A data-driven platform for predicting the position of future wind turbines

Olivier Parisot

Luxembourg Institute of Science and Technology (LIST)
5 Avenue des Hauts-Fourneaux
4362 Esch-sur-Alzette
Luxembourg
olivier.parisot@list.lu

Abstract. Optimal location of wind turbines is a complex decision problem involving environmental, performance, societal and other parameter. This paper investigates the domain by describing WindturbinesPlanner: by providing machine learning models trained on various data sources, the platform can help to anticipate the potential location of future onshore wind turbines in Luxembourg, France, Belgium and Germany.

Keywords: Wind turbines · Predictive analytics · Visualisation.

1 Introduction

Nowadays, the production of renewable energy is more than ever an ecological and political priority. Therefore, each citizen can observe the installation of new wind turbines everywhere.

To understand this trend from a data-driven approach, we have developed WindturbinesPlanner – a platform to analyse and anticipate the location of potential future wind turbines in Luxembourg, France, Belgium and Germany.

By applying machine learning techniques on heterogeneous data sources, WindturbinesPlanner provides predictions that may be useful for politics and energy actors while considering social acceptance by the public [6]. Weather conditions like *wind speed* are obviously important to explain the location of onshore wind farms. Nevertheless, WindturbinesPlanner could help to understand to what extent other criteria may be important in the covered territories.

The rest of this article is organized as follows. Firstly, related works about wind turbines planning are briefly presented (Section 2). Secondly, the input data sources are detailed (Section 3). Thirdly, a data-driven prediction approach is described (Section 4). Finally, the implementation is presented (Section 5) and the results of preliminary experiments are discussed (Section 6).

2 Related works

As the use of wind energy is a technology that is increasingly used worldwide, the scientific literature on this topic is very abundant.

For example, various computational methods for onshore wind farms placement have been proposed:

- Genetic algorithms have been developed to optimize the positioning of wind turbine in a single area [10, 4].
- Geographic Information System (GIS) have been applied to determine the areas that could be interesting for wind energy development in Northeast Nebraska (USA) [9]; a recent solution consider spatial preferences for offshore/onshore and farms locations in Denmark [7].
- Other works rather focus on the design of algorithms to optimize the layout of wind turbines in dedicated farms [12].

Nevertheless, there is no advanced work about the deducing of wind turbines positions with a purely data-driven approach.

3 Data sources

In order to build a meaningful and exploitable dataset for the prediction of the location of the next onshore wind turbines, different data sources have been aggregated for Luxembourg, Belgium, France, Germany:

- List of the current onshore wind turbines locations (for instance: 943 wind farms listed in France in April 2019) [1].
- Historical time series of daily minimal/maximal/average wind speed values: each time serie corresponds to a geolocated zone (with a width of 7.5km), two years of data have been considered.
- STRM digital elevation model [2]: it may help to check if the topology is concretly considered before installing wind turbines.
- Cities positions and populations: it may help to check the distance between existing wind turbines and town centers, for instance.
- Points of Interests (POI) positions [3]: it may have a direct/indirect impact on wind turbines installation.

Combining these heterogeneous data sources, we have built an aggregated dataset with the following structure:

- Coordinates of the center of the geographical zone (latitude and longitude).
- Average elevation of the geographical zone.
- Average and Maximum wind speed on a recent time period.
- Distance between the geographical zone center and the nearest POI.
- Count of POI in the considered geographical zone.
- Distance between the geographical zone center and the nearest city.
- Population of the nearest city.
- Class to predict: does the geographical zone accommodate wind turbine(s)?

[1] `https://data.open-power-system-data.org/renewable_power_plants/`
[2] `https://fr.wikipedia.org/wiki/Shuttle_Radar_Topography_Mission`
[3] `http://openpoimap.org/`

Additionally, different scales of precision were considered for the width of geographical zones: 7500 / 2500 / 1500 / 500 meters (Fig. 1). To give an overview of the amount of data, considering a width of 500 meters gives a dataset describing 2932088 geographical zones.

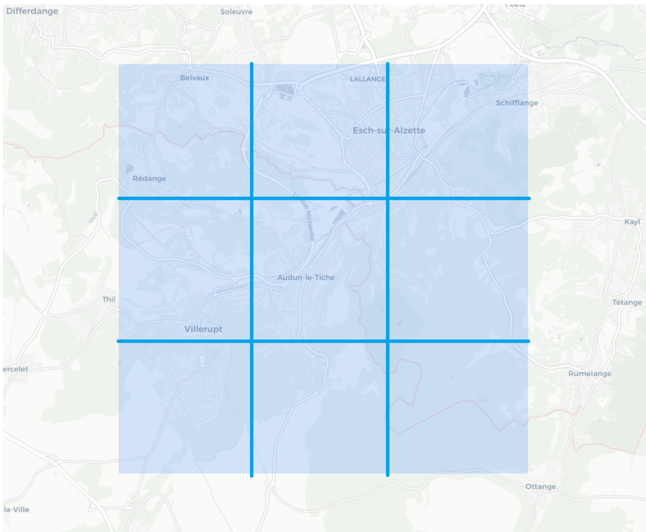

Fig. 1. Considering geographical zones near Audun-Le-Tiche (France): the large square has a width of 7500 meters while the little squares have a width of 2500 meters.

