



Data science

supporting industrial performance

Good understanding and use of data are excellent levers for performance and innovation in industry, whether to anticipate failures, reduce outage time, improve quality, reduce execution time, modelling empirical knowledge, etc. And yet, industrial firms are lagging behind in the use of data science, due to a lack of internal expertise and poor structuring of their data.



Mehdi Brahimi

Lead data scientist

PhD Engineer in Philosophy, automatic and IT signal processing, Mehdi has a strong experience in predictive maintenance, artificial intelligence and machine learning.



Data science is a discipline at the crossroads of information technology, client business activities and mathematical modelling. It aims at building statistical and machine learning models using big data, IoT and raw computing power. The purpose is therefore to explore and analyse raw data (whether structured or not) and transform them into relevant information and knowledge to resolve a business-related issue.

For example, in predictive maintenance, you need first to know how the equipment operates (its operating modes and main failure modes) in order to select the most appropriate sensors. Drawing a parallel with medicine, this is the patient examination phase, which aims to isolate the symptoms of an illness. A data scientist models the behaviour of the system using monitoring data (what is the state of health of my system?) and extracts useful knowledge from it using mathematical models or intelligence models (which component is not showing a nominal state of operation, and how will it evolve over time? This enables us to determine more accurately when it needs to be replaced before it fails and causes the infrastructure unavailability).

“Data science is at the crossroads of information technology, client business activities and mathematical modelling. It aims at building statistical models and machine learning models.”

Four types of data to structure and analyse

In today's industrial sectors, amongst others, there are four types of data. **Time-series**: mainly produced by sensors, they include weather data, financial data (performance regarding a site or a project) etc.; **2 or 3-dimensional data**: images, videos, digital mock-ups; **“text” documents** and **operational data**: these are structured (or formatted/data value types) and stored in a database table (ERP), such as production logs or part quality logs.

These four types of data correspond to the issues of the industrial sector, with four solutions.

Predictive maintenance

Regarding the issues encountered with time-series data, they can be resolved through predictive maintenance services. The aim is to be just-in-time, i.e. replacing parts or performing maintenance tasks at the right time before they fail. *“This requires the introduction of a monitoring system using IoT solutions enabling to extract indicators about the health of the equipment, in order to detect any anomalies. Then, we can implement an equipment diagnostic procedure (where does the anomaly occur, on which component, how often and why?) in order to identify the failure mode. Consequently, we can predict the evolution of how the system will degrade or better still, forecast it accurately.”*



Unlike prediction, forecasting is non-deterministic: we make a forecast over time with a degree of uncertainty. *“For example, I installed a sensor on my system (which reports its state of health over time). From this, I will be able to say that the bearing on this system will reach a critical state within the next three months, give or take two days. This real-life example shows that to achieve this, we must have good knowledge of information technology and understanding of the client business and modelling.”*

To implement a predictive maintenance approach, Assystem uses the “PHM” (Prognostics and Health Management) approach, which provides the tools required for the predictive maintenance modules (for data capture, anomaly detection, failure diagnostic, failure forecast, decision-making and applicable actions). *“For example, we had to monitor the volume of water taken into a nuclear plant at the ocean water inlet. The water passes through grids which can be obstructed by algae or waste carried by the seawater. This is a problem that can severely reduce the water flow into the plant’s primary cooling circuit. In this specific case, we developed the PHM approach to focus on the critical systems or components that will generate failures and outages more frequently.”*

This method enabled us to detect anomalies, which together with the analysis of the maintenance history of the grids, led to the implementation of a reliable diagnostic approach. We also identified a correlation between meteorological data and operational data. Depending on the tide, we can now recommend operating modes for the grid cleaning system and extend

“The regulated industries sector generates large volumes of documents and texts. «Document to Data» conversion is the first step in data structuring and demand for Assystem expertise is high.”

consequently the system’s life cycle, which generates significant time and money savings.

