Hugo Michel
DATA SCIENTIST/DATA ENGINEER

✉ hugo.michel@telecom-paris.fr   🌐 https://www.hugomichel.io/   ⚙ https://github.com/hugo-mi

Biography

My scientific background has allowed me to develop a logical way of thinking to convert business requirements into data analysis problems, to provide insights from data for decision support, thus building the bridge between the two worlds.

My recent experiences as a Data Scientist and Data Engineer have allowed me to become familiar with the entire data value chain. I combine my technical expertise in Bayesian statistics, DevOps and MLOps with my knowledge of financial industry processes and functional expertise in financial risk management to drive business outcomes. I am experienced at cracking real world problems end to end, from state of the art model implementation to deployment at scale and impact value measurement.

In addition to being a full-fledged and passionate scientist, I trust human interactions and feel at ease collaborating within a team both for idea generation and delivery.

Education

Télécom Paris   Sep. 2021 – December 2022
Advanced Master in Big Data   *Paris, Ile de France*

Program Highlights

- Massive data processing
- Distributed algorithms
- Statistical Learning
- Data/Text Mining
- Distributed Big Data architecture
- Data visualization
- Machine Learning and Optimization
- AI Ethics
- Deep Learning and NLP
- Econometric

Aalto University   Jan. 2020 – July 2020
Computer Science - Exchange semester in Finland   *Helsinki, Finland*

Program Highlights — GPA: 4,62/5

- Information security
- Cybersecurity
- Networking principles
- Engineering for Humans
- English improvement

University of Technology of Troyes   Sep. 2016 – July 2021
MEng - Computer Science and Information Systems specialized in Data and Knowledge Valuation   *Troyes, Aube*

Program Highlights — GPA: 3,99/5

- Graph theory
- Information System security
- Data Exploratory Analysis
- Usage Centered Design
- BI Architecture
- Web technologies
- Information System Analysis and Modelling
- Symbolic AI

Lycée Fabert   Sep. 2015 – July 2016
Intensive foundation degree in mathematics and physics   *Metz, Moselle*

Experiences

Capgemini Insights & Data   **December 2022 – Present**
Junior Data Engineer   *Paris, Ile de France*
- Pursuing of my end-of-studies internship mission for an international financial institution
- MLOps: Automation, and deployment at scale of various time-consuming operational financial processes:
    - Automating RWA (Risk-Weighted Assets) calculation process
    - Deploying at scale of the customer scoring model in collaboration with data scientists
    - Automating reporting of several accounting documents (balance sheet, ledger, financial audit trail,...)
- Building ETL pipelines to combine third party data from APIs with the datalake, with the objective of data enrichment for the creation of new digital financial product features in compliance with financial regulations
- Led the development of an internal engine for monitoring data quality and monitoring of operational activities (risk, marketing,...) to improve data observability and provide insights:
    - Scoping overhaul of data pipelines and data warehouse
    - Identifying data mart needs to improve data access
    - Conducting user interviews to understand the organization's data pain points to drive data quality monitoring and business KPIs
    - Implementing rules for calculating business KPIs to monitor marketing, accounting and risk activities in order to support business units managers in the decision-making process
    - Implementing rules for calculating KPIs to ensure data quality and data profiling monitoring of the automation engine and industrialization of reports
    - Implementing BI architecture for the orchestration of the calculation of various business KPIs and data quality KPIs
- Led release management and deployment in the production environment and maintaining the CI-CD pipeline

Capgemini Insights & Data   **July 2022 – December 2022**
Data Engineer Intern   *Paris, Ile de France*
- Designing the Data Lineage of the data platform
- Configuring data referential mostly in the cloud and Hadoop environment
- Building data pipelines mainly in the cloud and in hadoop environment to collect, transform and load data in collaboration with data scientists for a global leading banking group
- Management and deployment of releases in production environment thanks to the manipulation of CI-CD pipeline

Télécom Paris x Baalbek Management   **November 2021 – July 2022**
Data Scientist Intern   *Paris, Ile de France*
- Academic project in connection with a company: Research work on NILM (Non Intrusive Load Monitoring) technologies. Designing of an unsupervised end-to-end model for activity detection from the household's aggregated power signal

Aubay   **January 2020 – August 2020**
Machine Learning Engineer Intern   *Paris, Ile de France*
- Designing end-to-end application oriented summarization and automatic generation of French texts. Fine-tuning of state-of-the-art models based on a transformer architecture for the extractive synthesis task and the abstractive synthesis task

