A machine learning approach to evaluating renewable energy technology

Abstract

Reasons for expanding renewable systems are strong due to their impartiality in energy production, ability to reduce greenhouse gas emissions, and little impact on the surrounding environment. The increasing rate of harm from nuclear and fossil power sources has increased the attention paid to new energy systems. In recent years, ML techniques have quickly increased in several renewable energy-related applications, including those dealing with energy production and integration, energy consumption, and demand analysis. In this regard, this research is an attempt to better forecasts renewable energy sources of countries by using data-driven models. Multiple AI and ML approaches were studied to accurately forecast the availability of renewable energy sources.

Table of Contents

[**Chapter 1** 4](#_Toc113239494)

[**Introduction** 4](#_Toc113239495)

[**Common sources of renewable energy** 4](#_Toc113239496)

[Solar energy 4](#_Toc113239497)

[Wind energy 4](#_Toc113239498)

[Geothermal energy 5](#_Toc113239499)

[Hydropower 5](#_Toc113239500)

[Ocean energy 6](#_Toc113239501)

[Bioenergy 6](#_Toc113239502)

[**What is the machine learning approach?** 6](#_Toc113239503)

[**ML in evaluating renewable energy sources** 7](#_Toc113239504)

[**Research objectives** 7](#_Toc113239505)

[**Research Questions** 8](#_Toc113239506)

[**Significance of the research** 8](#_Toc113239507)

[**Formulation of the study** 9](#_Toc113239508)

[**Chapter 2** 10](#_Toc113239509)

[**Literature Analysis** 10](#_Toc113239510)

[**References** 11](#_Toc113239511)

# **Chapter 1**

# **Introduction**

Renewable energy is derived from non-exhaustible sources such as are the sun and the wind. The world is teeming with renewable energy options. Coal, oil, and gas are examples of fossil fuels and non-renewable resources since their formation requires hundreds of millions of years. Generating electricity from fossil fuels releases carbon dioxide and other greenhouse gases (Abualigah et al., 2022).

Renewable energy generation results in far fewer emissions than fossil fuel generation. Transitioning away from fossil fuels, the primary source of greenhouse gas emissions, and toward renewable energy sources is crucial to averting a climate disaster. In most regions, renewable energy sources have dropped in price and currently provide three times as much employment as traditional energy sources (Hon, 2020).

## **Common sources of renewable energy**

### Solar energy

Despite gloomy conditions, solar energy is still the most plentiful power source. Solar energy is converted into usable forms around 10,000 times faster than global energy demand. Solar energy may be used to provide the appropriate temperatures, airflow, natural lighting, electricity, and fuels for a broad range of applications. Solar energy systems convert sunlight into electricity with the help of photovoltaic panels or mirrors. (United Nations, 2021).

### Wind energy

Wind power is generated by installing enormous wind turbines on land (onshore) or in saltwater (offshore) or freshwater (inland) (offshore). Although people have been harnessing wind power for millennia, advances in onshore and offshore wind energy technology have allowed for more electricity to be generated by using turbines with larger rotor diameters and greater height. Although usual wind speeds do vary widely from place to place, the world's technological potential for wind energy much surpasses worldwide power output, and most places have the potential to permit considerable wind energy deployment.

### Hydropower

Hydroelectricity uses the potential energy of water flowing downhill. Water storage facilities and flowing water are viable sources. Two types of hydroelectric facilities exist those that use water already stored in a reservoir and those that use the river's natural flow to generate electricity. In addition to supplying electricity, hydropower reservoirs often provide various other functions, such as water storage for human use, agricultural irrigation, drought mitigation, navigation, and flood control. Regarding the electrical industry, hydropower is now the greatest renewable energy source. It is vulnerable to climate-induced droughts and ecosystem shifts that alter precipitation patterns since they depend on these factors.

### Ocean energy

To create electricity or heat, ocean energy uses the sea's kinetic and thermal energy, which may be harnessed from things like waves and currents. Ocean energy systems are still in the early stages of study, while various prototype wave and tidal current devices are being examined. Ocean energy has theoretical potential far beyond actual human energy needs.

### Bioenergy

Wood, charcoal, dung, and other manures are all examples of biomass that may be used to generate heat and electricity, while agricultural products can be converted into liquid biofuels. Most people who utilize biomass do so in rural settings, and those that do so tend to be lower-income residents of developing nations. Various organic waste streams, agricultural and forestry wastes, and specialized crops or trees comprise the modern biomass system. Greenhouse gas emissions are produced when biomass is burned to provide energy, although they are less than those produced when fossil fuels like coal, oil, or gas are burned. Therefore, bioenergy should only be used in limited situations because to the potential severe environmental repercussions associated to significant expansions in forest and bioenergy crops and the accompanying degradation and land-use shift (United Nations, 2021).

## **What is the machine learning approach?**

Machine learning (ML) is a subfield of AI that helps computers improve their predictive abilities without being explicitly taught how to do so. Predictions made by machine learning algorithms are grounded on past data. ML describes a group of procedures that can automatically extract patterns from data, often in enormous amounts (Lanio, 2018). These methods are now widely used in areas such as targeted promoting, the development of voice assistants for smartphones, and the analysis of medical images to aid in making diagnosis decisions, all thanks to recent advancements in methodology, computing framework, and data accessibility.

