LITERATURE REVIEW

Tackle climate change with machine learning is motivation behind this project. The problem being considered for the project is to predict or forecast energy demand accurately in Spain. Accurate forecast of energy demand helps in advance planning and dispatching of resources which in turn save environmental cost due to extra production of Green house gases (GHGs) such as carbon dioxide to meet energy demand by buring fossil fuels. There are many opportunities to reduce GHG emission using ML as shown in figure 1 (David et al.).

There are many entities involved in energy market. Most important one are transmission system operator, power plant, commercial and residential users. **A transmission system operator (TSO)https://en.wikipedia.org/wiki/Transmission_system_operator** (wikipedia) is an entity entrusted with transporting energy in the form of natural gas or electrical power on a national or regional level, using fixed infrastructure. The term is defined by the European Commission. As can be seen, forecasting is necessary for ISO or TSO as well as GENCO or power plant. The research questions are reiterated in Table 1.

Table 1: Literature Review Research:

No.	Research Question (RQ)
1	Which regression technique will accurately forecast the daily energy consumption demand using hourly period ?
2	How to accurately forecast energy demand 24 hour in advance compared to TSO?

3	**Using classification, determine what weather measurement and cities influence most the electric demand, prices, and generation capacity?**

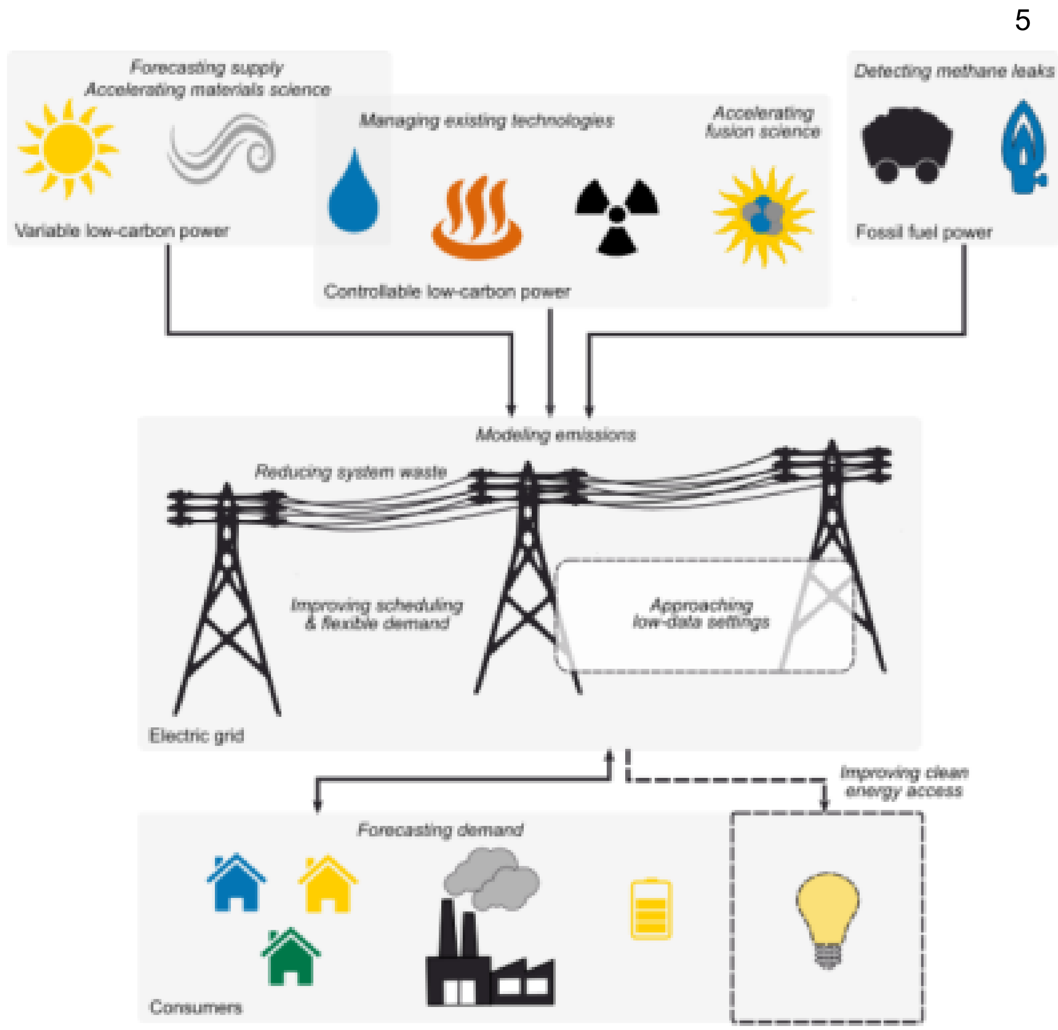

Figure 1: Selected opportunities to reduce GHG emissions from electricity systems using machine learning.