4 Approach

We considered the prediction of wind turbines location as a supervised classification problem with these characteristics:

– The input dataset is class-imbalanced [5] – there are much more zones without wind turbines (this kind of problem can be found in different domains like medical diagnosis or fraud detection).
– False positive should be *encouraged* in order to detect potentially interesting geographical zones to accommodate wind turbines.

Therefore, we have applied state-of-the-art machine learning techniques that are generally effective for class-imabalanced dataset [5]:

– Data Sampling coupled to well-known algorithms like Random Forest [1].
– Ensemble learning methods like Gradient Tree Boost [3].
– One-class Support Vector Machines [8] and One-class Neural Networks [2].

To check the efficiency of the predictive models obtained with these algorithms, we used the classical indicators: accuracy, precision, recall, F1. Additionally, we focused on the *False Discovery Rate* in order to control the proportion of potentially interesting geographical zones for wind turbines.

$$FalseDiscoveryRate = \frac{\text{False Positives Count}}{\text{False Positives Count} + \text{True Positives Count}} \quad (1)$$

5 Implementation

The proposed approach has been implemented into WindturbinesPlanner: a Backend for the computation and a Frontend for the interactive presentation of the results (Fig. 2).

More precisely, the Backend is a set of Javascript command-line tools to retrieve, preprocess and analyze the input data. Thus the predictive models are trained and served through a REST API. These scripts are based on LIMDU [4]: this library provides efficient implementations of the state-or-the-art machine learning algorithms.

The Frontend is a web application built uppon the recent React framework [5]). In order to efficiently show the data (several thousands of points and polygons), we have applied the WebGL technology through the recent DeckGL framework [11]. Thus, when running the web application on a computer with a decent graphics card, the user interface remains reactive whatever the amount of data to display (thanks to the GPU computation).

In practice, a typical usage scenario of WindturbinesPlanner is the following: after selecting an geographical zone on the map, the end-user can investigate to check why an area is potentially favorable or unfavorable to accommodate wind turbines by showing (or hiding): the existing onshore wind turbines, the cities, the points of interests, the weather data and the potential future wind turbines for various scales precisions (from 7500 to 500 meters) and different algorithms.

6 Experiments

Machine learning models have been trained and then integrated into WindturbinesPlanner for various configurations. To this end, the main dataset was splitted into a training dataset and a test dataset (*holdout* strategy).

According to the results (Table 1), we can observe that increasing the geographical precision of the prediction has the effect of greatly increasing the size of the resulting dataset. As a result, it affects the models training time (several minutes for a precision of 2500 meters, much slower for 500 meters).

Moreover, the experiments have shown that is easy to build accurate predictive models from the current datasets. To anticipate the installation of potential