## **ML in evaluating renewable energy sources**

Modern societies would collapse without their energy infrastructure, which is why many plans to ensure long-term ecological, financial, and social health centers on improving these systems. For example, energy availability is a cornerstone of economic growth. Transitioning to renewable and low-carbon energy sources will be crucial for achieving climate change and air-quality objectives (Lanio, 2018). It's no wonder academics, and practitioners have looked to machine learning and other fields to help them meet the pressing need for speed on these fronts (ML).

In the corporate world, machine learning and artificial intelligence are two of the most often discussed topics. As a result, companies in many sectors are exploring their potential for adoption in the hopes of enhancing and automating fundamental operations. The energy sector is no different. In reality, machine learning has shown to be quite useful for renewable energy firms (wind, solar, hydro, and nuclear) (Forootan et al., 2022). They cut expenditures, improved their forecasting, and saw a rise in their portfolio's return. This development is expected to accelerate in the future.

## **Research objectives**

**General Objective:**

The main objective of this work is to determine a method for predicting renewable energy resources, in order to allow the management of a smart electric network. For the management of this system we must determine a horizon of prediction of the grid renewable energy sources.

The specific objectives:

1. To understand the concept of renewable energy and to analyze the various types of energy management and their advantages over non-renewable energy technology.

2. Conduct a predictability study of renewable energy sources data to define prediction expectations.

3. Develop a method for predicting renewable energy sources. This step is the essential part of the job. It consists in carrying out a method for predicting renewable energy sources, meeting the needs of the management of the intelligent electrical micro-grid.

## **Research Questions**

* What is energy management, and what are the tools used to monitor and control energy consumption?
* What is the importance of predicting renewable energy sources for effective energy management?
* What are the required technological implementation for predicting renewable energy sources for monitoring and controlling energy consumption as a part of energy management?

## **Significance of the research**

Many developed and developing nations have used renewable energy technologies to generate power (Donti & Kolter, 2021). Interest in addressing energy insecurity, climate change, and pollution in the air has spurred the expansion of this technology's use in various areas (Khan et al., 2020). Reasons for expanding renewable systems are strong (Forootan et al., 2022) due to their impartiality in energy production, ability to reduce greenhouse gas emissions, and little impact on the surrounding environment. Yet policymakers need to consider the broader economic repercussions of such systems. Industrial, academic, engineering, government, civil society, and private sector professionals have all recently become more conscious of the job creation possibilities of contemporary energy production.

The increasing rate of harm from nuclear and fossil power sources has increased the attention paid to new energy systems. Environmental benefits, lower startup costs, fuel diversification, energy independence, improved energy enactment, and the potential for the development of power characteristics and safety (and in some cases, the postponement of grid expansion due to the likelihood of grid instability) are the primary drivers for the adoption of renewable energy sources like wind and solar (Lanio, 2018). In recent years, ML techniques have quickly increased in several renewable energy-related applications (Perera et al., 2014), including those dealing with energy production and integration, energy consumption, and demand analysis. Since it is costly to store energy from wind and solar PV, careful control of energy output is required. Throughout the many phases of a power grid incorporating renewable energy sources, different machine learning approaches are used as appropriate for the needs and nature of the situation. Several investigations have uncovered the use of various machine learning methods for forecasting the production of renewable energy sources (Vennila et al., 2022). Better forecasts for renewable energy may be made using data-driven models. Improved forecasts for renewable energy sources have also been achieved using hybrid machine learning techniques. Multiple AI and ML approaches were needed to accurately forecast the availability of renewable energy sources (Forootan, 2022).

## **Formulation of the study**

Chapter 1 introduces renewable energy sources and a machine learning approach. Chapter 2 provides a detailed analysis of the literature. Chapter 3 is the methodology, and chapter 4 is about results and discussion relevant to research aims. Chapter 5 is the conclusion and recommendations.

# **Chapter 2**

# **Literature Analysis**

2.1 Big Data and Artificial Intelligence

Large international companies are adopting Big Data (BD) and Artificial Intelligence (AI) initiatives on a sustained basis, as shown by the 2018 NewVantage Partners survey applied to executives from 60 Fortune 1000 companies. According to the study, 97.2% of companies are investing in the creation or launch of Big Data and Artificial Intelligence initiatives. Furthermore, among those surveyed, a growing consensus is emerging that AI and DB initiatives are closely intertwined. (NVP, 2018). However, the literature has provided very little empirical evidence on studies related to the use of Big Data in companies. When companies adopt new technological solutions, such as big data technologies, these companies can reap the benefits, but they must also be aware of the potential risks they may incur. On the one hand, investments in new technologies require changes that are observed over time in organizations, in order to reap the full benefits. In fact, companies need to maintain a range of managerial activities to ensure that they reap the benefits of a technological investment. On the other hand, they must improve their ability to manage the risks associated with a new technological investment, if they want to have successful implementations. But before continuing, what are we going to understand by Artificial Intelligence? By Big Data?

