"OUTAGE FORECAST" – A REAL APPLICATION OF MACHINE LEARNING ON GRID OPERATION MANAGEMENT STRATEGIES

Bernardo ALMEIDA
EDP Distribuição – Portugal
bernardo.almeida@edp.pt

Tiago SOARES
EDP Distribuição – Portugal
tiago.soares@edp.pt

Mário TEIXEIRA COUTO
EDP Distribuição – Portugal
mario.teixeiracouto@edp.pt

Tiago SANTOS
SmartWatt – Portugal
tiago.santos@smartwatt.pt

Isabel PRETO
SmartWatt – Portugal
isabel.preto@smartwatt.pt

Cláudio MONTEIRO
FEUP – Portugal
cdm@fe.up.pt

Gonçalo FARIA
EDP Distribuição – Portugal
goncalo.faria@edp.pt

Ricardo Jorge SANTOS
EDP Distribuição – Portugal
ricardojorge.santos@edp.pt

José FERREIRA PINTO
EDP Distribuição – Portugal
ferreira.pinto@edp.pt

ABSTRACT

The energy sector is under one of the biggest transformations ever. Driven by the current digitalization process being held in all sectors, energy utilities are embracing data analytics and artificial intelligence to face current challenges and create new opportunities. Utilities need to act fast to unlock the true potential of the digital grid. The use of real-time data to monitor and operate grid infrastructure under efforts to improve the efficiency and reliability of energy networks is increasing at a global scale.

A lot of utilities experience challenges associated with extreme weather conditions and its implications in quality of service, so a special focus on storm analytics, reliability, resiliency and outage management is required. With that in mind and being work force management (WFM) an extremely value-driving activity, EDP Distribuição (EDPD), the Portuguese DSO, started investing on new methodologies based on data analytics and machine learning to aid in the decision-making process of operational planning activities like the ones presented in this paper.

This paper presents the newly developed tool that uses machine learning algorithms to predict the number and location of outages on EDPD's high and medium voltage grid based on weather forecast. The paper shows in detail the used methodology and presents the project's first results.

INTRODUCTION

The energy sector is witnessing increased investments in research, development and implementation of analytics technologies, evidenced by the number of energy companies using smart technologies in grid operations [1].

Although the tremendous digital transformation that is now being carried out in every sector, the digitalization of the energy sector began a few years ago. The early adopters' utilities can now leverage from all the stored historical data and make use of modern data analytics algorithms to address the sector challenges towards a more reliable, smart and digital grid.

Current energy systems were designed more than 100 years ago and have been only incrementally modified since. They worked well in the time of vertically integrated, centrally supplied generation models, but this was when efficiency and resilience were less important. Today we need an energy systems that can offer reliability and withstand external threats such as extreme weather conditions (Figure 1). New grid optimization technology offers utilities the opportunity to improve efficiency and reduce costs, while improving quality of service [1].

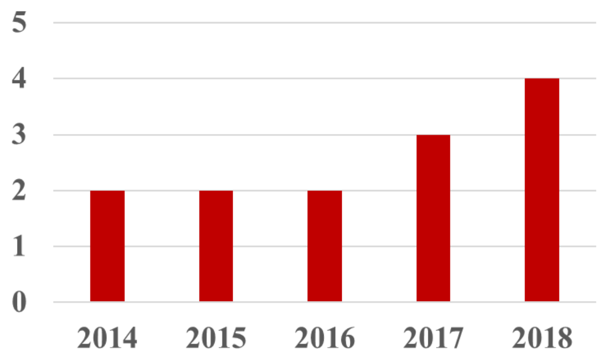

Figure 1 – Number of extreme weather events that affected Portugal mainland (2014-2018)

At EDP Distribuição about 80%, of the more than 80.000 km, of High Voltage (HV) and Medium Voltage (MV) network corresponds to overhead lines. Having an aged mainly overhead network, explains why weather forecast as always been one of the most important inputs in internal work force management strategies. Although this was always an empirical decision, based on historical data and operational experience, weather forecast, at EDPD, was always considered one of the most important factors in the decision-making process of work force allocation on the field and operational planning.

BUILDING KNOW-HOW

There is no doubt there is a correlation between weather conditions and the number of outages in the grid, but does it follow always the same pattern? Is it the same all over the country? What are the most relevant variables for each geographical area or for each feeder? How do each weather variable correlate with the number of outages? Having that in mind, in the last years, EDPD conducted several studies to answer to these questions based on statistical analysis of historical data (Figure 2). The findings helped us to build a better operational actuation plan throughout the years. However, recently, the need for the development of an intelligent tool that predicts the number of outages became mandatory.

