ASSIGNMENT

TECHNOLOGY PARK MALAYSIA

CT045-3-M-ABAV

ADVANCED BUSINESS ANALYTICS AND VISUALIZATION

APUMF1908DSBA

HAND OUT DATE: 03 JULY 2020

HAND IN DATE: 26 JULY 2020

WEIGHTAGE: 20%

A Story of Data and Energy: Renewable Energy Generation in Spain

NAME:	Ismail Dawoodjee
TP NUMBER:	TP054033
EMAIL:	tp054033@mail.apu.edu.my
LECTURER:	Dr. Preethi Subramanian

A Story of Data and Energy: Renewable Energy Generation in Spain

Ismail Esack Dawoodjee
Asia Pacific University of Technology and Innovation
tp054033@mail.apu.edu.my

Table of Contents

1. Introduction

According to the Intergovernmental Panel on Climate Change (IPCC), the electric power industry has been estimated to contribute around 25% of annual human-caused greenhouse gas emissions (IPCC, 2014). On the other hand, this industry also generates large amounts of data, making it an ideal sector for the application of artificial intelligence (AI) and machine learning (ML) techniques (Rolnick et al., 2019). Hence, ML has enormous potential to transform the electric power industry into a greener and more efficient system. This includes accelerating the research and development of clean energy sources (such as hydro, solar and wind power generation), improvements in energy demand forecasts, optimizing the supply and distribution of electricity with data-driven "smart grids", and reducing the wastage of electrical energy during transportation (Rolnick et al., 2019).

