MONTPELLIER

Data Engineering

#RESPONSIBLEDATASCIENCE
#DATAANALYTICS
#ARTIFICIALINTELLIGENCE
#BIGDATA
#COLLABORATIVEDEVELOPMENT

epf

ENGINEERING SCHOOL
Creating the future together

PROGRAM AIMS

The aim of the Data Engineering major is to train flexible and adaptable engineers both strong in Mathematics & Computer Sciences. They will be able **to help companies and laboratories to structure their data** and to **produce new insights** with Data analysis and Machine Learning approaches. Emphasis is placed on a systemic approach (cost/benefit) including **legal, human, economic and environmental aspects**.

Combining their **generalist education** with **an expertise on all the Data Journey**, the graduates will be able to address all the missions of the **Data Analyst**: data recovery, structuration, analysis & reporting, in direct interaction with the enduser but will also have a strong suit in **Data Science** allowing them to create models and advanced tools (predicter, classifier, etc.) and sufficient knowledge on the **Data Engineering** side to collaborate on the actual transformation of those models into services.

With a pedagogical approach **based on skills, learning-by-doing** and **life-long learning**, the graduates should be able to integrate small and big structures and to adapt to many different businesses.

PROGRAM STRUCTURE

The major extends over two academic years and is organised around two inclass semesters, framed by two internship semesters. (Note: for the international students, the first internship is replaced by an International Project semester which includes mechanics, energy, computer science and French.)

All the CU are offered **in English**. They are designed as independent credits so as to admit students from other programs or students attending vocational training.

In order to be as close as possible to employment conditions, the Major's project CU uses a project approach, thus confronting students to a real client specifications, teamwork and autonomy.

YEAR 4 . COMPULSORY CUS

Computer Science 101 🇬🇧 — 54 hours I 4 ECTS

· Web Programming
· Professional Programming

Hardware & software basics for future computer science's engineers.

Data architecture 🇬🇧 — 75 hours I 6 ECTS

· Infrastructure (OS & Virtualization)
· Data Models (Relational & Non-relational)
· Data Processing Development
  (Parallelization, Distributed computing)

Data architecture from client's need to physical. implementation and considerations on how to process large quantities of data.

Exploratory Data Analysis 🇬🇧 — 75 hours I 6 ECTS

· Basics of Probability & Statistics
· Data Manipulation & Visualisation (Pandas & co)

The mathematical and software tools to explore the data and produce insight for the clients.

Basics of Machine Learning 🇬🇧 — 69 hours I 5 ECTS

· Maths for Machine Learning
  (Algebra, Numerical optimization...)
· Data Cleaning
· Introduction to Machine Learning

The maths behind the magic. What are the types of problem that can be solved by Machine Learning. Starting to teach the computer how to play with data.

Support Digital Transformation 🇬🇧 — 54 hours I 4 ECTS

· Digital Transformation
· Change management
· Professional english

See the big picture of Digital Transformation and how to accompany it.

Major s project 🇬🇧 — 150 hours I 5 ECTS

· Project management · Learning-by-doing semester's project

Theory and practical application of AGILE methodology.

* One of the two ELECTIVE CUS must be chosen. Opening of the elective is subject to a minimal number of students.

YEAR 5 . COMPULSORY CUS

Data diversity 🇬🇧   75 hours I 6 ECTS

· Data Sources (IoT, Surveys, Digital traces & API)
· Geographical data (GIS & Remote sensing)

An exhaustive panorama of the sources of Data and a focus on less common ones: maps & satellite imagery.

Machine Learning: Theory & Practice 🇬🇧   75 hours I 6 ECTS

· Predictive modeling & clustering
· Times Series Analysis
· From PoC to production

Everything you always wanted to know about ML and its implementation in the real world.

Responsible Data Science 🇬🇧   57 hours I 4 ECTS

· Data Ethics
· Data Law
· Ecological impact of IT

Real world issues: take a step back from technology.

Data Strategy 🇬🇧   54 hours I 4 ECTS

· Business Intelligence
· Data Governance
· Data Protection (Cybersecurity, Block-chain)

How the company deals with the new datacentric paradigm and how the future Data engineer can contribute to it.

Major's project 🇬🇧   150 hours I 5 ECTS

· Project Management
· Professionalization
· Project monitoring & final presentation

Team project on behalf of a client.

YEAR 5 . ELECTIVE CUS*

Toward Machine Learning Engineer   63 hours I 5 ECTS

· Foundation of Machine Learning
· Dimensionnality Reduction
· Natural Language Processing
· Computer Vision with ML

From PAC Learning to the manipulation of unstructured data sources (Text, Images), everything you need to sharpen your expertise on Machine Learning.