Figure 1: [Selected opportunities to reduce GHG emission using ML]()

The general steps (Anton et al.) followed in review papers is shown in figure 2. As can be seen from the figure, ANN-based model is the most popular model in energy forecasting which is due to its non linearity in input to output matching. ANN is model inspired by human nervous system. ANN regression model comprises of input, weight, error, transfer function, activation function and output (Zakria et al.). The questions to considers for literature review are provided in Table 2. Six papesr are reviewed and key take-aways and how it can be useful to this project is provided below.

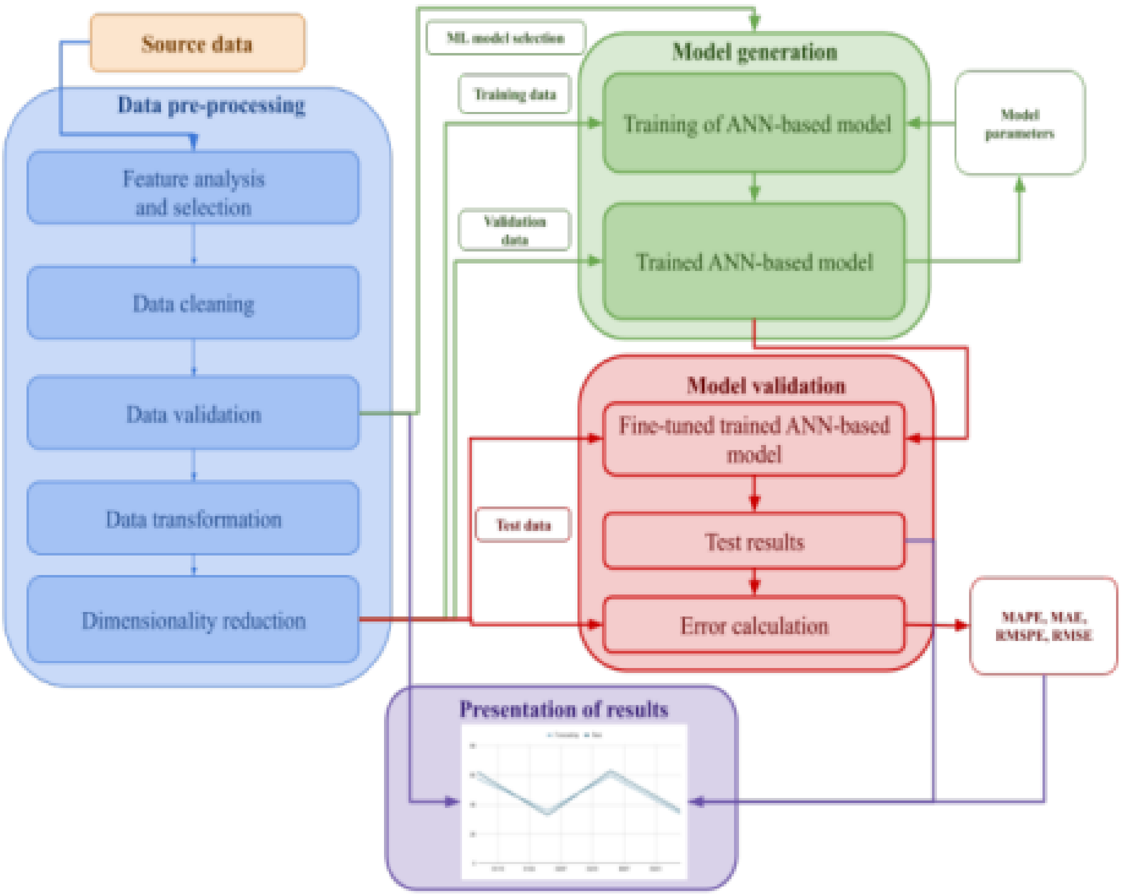

Figure 2: [Generalization of steps documented in review papers]

Table 2: Questions to consider for Literature Review

No.	**Literature Review Question**
1	**What do you already know about the topic?**
2	**What do you have to say critically about what is already known?**
3	**Has anyone else ever done anything exactly the same?**
4	**Has anyone else done anything that is related?**

5	**Where does your work fit in with what has gone before?**
6	**Why is your research worth doing in the light of what has already been done?**

4. DATA DESCRIPTION

Dataset Description

" energy demand generation and weather of spain"is the chosen dataset.

