL, Octave, Matlab.

**Big Data**: Hadoop, PySpark.

**ML Tools**: RapidMiner, Jupyter, Weka**.**

**BI Tools**: MS Excel.

**Statistical Methods**: Hypothesis Testing, Correspondence Analysis, Cross-Validation, PCA, EDA

**Other Tools**: GitHub, JIRA, AWS.