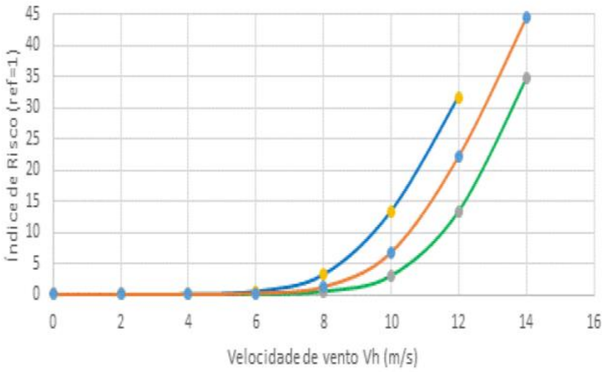

Figure 2 – Correlation between wind speed and the outage risk value (blue – wind speed decreasing in the previous 6h; orange – wind speed constant in the previous 6h; green – wind speed increasing in the previous 6h.

Over the last years, EDP Distribuição has been improved the internal plan for actuation in crisis scenarios. This plan aims to face extreme scenarios adapting organizational model considering the human resources (internal and external service providers), material resources as well as the relation with external stakeholders like Portuguese's weather institute, civil protection, municipalities, fireman, police, telecommunication operators and media. The evolution of this plan is supported in the business continuity strategy, in the experience acquired over the last years under disruptive events and post-event analysis.

Aligned with the digitalization process described before and leveraging from the current state of predictive analytics technics, EDPD established a partnership with SmartWatt, a Portuguese company specialized in developing system optimization solutions for the energy sector. From this partnership resulted the development the outage predictive tool shown in this paper, based on weather forecast, environmental and historical outage data that is currently being used on EDPD to assist the decision-making process of operational planning.

OUTAGE FORECAST TOOL DEVELOPMENT

The operation of the electric distribution grid in a strong disturbed level is featured by a high simultaneity of faults which can occur in large geographic areas or be locally circumscribed. Typically, in such days it is necessary to reinforce the number of human resources in the Dispatch Center for analysing and acting towards the large number of events that arrive at the SCADA and for prioritization of interventions. Consequently, a high number of maintenance teams in the field are required to restore the service as fast as possible.

Previous methods for work force and emergency equipment allocation when severe weather conditions are predicted were based on different weather service providers data and dispatch centre personal experience. However, as these methods have significant flaws, and the frequency of extreme weather events has been increasing, the need for a self-learning, fast and systematic method became urgent.

The trouble is caused by the absence of a proper decision-support tool with the ability to predict the level of faults at the distribution system as a function of the weather forecast. At the decision level, the allocation of additional human resources either in Dispatch Centres or in the field should be based on the predictable level of faults for the next hours/days in order to antedate the possible impact on the distributions system. For instance, when the grid incidents occur in a circumscribed area becomes important to understand the need to mobilize resources from other operational areas. Other key factor that influenced the urgent creation of this type of tool was the increase in number of weather phenomenon with high impact on the Portuguese grid (Figure 3). Indeed, there is a broad consensus that climate change will contribute to increasing the frequency and intensity of extremes weather phenomenon.

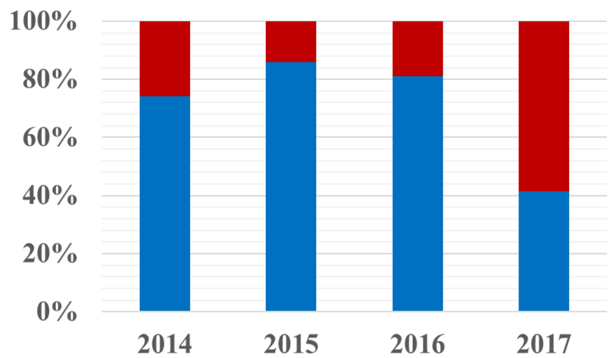

Figure 3 – Impact of extreme weather phenomenon (in red) on EDPD quality of service indicators (MV SAIDI).

The project main goal was to provide the operational areas of EDPD with a tool that could predict the number of outages for each 3h period on the next 3 days on each given area (municipality, operational area, region, etc.) with an associated level of trust.