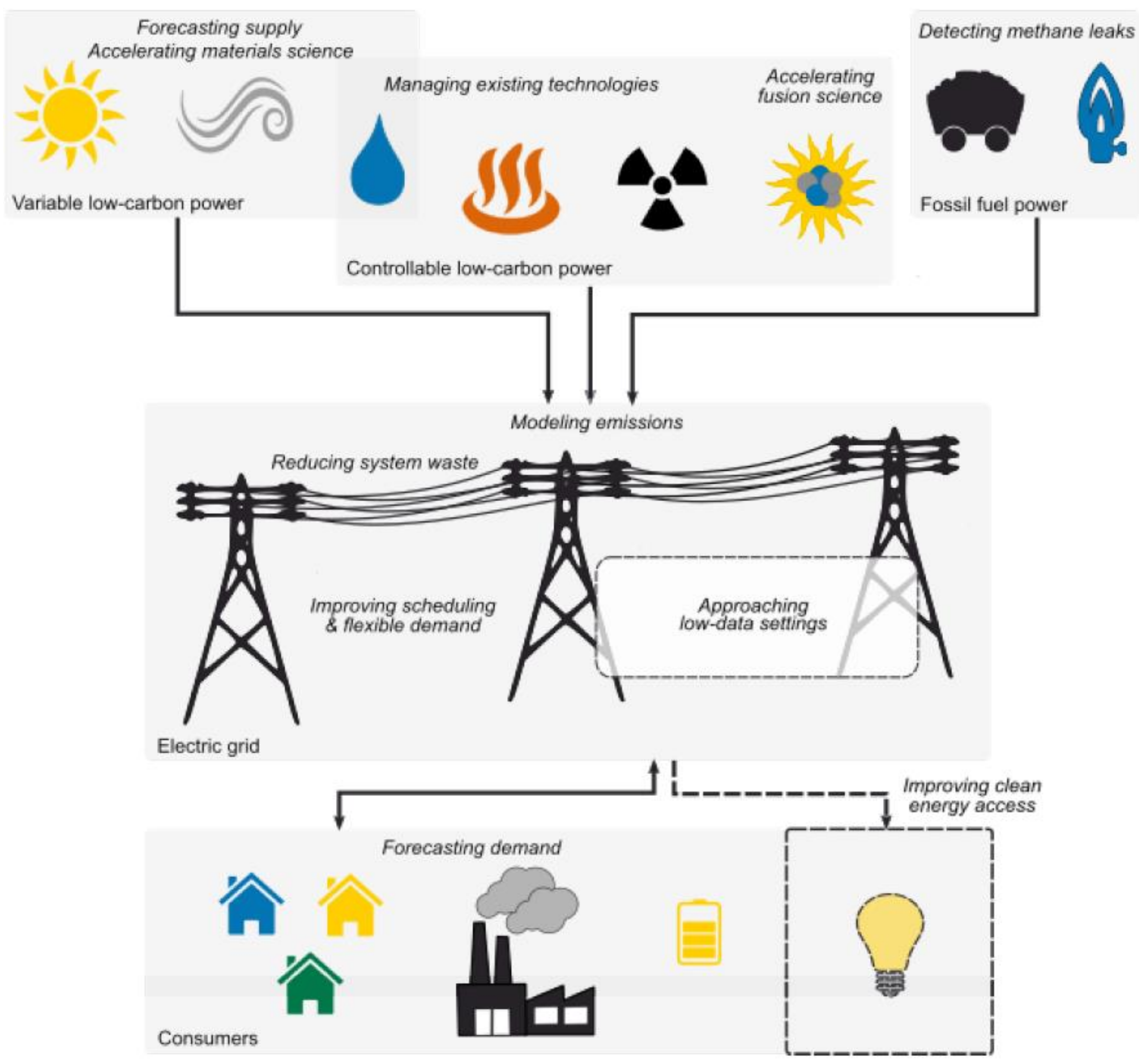

Figure 1: The variety of ways in which machine learning can assist the electric power industry in reducing greenhouse gas emissions (Rolnick et al., 2019).

2. Business Case

For this particular case study, the electric power industry in question comes from Spain, a country well-known for its sunny skies and beautiful coastal beaches. The energy service company, named as Red Eléctrica de España (REE), has gathered four years of electrical demand and price data (from January 2015 to December 2018), as well as the various types of non-renewable (coal, oil, gas, nuclear) and renewable energy generation data.

Figure 2: The land of sun and blue skies (Solucion Asesores XXI, 2016).

2.1. Business Goals

Currently, the company is looking to expand its renewables sector but has not yet decided on which energy source or where the energy generator (e.g. solar farm, wind farm, tidal power station) should be located. To facilitate this decision, weather data from the five largest cities in Spain were extracted from the Open Weather API over the same four year period (OpenWeather Ltd., 2020). This data contains information on each city's temperature, wind speed and direction, rainfall and cloudiness, among others. By analyzing both datasets, REE will be able to make an informed and data-driven decision, not only to figure out the best location to establish a clean energy generator but also to discover how weather patterns influence electricity demand and energy generation.

Justification: Investing in renewable energy resources is an essential stepping-stone to reducing greenhouse gas emissions and slowing down climate change (Rolnick et al., 2019). Spain, as a country surrounded by water on three sides, has already been subject to the ravages of climate change. The average temperature has risen by 8 °C, and overall precipitation received across the country is 25% less than that of 50 years ago (Heggie, 2020). In addition, about 90% of ice glaciers in the Pyrénées (a mountain range separating Spain from France) has completely disappeared (Heggie, 2020). Understandably, Spain has made and is still making outstanding efforts to combat climate change, and REE is contributing to this cause.

2.2. Aim and Objectives

The main aim of REE is to identify the optimum location (out of the five cities) in which a sustainable energy source can be placed to maximize electricity generation. For example, the city of Seville, which is located in the southern part of Spain, enjoys more sunshine than Bilbao, which is located in the northern mountainous region. Hence, constructing a solar farm near Seville could be *one* reasonable choice. Given the above aim, the objectives are:

1. Descriptive Analytics: To analyze, describe, and visualize trends in the energy generation sources, consumer energy demand, electricity price, and fluctuations in weather.
2. Predictive Analytics: To build multiple linear regression models to forecast consumer demand and energy generation, based on the city and weather predictor variables.
3. Prescriptive Analytics: To anticipate and prepare for future surges in demand, reduce the occurrence of surpluses and shortages in electricity, and put forward impactful recommendations for the location of several possible renewable energy generators.

2.3. Scope

The datasets that are used in this study can be obtained directly from Kaggle[1] (Jhana, 2019). Alternatively, the energy generation data can be manually extracted from the ENTSO-E (European Network of Transmission System Operators for Electricity) website (ENTSO-E, 2020), price data from the REE company website (REE, 2020), and the weather dataset can be purchased from the OpenWeather API page (OpenWeather Ltd., 2020).

The scope of this study will be limited to analyzing only two datasets: the energy and weather datasets directly obtained from the Kaggle page. No external datasets will be used, nor will there be an analysis of data that is outside the time period from January 2015 to December 2018. Moreover, the analysis will be limited only to business- and data-related problems that can be solved *before* the renewable energy generator construction takes place. Problems that occur during and after the construction takes place (that may still be solved with ML and AI), such as the logistics of transporting electricity, or the physics and engineering cross-disciplinary issues related with optimizing the generation of electricity, will not be considered. Finally, only weather

[1] https://www.kaggle.com/nicholasjhana/energy-consumption-generation-prices-and-weather

data from the five largest Spanish cities (namely, Madrid, Bilbao, Seville, Barcelona, and Valencia) will be analyzed. Other cities or regions in Spain will not be considered.