AkaBI   **July 2019 – December 2019**
Software Engineer Intern   *Luxembourg*
- Development of an expense report management module for consultants within the company's ERP intranet

Languages

French	English	German
	TOEIC (Level C1)	
Mother tongue	BULATS (Level C1)	Level B1

Projects

BI Architecture | *Bash, Python, PySpark, Cassandra, Zeppelin, Spark*   **June 2022**
- Designing a fault resilient Big Data architecture for distributed storage and distributed query execution based on the GDELT database

Econometrics | *Python, Statsmodel*   **May 2022**
- Significance test, heteroscedasticity test, equality of coefficients test, joint hypothesis test, structural change test, Chow test, Durbin-Watson test

Data Challenge IDEMIA: Face Recognition | *Python, Tensorflow, Scikit-Learn*   **April 2022**
- Designing of a model to predict whether or not two photos correspond to the same identity. Special attention is given to the fairness of the prediction model according to the gender of the person

ELK Stack | *ElasticSearch, Kibana, Logstash*   **March 2022**
- Automating collection, storage and visualization of a network log flow
- Analysis of large data sets from network logs

Designing distributed storage | *Bash, XML, pySpark, HDFS, Hive, Spark*   **January 2022**
- Implementing of a fault resilient distributed architecture
- Configuring software bricks: HDFS, YARN, Zookeeper, Hbase, Hive, Spark
- Using HDFS to store large files
- Using Hive to query large databases
- Using Spark for massive data pre-processing

Several data analysis projects (Vendée Globle, Road accidents) | *Python, Scikit-Learn, bs4*   **September 2021**

Bank loan risk assessment | *R, OWL, SPARQL, Protege*   **September 2020**
- Financial data mining, on a set of documents and a dataset about bank loans
- Text mining: term graph, ontology tree, ontology graph
- Data Mining: Application of classification algorithms (decision tree, k-NN, Apriori, Scoring,...)

Customer data management | *R, Jasper ETL*   **September 2020**
- Building data pipelines to collect and transform data for customer management thanks to an ETL tool

Instant messaging | *Python*   **February 2020**
- Developing instant chat application in python
- Developing features to send private messages
- Developing features to create a chat room for up to 10 people

Technical skills

Programming languages: Python, R, Java, Bash, HTML/CSS, JavaScript, SQL, LaTeX

Data Viz: Power BI, Jasper Report, Tableau, Seaborn, Matplotlib, Plotly, D3.js

Data Engineering: Hadoop Ecosystem (HDFS, Hbase, Hive), ElasticSearch, MongoDB/Cassandra, Neo4J, Microsoft SQL Server, pySpark, Talend ETL, ODI, Sqoop

CI-CD DevOps: Jenkins, Docker, Git, Jira, Azure DevOps

Data Science/Framework: NumPy, Pandas, Beautiful Soup, Scikit-Learn, Statsmodel, Keras, Tensorflow, Spacy, NLTK, HuggingFace API, Flask, ASP .Net Core, SvelteJS

Accomplishments

DataCamp Skill Track : Finance Fundamentals in Python **February 2023**
- Markowitz' portfolios optimization
- Calculating Portofolio Variance
- Maximum draw-down, downside risk, and the sortino ratio
- Manipulating Financial Time Series
- Exonantially weighted risk and return, semicovariance
- Investment Factors : Explain returns and sources of risk

IMT - Institut Mines-Télécom : Fundamentals for Big Data **June 2022**
- **Algebra:** Vector space, linear application, kernel, matrix operations, rank theorem, Singular Value Decomposition
- **Probability:** Conditional probability, Joint probability, Cauchy-Schwartz inequality, Normal and Gaussian law, Gaussian vector
- **Analysis:** Partial derivation, Convex function, Optimization problems, Local minimum, Fermat rule, Weierstrass theorem, KKT conditions, Gradient descent algorithm
- **Statistic:** Empirical variance and Standard deviation, Correlation analysis, Bias-Variance trade-off, Study Gaussian Case, Compute estimator from a sample, Compute empricial risk of an estimator, Hypothesis tests, Regression model for multi-dimensional cases
- **Binary Classification problem:** Properties of the empirical risk, Determine the separating hyperplane, Use a Perceptron algorithm in a classification problem

Readings

CFA Level 1 - Schweser Notes
- Fixed Income
- Derivatives
- Alternative Investments