Text mining and Natural Language Processing (NLP): the “Google” of industrial systems

Assystem’s “Document to Data” capabilities are in high demand. *“We meet these requirements with Text Mining solutions to retrieve information from documents and from NLP (Natural Language Processing) to understand the meaning or underlying theme of the text.”*

The industrial sector, and more specifically the regulated sectors (nuclear, transportation, pharmaceuticals), generates large volumes of documents and texts (incident reports, requirements,

design specifications, project deliverables, etc.), so we have created sector-specific search engines to make searches more efficient. The possible applications are extensive. For example, we worked on railway station audits using a variety of documents (Word, Excel, e-mails, PowerPoint, PDF, drawings, images etc.) supplied by SNCF, the French National Railway Company. Then our system automatically indexed the documents by category, depending on the business issue encountered.

In the case of nuclear plant decommissioning, our client provided over 50 years’ worth of documents related to the lifetime of the plant. This represented thousands of documents to be scanned and analysed by our decommissioning experts. *“In this specific case, we developed an original method for collecting, processing and analysing this corpus*



of heterogeneous data. The method hinged around three solutions.”

Firstly, a search engine that integrated the ontology of nuclear decommissioning, enabling the documents to be indexed according to the cycle of decommissioning activities. Then an OCR (Optical Character Recognition) module used deep learning techniques to automatically extract all tables from PDF documents. Lastly, a NER (Name Entity Recognition) module was used to recognize names related to the systems and subsystems in these tables. *“In this way, we rebuilt the description of the facility and its architecture, verified all incidents linked to these systems and retraced their history.”*

The search engine developed is based on an artificial intelligence algorithm developed by Facebook (Dr. Q&A) which responds to open questions using a documentary corpus (Wikipedia). *“We decided to use this Open Source solution and adapt it to French on Wikipedia. Then we applied it to the nuclear industry and more specifically decommissioning, using a chatbot. We use it internally, as do our clients, to ensure greater efficiency when an infrastructure is decommissioned. Using this solution, we can achieve around 80 % reliability in responses for the time being, which raises our efficiency in project implementation on behalf of our clients.”*

Image and video processing

In terms of images or videos in video surveillance, deep learning approaches enable us to detect intrusions on sensitive sites (nuclear, defence and related sites, industrial sites, railway stations, dams, etc.). For example, on some of these facilities, the current systems are affected by many false alarms due to diagnostic errors (e.g. confusion between an animal and a person entering a site). *“Using deep learning approaches, we can reduce the rate of false alarms, and support decision-making for security teams of client sites.”*

Analysis of operational data

For structured data (quality, production, schedule etc.), we use solutions from the field of operational research. Assystem models a process in order to optimise it. *“We work on these matters in partnership with*

Cosmo Tech to analyse the execution and sequencing of various tasks over time in order to improve scheduling and for example optimise unit shutdowns on a nuclear power plant upstream, in light of the local human resources available. This enables us to anticipate delays on certain operations, and lower the associated costs.”

We can also use the Root Cause Analysis (RCA) approach to identify the origin of quality-related or production issues, mainly in the nuclear, transport, manufacturing and process sectors. For example, a medicine manufacturer experienced a difference in diameter on tablets in the same production run. We analysed all the data available from production to identify the causal link between quality and tablet diameter based on the parameters of the system set in place. *“We identified the cause and consequently introduced a mathematical model including humidity and temperature to predict*

“The use of artificial intelligence and computing power have drastically altered our work methods and the execution of industrial projects.”



the tablet diameter and thereby ensure a more effective configuration of the production system.”

Data structuring is underway in the industrial sector

The use of artificial intelligence and computing power have drastically altered our work methods and the execution of industrial projects. Nonetheless, access to structured data remains scarce. Yet data structuring is the first step to set up when getting to grips with data science. This structuring step also requires strong understanding of the client's business activities and the creation of algorithms to process the mass of text data. Knowledge of industrial activities is therefore key in finding a solution suited to each process and specific aspect of the industrial sector.

“Today, we are seeing the spreading of structuring of data. Yet the digital culture of industrial firms still has a long way to go. The mistake in this sector is to believe that the volume of data we have at our disposal is immediately available for use... or believing that we possess data that we don't have.”

It is worth noting that the volume of data is often subject to an increasing amount of false alarms. Therefore, to offset this phenomenon, industrial firms are investing in the digital twin field, which combines physical models, experts knowledge and data to create a dynamic in-vivo model of a piece of equipment or a system which reflects its properties more faithfully. ■

“ The mistake in this sector is to believe that the volume of data we have at our disposal is immediately available for use... or believing that we possess data that we don't have.”