Data Sets

The first step on the project, before starting to be developed the model, was to select, collect and analyse all relevant data. For the project, the following data sets were used:

- Network data form EDPD's Geographical Information System (GIS) like lines voltage level, length, location of each segment, age, section, type, etc;
- EDPD's Outage Management System (OMS) historical data with cause characterization;
- Terrain characteristics, distance and type of surrounding vegetation and water sources (Figure 4);
- Historical weather forecast data for different points across Portugal (meteorological points) provided by SmartWatt (wind speed, direction and gust, temperature, precipitation, radiation, relative humidity, surface lifted index, convective available potential energy and convective inhibition) – see Figure 5.

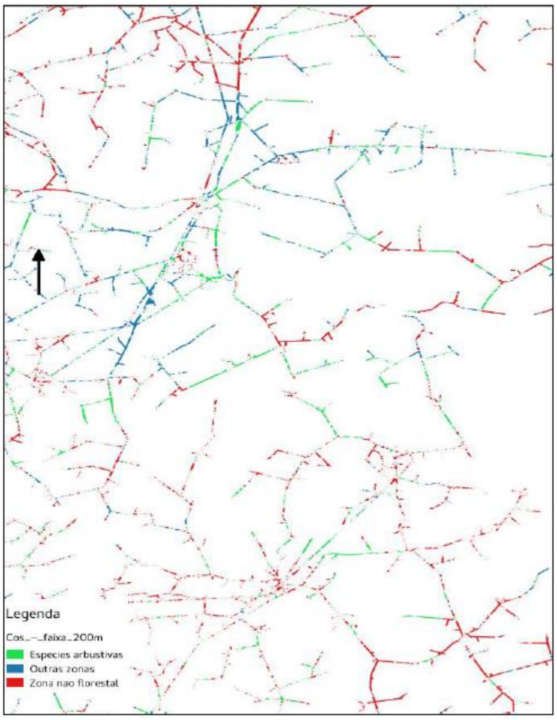

Figure 4 – Detailed example of a rural part of Portuguese network characterization according to vegetation (green – zones with vegetation, red – zone without vegetation, blue – other zones).

Lines Discretization

To create the algorithms, the first step consisted on discretizing each line from the EDPD's high and medium voltage grid. Each line was discretized considering equally-spaced points, with approximately 100 m between them (discretization points), allowing the synchronization between each line and the respective weather forecast information as well as to characterize each line according to terrain characteristics, vegetation and water sources.

Using information from the GIS, it was possible to assess which meteorological point was closer to each discretization point. Therefore, it was assigned a weight to the meteorological points associated to each line, according to the number of closest discretization points. Using these weights, the weather variables for each line were computed as a weighted average of the weather variables of the associated meteorological points. Thus, it was possible to transform each line in just one point containing all the available information about the line.

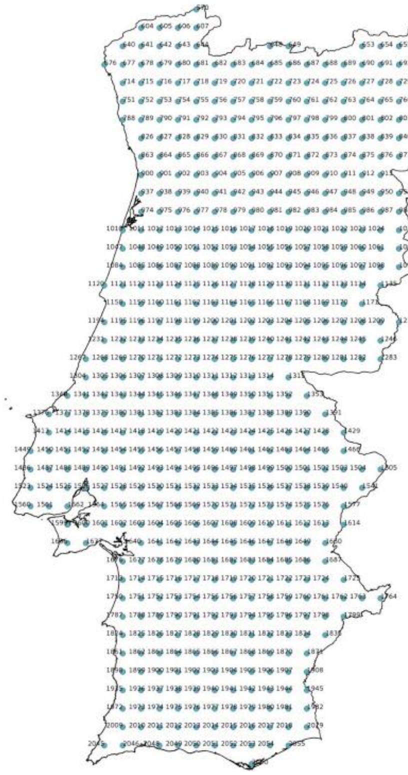

Figure 5 – Representation of Portuguese grid meteorological data points.

This information about all the lines was then matched with the corresponding information about the outages in the grid.

Knowledge Database

For first stage algorithm training a period of 3 years was considered (2018 was only used later). It was analysed the historical data of weather forecast and outages considering each 3-hour interval.

Outages in the grid are rare events, representing only 3% of time when compared with the total time of analysed historical data. This means we would have a very limited data set to create machine learning algorithms. To overcome this issue, a knowledge database was created, using **Bayesian inference** and **kernel density estimation**.

This knowledge database consists on information about weather variables (wind speed, precipitation, atmospheric pressure and cape and convective available potential energy) and the corresponding estimation of a risk index through conditional probabilities (Bayes' theorem) and kernel density estimation. This knowledge database was the basis for risk model creation.