Artificial Intelligence:

Artificial intelligence is a broad spectrum of technologies that allow independent interaction with the environment. To put it in some way, computers are capable of “thinking”. Within this broad spectrum we find (1) Natural Language Processing (NPL) which allows computers the ability to understand oral and written discourse; (2) machine reasoning, which allows systems to execute logical reasoning to perform data inference; (3) automatic process planning, in which the systems are capable of autonomously defining the sequence of processes to reach a final goal; (4) machine learning, which uses a variety of algorithms that recognize the structure of the data to improve the information, describe information and make predictions in different scenarios. These algorithms are trained based on a set of data to determine the existing relationships between them and, later, to be able to deliver a result that depends on new data groups used in the training of the algorithm. It is from Machine Learning that we can start talking about Big Data, since these Artificial Intelligence methods allow us to analyze large amounts of data and give them coherence.

Big Data:

Today we see that this technology is a field that is growing and opening space rapidly, its general description deals with the analysis of large databases using artificial intelligence or machine learning, however, "big" data, becomes more acute when the development in the scientific community is slower than the development that we see commercially, both in innovation and conceptual development, which leads the Big Data concept to tend to be a salable wrapper or fanzine. Finally, it is recurrent to define BD through the "Vs" of BD, which include (1) volume (amount of information), (2) variety (heterogeneity or complexity of information), (3) speed (speed production and dissemination of information), (4) value (newness of the information) and (5) veracity (quality or pedigree of the information). However, this classification also has its limitations, since there are descriptions where there are 10 "Vs", but most agree on these 5 "Vs".

2.2 Energy industry:

The electricity generation industry is characterized by a good source of non-conventional renewable energy (NCRE) compared to the world average and a high concentration in a few companies that have a large part of the generation market. Now if we go into detail how the composition of generation is distributed by type of energy, we have to analyze it based on the installed capacity they have. The concept of installed generation capacity refers to the electricity generation capacity of the system, in terms of gross power installed in the infrastructure. The national installed capacity for the year 2019 amounts to 25,558.74[𝑀𝑊]. It is categorized approximately in 54% of conventional thermal energy (from fossil fuels), 27% of conventional hydroelectricity (greater than 20 [𝑀𝑊]) and 19% of NCRE (Lia, 2020).

Non-conventional renewable energy industry (NCRE):

The use of Big Data in relation to the energy industry focuses on non-conventional renewable energies (NCRE). These, according to the Ministry of Energy, are less developed resources and technologies and/or with little penetration in the current energy markets. Specifically, according to Law 20,257, wind energy, small hydroelectric power (plants up to 20 [𝑀𝑊]), biomass, biogas, geothermal, solar and marine energy are considered NCRE. For the specific case of waste-based energy (WBE), Law 20,257 establishes that NCRE is considered to be those whose energy sources are biomass, understood as including the biodegradable fraction of household solid waste and not home. Additionally, the National Energy Commission (NEC) can qualify as NCRE those plants that use renewable energies for the generation of electricity, that contribute to diversifying the sources of energy supply in the electrical systems and that cause a low environmental impact. Consequently, the treatment of the biodegradable part of Municipal Solid Waste (MSW) in WBE plants qualifies as an NCRE source while the classification of a WBE plant that treats all MSW fractions is subject to the decision of the NEC.

The United States Environmental Protection Agency (EPA) classifies energy sources in a similar way, between conventional (or non-renewable) and renewable. Among the renewables, it establishes a subset called green energy, equivalent to the NCRE classification, but not only subclasses them by the nature of maturity, but also they are characterized by providing a greater environmental benefit compared to the rest.

2.3 Big Data and Artificial Intelligence applications in the energy industry worldwide:

In this section, the potential of Big Data to contribute to the improvement of existing models of energy economics and for policy analysis, both in the near future and in the long term, is explored. Current developments and innovations in data collection, storage, retrieval, and processing have already begun to transform science in a number of areas, notably high-energy physics and astronomy. In the near future, these innovations may offer the potential to change the operating environment of energy models from one characterized by data scarcity to one characterized by data abundance. However, in the field of energy system modelling, it has not yet begun to grapple with what future advances in information science mean for established practices, tools and techniques.

Most common DB and AI applications in an energy system:

1. Production forecast:

This application is relevant for NCRE within the energy system. The programming and operation of the energy system face the challenge that when entering the NCRE, uncertainty increases, since the amount of rainfall per year is variable and there is no certainty of energy generation given daily weather conditions, even hourly. Therefore, accurate forecasts of renewable energy production and energy demands at different levels are critical. Although many methods have existed for wind and photovoltaic energy production forecasting and electric load prediction, most of them are based on small data sets, without using large volumes of data, such as those provided by electricity meters (Smart Metering).

Big Data analytics is most effective when it comes to energy production and consumption forecasting based on large databases. The deployment of smart meters and the establishment of numerical weather forecasting and GPS systems make it possible to give more detailed and accurate forecasts through big data analysis (Zitnick et al, 2020).