Toward Data Engineer 🇬🇧   63 hours I 5 ECTS

· Datalakes & Pipelines
· Cloud-based production

How to build your own Datalake, on premise or in the Cloud, a good springboard toward the Data Engineer's jobs.

* One of the two ELECTIVE CUS must be chosen. Opening of the elective is subject to a minimal number of students.

YOUR FUTURE AFTER THE DATA ENGINEERING MAJOR

The array of business sectors of interest is extremely wide and concerns large industrial groups as much as SME/SMIs, research centers or start-ups.

TARGETED POSITIONS
• Data Analyst
• ML Engineer
• Data Engineer
• Data Quality Manager
And many other positions in the IT and Data sector.

FIELDS
• Companies in the digital industry
• Insurance and health companies
• Banks/ Financial industry
• Sales, distribution/ Marketing
• Medical/ pharmaceutical industry
• Energy
• Communal services
• Industry Transport industry
  Life sciences Natural Sciences
  Engineering Journalism

PROJECTS

A project is carried out during both academic semesters in collaboration with a company. **It is used as a guideline for the whole semester and serves as support to many lectures.**

Some examples of projects conducted in the major:

▸ **4TH YEAR**
• Development of a **web tool** allowing the co-design between volunteers and researchers of a common **OLAP-cube representation of a biodiversity** database.
• **Data structuration** (model & tools) of a **financial** analysis process in renewable energy sector

▸ **5TH YEAR**
• Use of **Natural Language Processing** techniques for the **analysis and clustering** of a large **medical's** survey database.
• Development of **Deep Learning** tools for **Object detection** in aerial imagery applied on **ecological study** of cetaceans.
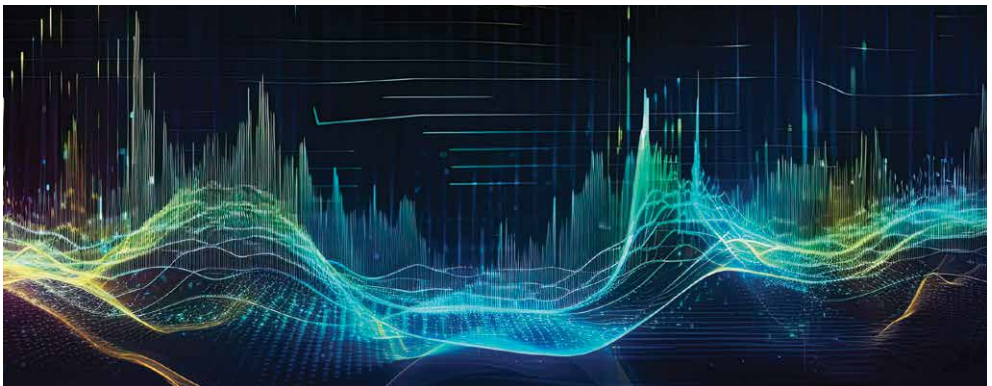
Jade KREISBERGER

2021 School Year

❝ EPF's generalist course allowed me to discover the different fields of engineering. Choosing the right course can be difficult. I have always admired people who develop computer tools, but despite my curiosity, I did not know if this area would suit me. **During my first 3 years at EPF in Montpellier, I took my first steps in programming, development and data analysis.** We had specific plans and guidelines to follow while retaining a certain freedom that allowed us to express our individuality. We created games and tools for data management from scratch. This experience encouraged me to specialize in data engineering. As part of this major, we studied many different subjects: data analysis, machine learning, big data and ecodesign. Everything we did was amazing; from a simple survey, it was possible to extract conclusive information!

For example, my 4th year project consisted of carrying out a study on companies that use connected objects in their work. We were able to evaluate the sectors that utilized them the most and assess why not all companies were able to do so. The major in data engineering and the semester project allowed me to discover IT project management, which combines team management and work progress. I was able to build on skills that I had acquired during my internship at Bolloré Logistics in Singapore.

I had the opportunity to develop a tool for estimating the cost and CO_2 emissions of air and maritime transport for customers in order to reduce the number of company e-mail exchanges. **Yes, IT can be useful for our planet!** When I was in 5th year, I undertook a professionalization contract at Accenture, within the Artificial Intelligence/Data Department where I assisted a data architect. I no longer have any doubts that I belong in this exciting field because I participate in projects with a sense of meaning! **❞**

ANY QUESTIONS?

Antoine GADEMER

Robert RAPADAMNABA

majeure-data@epf.fr