Models Description

Predictive models for risk index (RI) and for the number of outages were developed based on artificial neural networks. Wind, precipitation and thunderstorms events are the most common causes of outages, justifying the creation of a specific risk index model for each variable.

The following predictive models were created, for each EDPD's supervision geographical area (total of 6) and for each geographically operational area (total of 21) – see Figure 6:

1. Risk Index
    a. Global Risk Index
    b. Wind Risk Index
    c. Precipitation Risk Index
    d. Storm Risk Index
2. Number of outages

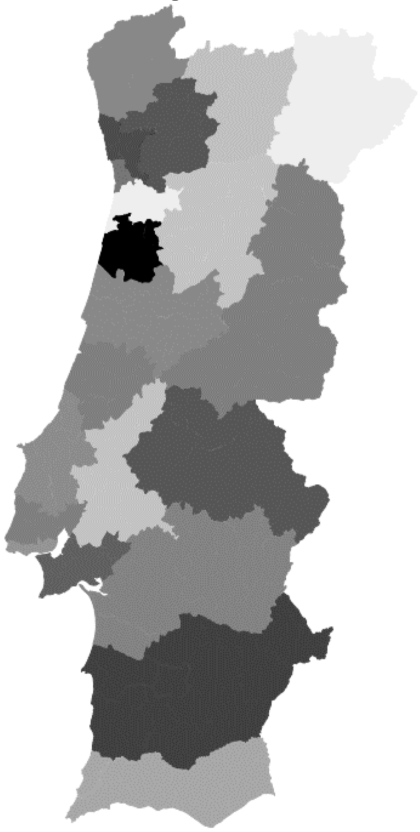

Figure 6 – Portugal divided in the 21 existing operational areas.

Risk Index Model

The risk index model intends to predict the number of times the risk increases, in the analysed period, when compared to a typical day with normal weather conditions. To create the risk index model the knowledge database described before was used as historical data. Data input to the model consisted on the weather variables forecast in 3h periods for the next 3 days and updated every 12 hours. The applied weather variables forecast and its weight varies according to the respective model. Wind speed, precipitation, convective available potential energy and atmospheric pressure were used for the global risk index. For the other models, different combinations of those variables were used.

To calculate a given area risk index, aggregate weighted average needs to be computed taking into consideration all the lines that make part of the respective area. As each line has its own risk index, this means that, more problematic lines are more important when assessing the overall area risk index.

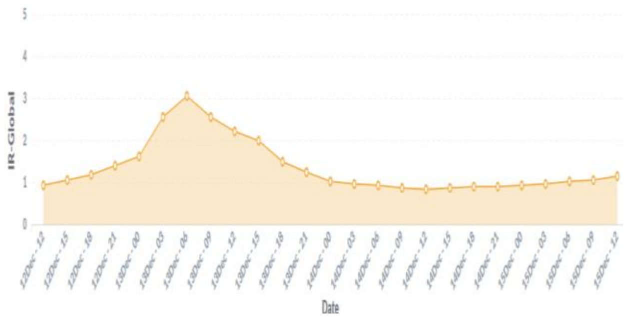

Figure 7 – Risk index output example for 3-day period on a given geographical area.

Number of Outages Model

The number of outages model intends to translate the risk index value in the number of lines that are predicted to fail in each period (only considering long duration outages <3 min). For this model, historical data used consists on the historical forecasted risk index matched with the respective number of faults. Data input consists on the output resulting from the beforementioned five risk index predictive models.

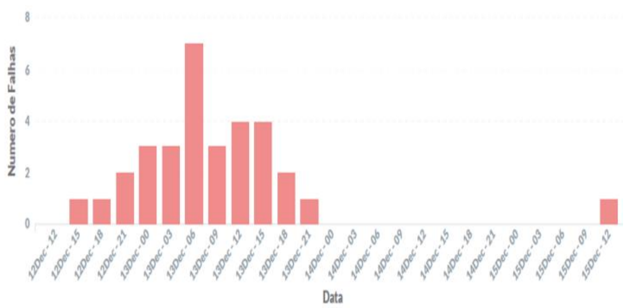

Figure 8 – Predicted number of outages for the same 3-day period and the same geographical area of Figure 7.

Model Output

As said before, the risk index intends to reflect the number of times risk increases when compared to normal behaviour in weather variables. The risk index and number of outages must be carefully analysed according to the respective calculated area. The average risk index is 1, so RI=1 is the reference value, representing a normal situation. If, for example, RI = 2 it represents a double risk of occurring outages in the grid.