2. Analysis of the behavior of energy consumption.

The identification and prediction of energy flexibility on the demand side are the foundations for implementing demand response. Understanding consumer preference and energy production/consumption behavior is the basis for analyzing their willingness and potential to engage in demand response and energy trading.

3. Security risk analysis.

The identification of disturbances in the electrical system, the evaluation of stability and emergency control are essential to guarantee security of supply. With the deployment of Wide Area Measurement Systems (WAMS) control centers, large volumes of data are being generated, therefore, transforming data into knowledge, preferably automatically, is a real challenge for operators of systems.

4. Operation Mode Layout

How to estimate energy system flexibility at various levels and effectively manage flexible sources is essential for energy balance. Data analytics can play an important role in this aspect, from energy management in the micro-grid based on deep learning to all kinds of flexibility estimation in all kinds of flexible sources, such as energy conversion / storage. and flexible loads.

2.4. Why Big Data and NCRE?

With the background, we can see that the greatest development or intersection of Big Data and the energy industry are found: (1) in the perspective of the energy system and the development of the energy economy, which leads us to the intensive use of Big Data; (2) in NCRE given their level of uncertainty with current models and the variability of the energy source; and (3) maintenance of generation systems. As we see in Figure 2.1, based on a study [8] on research that intersects energy efficiency and Big Data, the largest group of research (34.3%) explored the informatics of energy efficiency, such as techniques for maximize performance per watt when processing large sets of data. The second most important group (25.5%) is energy efficiency in buildings. The third (16.8%) is from Smart city and the fourth (11.7%) and the fifth (11.7%) sustainable manufacturing and the reduction of industrial use of energy.

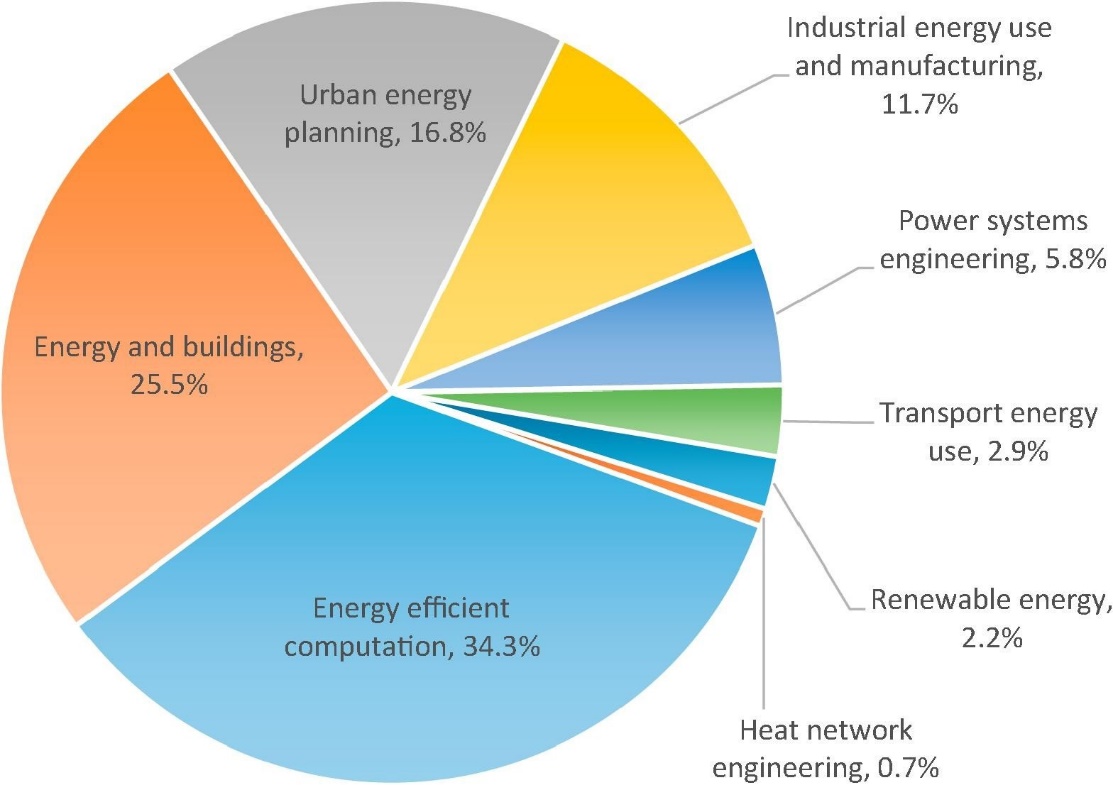


Figure 2.1: Percentages of Big Data and NCRE research topics in percentages

The remaining 12% of the sample reveals several interesting use cases for Big Data and energy research. One of these is the electricity system, where scientists are investigating possible design configurations for future "smart grids" and exploring the benefits for system management, stability and operational flexibility that can be achieved. This could be important in future networks that combine intermittent renewable generation (both distributed and centralized) with other types of generation. Another is a nexus of Big Data research focused on analyzing the complexity of transportation patterns as a means of anticipating how they might change or be managed in the future. This work focuses on the use of Big Data and Artificial Intelligence in the energy industry and in particular in NCREs, given the value that this technology has in improving high levels of uncertainty. Although today there are many methods for forecasting of wind and photovoltaic energy production and forecasting of electrical load, most of them are based on small data sets, without using large volumes of data, such as those provided by the deployment of smart meters and the establishment of forecasts numerical meteorological and GPS systems that make it possible to give more detailed and precise forecasts through the analysis of Big Data (Gu, 2019).