- Dataset has 35064 rows and 29 columns.
- The dataset have 4 years of electrical consumption, generation, pricing and weather data for Spain.
- The data is retrieved from (Entsoe, Esios and Openweather). The links are given in references.
- The dataset have hourly data for electrical consumption and respective forecast by Transmission Service Opersator (TSO) such as Spanish esios Red Electric Espana (REE) for consumption and pricing.
- The attributes type is given in table 1.
- Here chr mean character and num means numeric.
- Out of 29 attributes, 26 are numeric, 3 are non-numeric
- Out of 3 non-numeric attributes, 2 are logical and 1 is character
- Out of 26 numeric attributes, 6 attributes have either zero or missing values. These attributes are provided in table 4. They can be removed from analysis as there is no information regarding these attributes.
- Using python, descriptive Statistics for remaining 20 numeric attributes is provided in Table 3
- For regression analysis, **total.load.forecast** (row 17 in table 5) is selected as

dependent variable.

- - For classification analysis, **forecast.wind.offshore.eday.ahead** is selected as dependent variable
- **Table 1: Attributes Type**

Attribute	Type Attribute	Type
time	chr generation.nuclear	num
generation.biomass	num generation.other	num
generation.fossil.brown.coal.lignite	num generation.other.renewable	num
generation.fossil.coal.derived.gas	num generation.solar	num
generation.fossil.gas generation.fossil.hard.coal	num generation.waste num generation.wind.offshore	num num
generation.fossil.oil generation.fossil.oil.shale generation.fossil.peat generation.geothermal	num generation.wind.onshore num forecast.solar.day.ahead num forecast.wind.offshore.eday.ahead num forecast.wind.onshore.day.ahead	num num logical num
generation.hydro.pumped.storage.ag gregated	logical total.load.forecast	num
generation.hydro.pumped.storage.co nsumption	num total.load.actual	num

generation.hydro.run.of.river.and.po undage	**num price.day.ahead**	**num**
generation.hydro.water.reservoir	**num price.actual**	**num**
price.actual	**num**	

Table 2: Numeric attributes with zero or missing values

Numerical Attribute
generation.fossil.coal.derived.gas
generation.fossil.oil.shale
generation.fossil.peat
generation.geothermal
generation.marine
generation.wind.offshore

Table 3: Descriptive Statistic of Numerical attribute with non-zero values

No.	Arrtibute Mean St dev Min	Max	NA
1	generation.biomass 383.5 85.4 0	592	19
2	generation.fossil.brown.coal.lignite 448.1 354.6 0	999	18
3	generation.fossil.gas 5623 2201.8 0	20034	18
4	generation.fossil.hard.coal 4256 1961.6 0	8359	18
5	generation.fossil.oil 298.3 52.5 0	449	19
6	generation.hydro.pumped.storage. 475.6 792.4 0 consumption	4523	19
7	generation.hydro.run.of.river.and. 972.1 400.8 0 poundage	2000	19
8	generation.hydro.water.reservoir 2605 1835.2 0	9728	18
9	generation.nuclear 6264 839.7 0	7117	17
10	generation.other 60.23 20.2 0	106	18
11	generation.other.renewable 85.64 14.1 0	119	18
12	generation.solar 1433 1680.1 0	5792	18
13	generation.waste 269.5 50.2 0	357	19
14	generation.wind.onshore 5464 3213.7 0	17436	18
15	forecast.solar.day.ahead 1439 1677.703 0	5836	0

16	**forecast.wind.onshore.day.ahead 5471 3176.313 237**	**17430**	**0**

Table 5 continued

17	**total.load.forecast 28712 4594.101 18105**	**41390**	**0**
18	**total.load.actual 28697 NA 18041**	**41015**	**36**
19	**price.day.ahead 49.87 14.619 2.06**	**101.99**	**0**
20	**price.actual 57.88 14.204 9.33**	**116.80**	**0**

Entsoe retrieved from
https://transparency.entsoe.eu/dashboard/show

Openweather retrieved from
https://openweathermap.org/api