3. Methodology

The data mining approach to achieve the aim and objectives described above will be the CRISP-DM (CRoss Industry Standard Process for Data Mining) analytics model. This model includes the vital first step of understanding the business and setting goals for it to accomplish. Henceforth, all subsequent steps and objectives in the case study can be meaningfully retraced back to the original goal that REE has set out to achieve: to expand its renewables sector in the name of slowing down climate change, but also making a good profit from it.

For the first objective, the visualization and data mining tools Tableau and SAS Enterprise Miner will be used for preliminary data exploration, preprocessing and feature engineering. The trend over the years can be visualized through the use of time series line plots, with time aggregation (resampling by day, week or month) and variable aggregation (aggregating over all renewable or non-renewable energy sources). An interactive dashboard linking together some of the most important dependencies between city weather, energy, demand, and price would be an excellent visualization of the hidden patterns that exist within the two datasets.

Next, linear regression forecasting models can be used to forecast for future energy demand and generation, which will be taken as the dependent variables. The reason for using a regression model is because sustainable energy generation is heavily dependent on the weather (e.g. the sun needs to be shining on the photovoltaic cells to generate energy and the wind needs to be blowing at a certain speed for the wind turbines to rotate and produce electricity). Hence, weather variables can be inserted into the regression model as the independent predictor variables, after accounting for their predictive power[2] using the adjusted R^2 value, Akaike's Information Criterion, and other measures of predictive accuracy (Hyndman and Athanasopoulos, 2018).

In addition, knowledge about how consumer energy demand varies with weather variables can be obtained from a similar regression model. For instance, people tend to use heaters on cold days and air conditioners on hot days, but what about on windy or cloudy days? Evidently, the effect of weather on human behavior may not be as predictable as that on mechanical objects. Nevertheless, the regression models will make a brave attempt at modelling both energy demand and energy generation.

Finally, based on the regression models, forecasts can be made for the next 24 hours, 7 days, 4 weeks and 3 months. Although it is an inevitable fact that forecasts become less accurate as the forecast horizon gets larger, the overall trend and seasonality data obtained from the

---

[2] Predictive power: how important an independent variable is in influencing the dependent variable.

predictions will help REE make better decisions in the future. In other words, the company's response to consumer demand can be more effective and less wasteful due to closing the occasional gap between energy supply and demand.

References

ENTSO-E, 2020. *ENTSO-E Transparency Platform.* [Online] Available at: https://transparency.entsoe.eu/dashboard/show [Accessed 16 July 2020].

Heggie, J., 2020. *Spain: taking sustainable energy to the next level.* [Online] Available at: https://www.nationalgeographic.com/science/2020/02/partner-content-setting-standard-for-sustainability/ [Accessed 17 July 2020].

Hyndman, R.J., & Athanasopoulos, G. (2018) *Forecasting: principles and practice*, 2nd edition, OTexts: Melbourne, Australia. OTexts.com/fpp2 [Accessed 17 July 2020].

IPCC, 2014: *Climate Change 2014: Mitigation of Climate Change. Contribution of Working Group III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* [Edenhofer, O., R. Pichs-Madruga, Y. Sokona, E. Farahani, S. Kadner, K. Seyboth, A. Adler, I. Baum, S. Brunner, P. Eickemeier, B. Kriemann, J. Savolainen, S. Schlömer, C. von Stechow, T. Zwickel and J.C. Minx (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Jhana, N., 2019. *Hourly energy demand generation and weather | Kaggle.* [Online] Available at: https://www.kaggle.com/nicholasjhana/energy-consumption-generation-prices-and-weather [Accessed 15 July 2020].

OpenWeather Ltd., 2020. *Weather API - OpenWeatherMap.* [Online] Available at: https://openweathermap.org/api [Accessed 17 July 2020].

REE, 2020. *Markets and prices | ESIOS electricity · data · transparency.* [Online] Available at: https://www.esios.ree.es/en/market-and-prices [Accessed 16 July 2020].

Rolnick, D., Donti, P.L., Kaack, L.H., Kochanski, K., Lacoste, A., Sankaran, K., Ross, A.S., Milojevic-Dupont, N., Jaques, N., Waldman-Brown, A., Luccioni, A., Maharaj, T., Sherwin, E.D., Mukkavilli, S.K., Kording, K.P., Gomes, C., Ng, A.Y., Hassabis, D., Platt, J.C., Creutzig, F., Chayes, J. & Bengio, Y. (2019). Tackling Climate Change with Machine Learning. *CoRR.* [Online]. abs/1906.05433. Available from: http://arxiv.org/abs/1906.05433

Solucion Asesores XXI, 2016. *WHY SPAIN - Solucion Asesores XXI.* [Online] Available at: https://solucionasesoressl.com/en/why-spain [Accessed 16 July 2020].