2.5 Uses of Artificial Intelligence and Big Data.

The uses of Big Data in formal terms, given the volumes that are technically defined for Big Data, are null within the world of energy in Chile. This can only be developed in this way in areas such as astronomy currently in Chile and in some cases in retail. However, there is intelligent use of large databases, with machine learning and artificial intelligence. That said, the applications are quite varied in the various companies. At least 4 main ones are observed: process optimization, energy market analysis, maintenance and weather forecasting for NCRE plants. Next, we will analyze the different applications.

2.5.1 Process optimizations.

The widespread application is that, by taking data from processes carried out in the industry, certain decisions can be automated through machines that make decisions with the data, in an "intelligent" way and with fewer resources than before through the data they collect for several months or years. The data in itself does not have a value, it begins to have value at the moment that it has the possibility of predicting certain decisions and what will happen if they are taken in one way or another. The nature of the decisions that are beginning to be predicted and optimized are decisions that are made every day, hours and minutes, they are micro decisions, which do not involve the company's board in them. They are logistics decisions, when to repair a machine, when an input enters a boiler, when to turn on a machine, that is, they are routine in the industry and decisions that are also supported by a lot of data. There is a lot of evidence of what was happening, what was done, what happened. It is almost always possible, with the abundance of data that exists today, to systematize that decision making so that by formalizing how that decision is made in terms of a model of some kind, help that decision making converge over time towards an optimal. In addition, this allows you to automate those decisions or simply allow you to see the processes more generally and detect other problems in the future (Wang, 2019).

2.5.2 Analysis of the energy market (electricity)

Within all the interviews, it is diagnosed that the long-term programming model (LPM) is exhausted, that the prediction of the energy market only with the projection of the hydrological basins with the history of the 1960s is an inaccurate projection given the phenomena of today's climate crisis; and that the entry of non-conventional renewable energies generates other meteorological variability that is less stable than just the amount of rain each year. On the other hand, a change is identified in the industry where at the beginning there were few clients, but given the technological and legislative changes that the country is experiencing (every time it is beginning to be more similar to a retail), there are more clients and greater possibility of preferences between one energy and another. Somehow the energy market is being liberalized. Before, each generation company had 5 clients, a maximum of 10, which is changing today, which makes the market more complex and more marketable.

Given the above, the different companies and study centers are needing much more data and better infrastructures to be able to make models of the system and thus be able to predict the average price of energy each year, how much each plant will produce, how much the demand of each of the actors in the system. This has pushed the commercial area of companies to start with an intensive use of data. Specifically, the utility for companies that today has analysis of the energy market is to be able to evaluate new projects by generating companies with much greater precision than the current one, in addition to preparing for the future when the client portfolio increases and the possibility of to predict to give a guarantee of the energy that they are going to produce. On the other hand, the electrical coordinator is in an intensive study to have the possibility of obtaining a new model that improves the bidding processes and the current price system.

2.5.3 Asset Management and Maintenance

Throughout the research process, 2 applications are observed in the management of physical assets and maintenance using data intelligence. The first has to do with having sensors inside the machinery to be able to constantly monitor and be able to do types of predictive maintenance. In order to optimize the costs of operations and maintenance. This is more traditional and is dealt with in many other industries. Making use of Prognostics and Health Management (PHM) or predictive maintenance and other types of logic. The second is a very punctual problem and it is the case where there is a Big Data application. The challenge responds to the problem they faced in the transfer of oil in the pipelines. The company has diagnosed “mega risks”, where there is one in particular, which is the loss of containment of a 350 km oil pipeline. The law requires that the pipeline be monitored since there may be a perforation, with great complications for the environment, with economic and reputational costs, etc. So, the law requires that "you must have people walking, watching", and that is neither economically viable nor socially responsible. A year ago, there were planes that passed by there every day and took pictures and generated reports about it. This was changed to a drone surveillance operation with artificial intelligence, taking pictures or videos and broadcast this in real time to a server that processes the data and is capable of discriminating whether or not there is a risk in what you are seeing. Basically, it was possible to understand, classify and train the algorithm on the risks that were being displayed. If there was a machine on top of the pipeline, then that is a risk, or if there is a fire, or some earth movement nearby, it is possible to discriminate the degree of danger of the situation (green, yellow or red). All this on a platform where all this information is stored and gives warnings when it is yellow or red, without the need for permanent staff. If there was a machine on top of the pipeline, then that is a risk, or if there is a fire, or some earth movement nearby, it is possible to discriminate the degree of danger of the situation (green, yellow or red).