[4] `https://github.com/erelsgl/limdu`

[5] `https://reactjs.org`

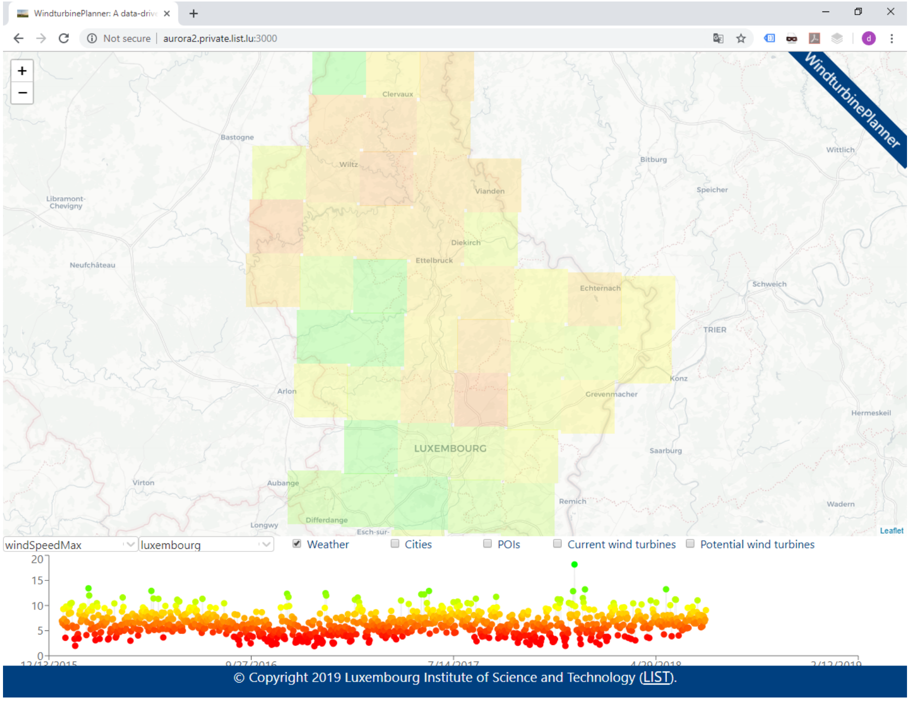

Fig. 2. WindturbinesPlanner provides an interactive map highlighting geographical zones that could contain wind turbines (green color) or not (red color). The predictions can be interpreted by inspecting the input data (wind speed, cities, points of interests).

but not-yet-existing wind turbines, we think it's better to select highly accurate models that produce a *reasonnable* rate of false negative by selecting a geographical zone width (for instance: 1500 meters).

Table 1. Several machine learning models trained to predict if a geographical zone accomodates wind turbine(s). The indicators have been obtained with the test dataset.

Zone width	Dataset size	Algorithm	Accuracy	Precision	Recall	F1	False Positive Rate
7500	13036	XGBoost	0.94	0.68	0.93	0.79	0.31
2500	117278	SVN	0.99	0.98	0.99	0.99	0.01
1500	325786	SVN	0.99	0.84	0.99	0.91	0.15
500	2932088	SVN	0.99	0.27	0.70	0.40	0.72

7 Conclusion and perspectives

In this paper, we presented the WindturbinesPlanner platform in order to anticipate the installation of next onshore wind turbines in a given geographical area. A meaningful dataset was built, machine learning models have been trained and an interactive user interface was developed.

In future work, we will improve the platform to dynamically highlight the evolution over time of wind turbines installation policies carried out by professionals in the energy sector. Moreover, we plan to speed-up the models learning phase by training the model on a High-Performance supercomputer (HPC). This should help refine predictions geographically without sacrificing performance.

Acknowledgments: this work was carried out as part of the FEDER Data Analytics Platform project [6]. Special thanks to Anne Hendrick for her support.

References

1. Bellinger, C., Sharma, S., Japkowicz, N., Zaïane, O.R.: Framework for extreme imbalance classification: Swim—sampling with the majority class. Knowledge and Information Systems pp. 1–26 (2019)
2. Chalapathy, R., Menon, A.K., Chawla, S.: Anomaly detection using one-class neural networks. arXiv preprint arXiv:1802.06360 (2018)
3. Friedman, J.H.: Greedy function approximation: a gradient boosting machine. Annals of statistics pp. 1189–1232 (2001)
4. Grady, S., Hussaini, M., Abdullah, M.M.: Placement of wind turbines using genetic algorithms. Renewable energy **30**(2), 259–270 (2005)
5. Haixiang, G., Yijing, L., Shang, J., Mingyun, G., Yuanyue, H., Bing, G.: Learning from class-imbalanced data: Review of methods and applications. Expert Systems with Applications **73**, 220–239 (2017)
6. Hevia-Koch, P., Ladenburg, J.: Where should wind energy be located? a review of preferences and visualisation approaches for wind turbine locations. Energy Research and Social Science **53**, 23–33 (2019)
7. Ladenburg, J., Hevia-Koch, P., Petrovic, S., Knapp, L.: The offshore-onshore conundrum: Preferences for wind energy considering spatial data in Denmark. Renewable and Sustainable Energy Reviews **121**, 109711 (2020)
8. Li, K.L., Huang, H.K., Tian, S.F., Xu, W.: Improving one-class svm for anomaly detection. In: Proceedings of the 2003 International Conference on Machine Learning and Cybernetics (IEEE Cat. No. 03EX693). vol. 5, pp. 3077–3081. IEEE (2003)
9. Miller, A., Li, R.: A geospatial approach for prioritizing wind farm development in Northeast Nebraska. ISPRS Int. Journ. of Geo-Information **3**(3), 968–979 (2014)
10. Mosetti, G., Poloni, C., Diviacco, B.: Optimization of wind turbine positioning in large windfarms by means of a genetic algorithm. Journal of Wind Engineering and Industrial Aerodynamics **51**(1), 105–116 (1994)
11. Wang, Y.: Deck. gl: Large-scale web-based visual analytics made easy. arXiv preprint arXiv:1910.08865 (2019)
12. Yang, K., Kwak, G., Cho, K., Huh, J.: Wind farm layout optimization for wake effect uniformity. Energy **183**, 983 – 995 (2019)

[6] `http://tiny.cc/feder-dap-project`