Enedis (Formerly ERDF) is a subsidiary of EDF (Electricité de France) in charge of addressing the country’s needs in electricity, which is done today mainly through nuclear power plants. Enedis is the French main electrical DSO (Distribuition System Operator). With 35 million customers, 95% of the French distribution network, ENEDIS is the largest DSO on the European Continent.

DCbrain is a French start-up incubated at the Telecom ParisTech engineer school, member of the ParisTech & Systematic clusters. Specialized in Artificial Intelligence applied to the digitalization of energy networks, DCbrain provides agile capabilities in terms of flow visualization, monitoring, prediction and simulation to complex network managers. DCbrain was the winner of the “Smart Grid Contest” of Enedis, in the Big Data Field.

Three main functionalities

DCbrain was selected to work on a predictive maintenance project. The idea was to integrate into classical machine learning model a new piece of information: the history of the topology and the premature aging of assets due to electrical impacts. In fact, operational teams had the feeling that an incident on the electrical network would impact and age any asset that would be electrically adjacent to the incident.

Thus, the objectives of the project were as follow:

- Gather within a unique model all the patrimonial references, steering, and flows dataset of the grid in order to reconstruct the history of the topology
- Propagate incidents according to the steering policy in place at the moment the incident took place
- Use that model to identify the root cause leading to these incidents on the cables
DCbrain created a unique technology, the Deep Flow Engine, that has the capacity to modelize under the form of a graph of flows all of the flows circulating on the grid.

- The links matching the elements represent the cables and they carry with them the measured flows at the given time-stamp. They also carry all their physical and electrical characteristics.

- The nodes, elements of the network, carry both their steering and state history.

Such a model therefore allows the users to browse, thanks to computing capabilities, an evolutive representation of a network. It becomes quite simple to calculate, for example, intensity propagation or evaluate the impact of an outage by identifying all the « children nodes » or downstream nodes from that outage.

Our Graph Approach, coupled with our capacity to deal with large masses of data allows to regroup different equipment data bases within a single model. In this case, the physical data-base (equipment/assets), the electrical data-base, the steering and incident history data-base.

**KEY TAKE AWAY :**

- The Deep Flow engine has the ability to easily aggregate different datasets: electrical measures, asset information, steering journals, incidents logs, ...
- It allows to browse very easily, within a time dimension, a complex image of the network
- This approach is able to generate information such as indirect impacts over cables linked to outage incidents (ex: propagation of homopolar faults)
We began with raw data coming from Enedis. We used, cleaned and aggregated the different datasets, following a standard process.

This data treatment process follows 3 main stages:
- **Data crunching**: cleaning, extraction of the meaning imbedded into data lines and formatting
- **Graph mining**: data treatment with the graph technology in order to represent the different layers of the Grid
- **Machine learning**: preparation and set up of self-learning algorithms in order to determine reliability scores for cables

**DATA CRUNCHING / DETAILS**

This first stage is crucial. It consists mainly in gathering the raw data-sets coming out of the DSO’s Information System and transform them into usable material, just as we were converting a text file containing encrypted data into an Excel file. Then, we suppress aberrant or false data lines and we control the overall coherence thanks to dedicated algorithms. Once this set of processes has been done, we can construct a first electrical graph, a first physical graph as well as a set of measurement files and steering files.

Example of data validation: identification of steering modifications by derivative calculations on non-aggregated power measurements of arteries.
Graph mining consists in relying on the Graph structure to merge the different data sources cleaned during data crunching. We are able to reconstruct the missing data regarding premature aging due to an electrical impact. The idea is that, in the event of a direct incident, an over-voltage / over-current runs through the network and prematurely degrades the adjacent cables. The difficulty is that only electrically adjacent cables are impacted by this phenomenon. It is therefore necessary to know the topology of the network as well as the way it was operated so far to propagate correctly this electrical impact along the cables and arteries. Here is an example illustrated by the network driving interface provided by DCbrain to the grid:

KEY TAKE AWAY:

▷ The data cleaning phase is key to the success of a model scoring project
▷ DCbrain has developed proprietary data cleaning tool to assess data usability rapidly and to aggregate heterogenous datasets

KEY TAKE AWAY:

▷ The model needs to be tested with experts in order to be fully operational
The red parts are theses impacted by the incident. Full nodes are closed and transparent ones are open. That is the state of the networks steering at the time of incident. Because the entire steering history since 2012 has been integrated into the dynamic repository in the Deep Flow Engine, it is possible, for each incident, to identify the cables sections receiving current impacts and voltage impacts. It is also possible to know the physical characteristics of the electrical sections affected by the incidents and impacts of propagated current / voltage, since the physical data-base system has also been integrated into the dynamic repository.

**KEY TAKE AWAY:**

- Building a graph vision of the synoptic and its historical evolution is key in identifying premature aging linked to an impact propagation.
- Using Graph data base allows a fast construction of this dynamic repository.

**PREDICTIVE MODEL CREATION**

Once all the data is collected, cleaned, processed, assembled in the transverse and enriched repository, we can use them to train a supervised predictive model.

The purpose of a predictive model of classifier type is to estimate whether an element belongs to x or y category:

Here, the two categories are «incident» and «non-incident», and we try to determine the distance (the score) that separates a stretch from «incident» to «non-incident». We start by giving the model the input columns as well as the output: the incident column. The model can thus discover the correlations between the inputs and the output: it learns to separate the data.

**Each entry column is a dimension**

Then, the model is tested with a new data set of which it does not know the answers and compares its prediction with the known result. We then score the model. If the score is not good, the model and the input data are modified and the first stage is done again until a model is deemed valid and robust. At the end of the day, the users can inject in the model new data so that the model can make its prediction.

**KEY TAKE AWAY:**

- Many models exist: logistic regression, decision trees, neural networks ...
  But much of the work lies in the preparation of input data.
3 RESULTS

This project proved DCbrain’s ability to aggregate and rebuild data in order to feed and improve scoring models. Our vision is that a scoring model of a physical network should be largely fed by «network» data such as the “adjacent impacts” feature or flows. This project proved the relevance of this type of data:

Without the feature

With the feature

These recent approaches to define maintenance plans are now being used to challenge the statistical (or even empirical) methodologies in use.