2.5.4 Weather forecast, guarantee energy.

This application is recent and experimental and was being carried out by a company in the context of the change in the distribution law where prices are going to have greater variation. What you are looking for is a good forecast management analysis, to be able to know if the day is cloudy or if it is sunny and in this way project photovoltaic generations. The most important thing for this is to be able to guarantee the generation of energy, in order to have a better price. This application could also be for wind power, but the application was not seen in the interviews conducted. Despite these 4 types of applications, each of them could be many more and more developed. Later, the obstacles that are facing the incorporation of BD and AI in companies and organizations will be analyzed.

2.6 Obstacles and Challenges for the development of Big Data and Artificial Intelligence in the energy industry

Below are the main obstacles and challenges for the development of BD and AI in the sector (Perea, 2014).

2.6.1 Lack of data source, infrastructure and possibilities of processing of that data.

This difficulty is mainly related to the fact that there is little monitoring of variables that today could be measurable in the energy system, in order to have better modeling of the market, prices, generation and to be able to better coordinate demand and supply of energy. Although there is plenty of data in terms of operation and maintenance of the generating plants, there is very little data regarding demand, regarding weather changes and wiring conditions, etc.

2.6.2 Organizational culture

This is the most recurring obstacle of all, this not only occurs in the energy industry but also in other industries that are already more advanced in the intensive use of data analytics. The main obstacles in this case have to do with the credibility of the technology, after its use in decision-making and in the operation of processes, whether or not it is considered, in addition to the resistance to change that every organization has.

2.7 About current applications Where to start?

Although it is necessary to take certain risks, or invest, the application of machine learning and Artificial Intelligence must be to face a practical and profitable problem. It seems like a truism definition, but there are many companies that have taken the definition of joining the BigData trend simply because it is something that is fashionable. Therefore, the first applications that are being seen and recommended are to solve part of the strategic problems seen in the previous point. The major applications applied to specific problems are (Khan, 2020):

1) Low returns:

To face the low profitability of the companies, measures are taken to lower the generation costs of the plants, which leads to optimization of processes and maintenance.

a) Optimization of generation processes.

The widespread application is that by taking data from processes carried out in the industry, certain decisions can be automated through machines that make decisions with the data, in an "intelligent" way and with fewer resources than before through the data that they collect for several months or years. The data in itself does not have a value, it begins to have value at the moment that it has the possibility of predicting certain decisions and what will happen if you make them one way or another. The decisions that are beginning to be predicted and optimized are decisions of an operational nature, which are made every day, hours and minutes, they are daily decisions, which do not involve company directors. They are logistical decisions about when to put an input into a boiler, when to turn on a machine, when to repair it, that is, they are micro and routine decisions that also have the grace that they are often backed by a lot of data. There is a lot of evidence that it was happening, of what was done, what happened, it is almost always possible with the abundance of data to systematize that decision making so that by modeling that decision making over time and iterations can converge to an optimum level. Specific examples at this point are, optimization of a boiler, a gasoline stove, the use of a Biomass Battery in a Biomass power plant (Khan, 2020).

b) Optimization of the Maintenance of physical assets.

This application tries to solve 2 main types of problems, one that is more studied, which has to do with predictive maintenance using data intelligence, and the other is how to face complex maintenance problems where it is openly sought to replace processes that require human judgment. The first application is related to a more traditional use that is also used in other industries, which is basically the use of PHM or predictive maintenance. It is shown that you have sensors inside the machinery to be able to constantly monitor and predict maintenance to optimize operational costs and logistics itself. The second application has to do with a specific maintenance problem where image processing is used to predict possible failures and it is the case where it is more possible to use Big Data. The challenge consisted of the following case: There is an oil transfer through pipelines of 350 km or more. The law requires that they be monitored, where it is made explicit that "you must have people walking, watching", and that is neither economically viable nor socially responsible, but on the other hand there may be a perforation of the pipeline and this may be very complicated for the environment, with economic, reputational costs, etc. For the company, this classifies as “Mega Risks”. The solution that was done a few years ago was that planes flew over the pipeline every day and took pictures and generated reports. Today an operation is carried out with drones with artificial intelligence. The drones monitor the pipelines and take videos and photographs and send this information in real time to a central server that processes the data and is "trained" to discriminate whether or not there is a risk in what it is seeing. Basically, if there was a machine on top of the pipeline, then that is a risk, or if you have a fire, or some earth movement nearby, we are able to discriminate if you are green, yellow or red (Ahmed, 2020).

2) Strong income from renewable energy is just in the experimental phase of weather forecasting.

This application is very limited and recently experimental. Based on the challenge that the new distribution law opens, where prices are going to have greater variation, based on this, a good generation forecast analysis is being sought. Specifically, to be able to know if the day is projected to be cloudy or if it is projected to be sunny and in this way project when the photovoltaic generation will be. This assumes that if power generation can be guaranteed, there will be a better price.

3) About the infrastructure for data analytics.