FIRST RESULTS AND PERFORMANCE EVALUATION

Although it is too early to have already the final performance evaluation of the tool described in this paper, as most of development was made from middle 2018 forward, the first results are very encouraging.

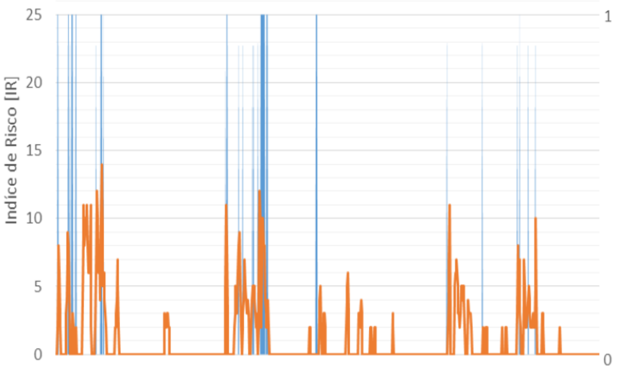

Figure 9 – Risk index (orange) and outage occurrence (blue – binary value) comparison for one operational area.

On Figure 9 for a select evaluation period is shown the correlation between the forecasted risk index (in orange), calculated by the developed model, and the observation of outages occurrence on the same period. From the image is already possible to see that specific operational area above risk index close to 5 the occurrence of outages is almost certain. This kind of risk index sensibility assessment is of enormous importance when work force allocation decisions are necessary to make. If the number is higher than a defined threshold for area mobilization actions must be considered.

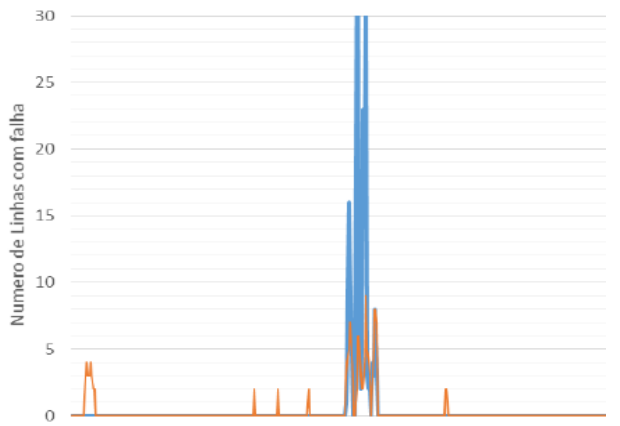

Figure 10 – Predicted number of outages (orange) and registered outages (blue) comparison for one operational area.

On Figure 10 the correlation between the predicted number of outages (in orange) and the number of registered outages is shown. Although there are some false positives one can also observe a big difference between the predicted number and the real outage number. After analysing the results in detail, it was possible to conclude that the difference was because the model output does not consider short duration outages (like the one consequent from successful auto reclose) and in blue all outages are represented.

Table 1 – Result assessment for 3 different areas and the same 4 consecutive periods.

Area	Period	Risk Index	Predicted Outages	Registered Outages
A	1	1,8	3	2
	2	2,59	3	3
	3	2,89	5	1
	4	2,48	3	0
B	1	2,57	2	1
	2	3,66	4	3
	3	3,23	3	4
	4	2,64	1	3
C	1	1,6	2	9
	2	2,4	4	10
	3	2,81	6	5
	4	2,25	4	5

The algorithm is still under performance evaluation and still lack of fine tuning as some assumptions are currently being reviewed. Table 1 present both risk index and predicted long duration outages (from the algorithm output) and shows the registered number of long duration outages. These results show the enormous potential of the developed models as the results only tend to get better as everyday more data is being created to train the algorithm.

CONCLUSIONS

The energy utilities of the future now operate in a new value chain leveraged by digital technologies, where both power and information flow in both directions. The experience of handling large quantities of data has given a new role to the DSO as a data manager but also presented a series of challenges that need to be quickly addressed.

This paper shows how the current state of technology offers a series of opportunities to help face past and future challenges. Machine learning tools like the one present in this paper can help DSO to make more fast, simple and informed decisions for operational planning, with direct impact on quality of service, operational costs and teams response time. The development of these types of tools also assume an important of on internalizing know-how and in continuous improvement of the grid operation and maintenance.

REFERENCES

[1] Ernst & Young, 2016, *Digital grid: powering the future of utilities*, EYGM Limited.