This recommendation is pertinent for companies that want to take the definition of taking charge of their own processes and do not outsource it to another company that does the storage and analysis of company data, examples of companies that do this EY, Accenture, universities, etc. There are three elements to consider in this case: (1) where I store the data that is collected, (2) what processors the data is analyzed with, and (3) what type of software is used. The first thing is that this decision is strategic and for point 1 and 2 there are two paths:

(1) Have your own infrastructure, that is, have your own data center or (2) Outsource and occupy space in the "cloud". The definition of the cloud may seem vague, but basically it is a term used to describe a worldwide network of servers, each with a unique function. The cloud is not a physical entity, but a huge network of remote servers around the world that are connected to function as a single ecosystem. These servers are designed to store and manage data, run applications, or deliver content or services, such as streaming video, webmail, office software, or social media. The first thing to clarify is that due to the use of the cloud there is no risk of data vulnerability, so far there are no cases where this has happened. Therefore, the definition of whether or not it is carried out with external infrastructure is of a strategic nature and there is no correct answer a priori. It is important that the expansion of storage and/or processing is in line with the specific challenges that the company faces and is economically justified. On the other hand, most companies started with smaller pilots and, therefore, a limited survey of internal infrastructure and then made the decision to migrate to the cloud. Depending on the nature of the use of the data, if for example they become a pillar of the company, they also sometimes make hybrid use cases.

2.8 On organizational challenges.

This section will assess the main organizational obstacles that companies face and possible ways to address them. There are two possible scenarios that the organization faces.

1) The outsourcing of the data analytics service, which brings very few organizational challenges since the intervention, not being permanent, does not represent a change in the organizational culture. On the other hand, if it becomes permanent, it usually requires some significant degree of internal development within the company. For this reason, it becomes the second scenario where the organization begins to experience a change in culture and capabilities. This definition of internalizing or not is a strategic definition since it represents the degree of depth of the problem being addressed. Examples of this type of case are the development of the maintenance of certain equipment with PHM, or the gas pipeline, etc.

2) The second scenario that is better explored is related to the internalization of these skills and how the different companies have faced it. It will begin by showing a general diagnosis and particular obstacles, to move on to recommendations (Gu, 2019).

General diagnosis

Most companies start by developing a model in an innovation, digital transformation or specific project development department. Once developed there, they scale it or opt for a complete area of intensive data use where they include data scientists and interdisciplinary groups that coordinate with the rest of the company's departments and areas.

Main obstacles:

1. It is not a conversation of the directories of the companies:

In many companies, the decision to implement data analytics technology and/or Big Data is not a decision or conversation that has a company-wide perspective. This occurs when the application of technology occurs in specific departments and/or for specific tasks. This brings difficulties when scaling a solution, obtaining more financing or obtaining strategic data from the company or from other areas.

2. Resistance to change: Lack of skills and ignorance of technology:

This obstacle has to do with the few skills developed in professionals and workers within the company regarding the terminologies, technical implications, scope and meanings of Big Data and Artificial Intelligence. This often prevents speaking a common language or posing challenges that only exceed the bets of a specific department, since the lack of qualified personnel prevents further growth with technology. This leads to various problems, for example: the model is not implemented in the operational cycle or the decision makers do not consider the analyzes provided by the data. Being a new and unknown technology for the technical and managerial staff, many times a very good model or data analysis is left without being executed.

3. The lack of coordination of areas within the company.

The intensive use of data analytics is a technology that can be used in various areas of the company and requires various data from the same company. Given this, the relationship between areas, headquarters and resources generates problems in the process of change. There are many times that the areas of a company are suspicious of their resources, data and chains of command of each one, which makes change actions difficult due to lack of coordination and lack of awareness of the challenge or problem to be faced. In addition, the problem that arises in areas of digital transformation is manifested, which part of its functions is the development of data analytics, but these do not have sufficient resources to implement the challenges or their chain of command is not well defined with respect to other areas, which leaves this task immobile.

2.9 Proposals on how to address the change in organizational culture:

This is one of the problems that companies face and of which they do have control, unlike the regulation of the State, the meteorological and technological infrastructure of the users. So, it is important to address this issue in more depth, since the cost of not taking care of this problem is that the investment of resources and technique made remains a dead letter within the company. The first thing is that the design of the implementation of a culture of data analytics within the company requires that it be coupled to the strategic plans of the company and, as shown at the beginning of the recommendations, there are 2 strategic challenges for which electricity generation companies are going through, which has direct implications in the use of intensive data analytics, so any attempt to introduce this technology within companies must dialogue with these specific challenges that are manifested.

Therefore, there will be an important change within the companies given that by changing some strategic visions, they begin to touch the ways of functioning of the same organizations. In this sense, they face a change management problem within which there are certain minimum principles that they must face (Forootan, 2022).

1. Change processes are “processes”

We must understand that workers and managers within the company are facing something new and that the internalization of new technologies is not going to do overnight, where also mistakes are likely to be made. At this point it is very important that anxiety is not what drives the process but rather understand that this transformation is going to take time, but that the pressure for transformation must be continuous, have a procedure or a clear plan on how to address the technical and human areas, but in spite of that the process can have ups and downs, that is to say, it is a process.

2. There is a previous culture of the organization.

As we have described above, there are elements of the company culture or traditions that often prevent openings to change from being generated. The important thing about this is that you don't start from a blank sheet of paper, but you have to be able to make the changes based on the current organization chart, the customs and desires that the organization previously had and from those same power to promote the change that we are searching. In fact, in many companies the motivation of not being destroyed by a technological advance like this is enough to push this type of transformation, or if the vision of innovation manages to enter into the same dialogue. In this case, there is a general culture regarding the law of distribution and climate change that is pushing these strategic changes defined above.

3. The design must be carried out by the organization

The design of the change must be a company decision, that is, from the board of directors or from the highest possible leadership, it is very difficult to successfully implement these technologies if it is not a decision that comes from above and is agreed upon as a company policy as well as being backed by resources (Hon, 2020).

4. Implementation must be done by the organization

At this point it is not recommended that the implementation of this be carried out by external companies, ideally it is the same company with its own chains of command and bureaucratic procedures that implements the transformation that is to be carried out. For this it is important to create new procedures and instructions for the new functions, tasks and responsibilities.

5. There is a continuous process of design and redesign

Being a process of change and new to the organizational culture of the company, mistakes will probably be made, so there will be constant work to design proposals and redesign them. For example, in a company at the beginning, an implementation was carried out with a specific area of digital transformation that contained the task force of data analytics, had the power to work with all the areas and even in the organizational chart was above the others areas. But, since there is no formal space where all areas will make decisions regarding the challenges of technology, the measures ended up not being implemented. From this, the initial design was rectified and a new space was proposed that was a transversal and interdisciplinary area and with its own resources to face problems that can be solved with BD.

6. It is essential to care about people throughout the process.

This is important, organizations are not abstract entities, they are made up of people. In several companies, managers state that the main problem for the implementation of technology is people. On the other hand, probably the implementation of these processes in some cases will bring relocation of functions, new tasks or even dismissals. If these types of things are not dealt with well within the staff, they can stop an entire process of change.

7. All process of change is a learning process. It is important to reveal what learning?

It is important to make clear what is expected of people, what knowledge they should acquire. For that, the minimum things are training to combine nomenclature, technical concepts, and in the case of more specific roles, technical skills. This is a task that is not usually addressed and it is of tremendous importance that the technical concepts are not only handled by the data team but also by all the counterparties with which they are related in order to better project the applications in the different areas. Training and technical needs. Do not start with Big Data or AI all at once, start with issues of digital transformation, to gradually install a data culture within the company.

8. People and emotions are valid

It is necessary to consider that one of the variables and obstacles present is the fear that the “Machines will take away the work”; it is very important to give space to that fear and to be able to give certainty in this regard. Also another fear is to be wrong that is often present in companies. If you want to innovate and test new technologies, you have to give yourself a range and a space for these feelings to manifest.

2.10 Recommendations to the State.

At this point, rather general aspects will be addressed, collecting elements that were repeatedly expressed in the interviews, given that each of these recommendations could be a complete document or even an investigation with much more detailed proposals (Hon, 2020).

Use of smart meters or Smart Metering.

A smart meter or smart meter is a type of advanced electricity meter (watt hour meter), which calculates consumption (or production) in a more detailed way than conventional or analog meters. These devices also offer the possibility of communicating this information through a telecommunication network to a data processing center of the local service company, which can use the data for the purpose of billing, monitoring, providing its users with greater control over their consumption, or even being able to offer personalized services to customers. Likewise, these devices have the ability to configure the service to measure, or interrupt the supply remotely, in the event, for example, that the monthly payment for the contracted service has not been made.

These are its advantages:

1. They enable rapid identification of problems and failures whose solution, therefore, will be faster.

2. They offer a simple option to modify the contracted power remotely, without the need for a technician to attend.

3. They provide greater security to prevent tampering, as the smart meter is continuously and automatically monitored.

4. They leave estimated readings behind to make way for actual readings.

5. Its operation can be checked by means of a lighting system: off (there is no electricity consumption in the house), it flashes slightly (electricity consumption is taking place at this moment), it flashes quickly (the electricity has jumped) or on and fixed (the contracted electrical power has been exceeded in the home or premises).

The proposal to install smart meters is made to improve the data we have regarding consumers and in this way be able to generate more precise data regarding the input and output that each home or industry in the country will have and in this way improve hours and time higher demand and generation, in addition to allowing cogeneration in homes or small businesses (Gu, 2019).

Improve meteorological infrastructure.

Substantially improve national meteorological monitoring, which allows for real-time data, achieving an atmospheric studies-oriented approach that simulates the earth's atmosphere and the entire atmospheric and temporal system of the earth to forecast weather conditions. Currently there are 1299 registered Meteorological Stations, more than half only have paper records and most have data collection that only allows daily projections but not hour by hour and less minute by minute, there are only between 30 and 50 online. This makes it possible to have forecasts for the agribusiness but it is impossible for the energy industry. In addition to this, it allows generating the infrastructure to be able to store this amount of data and make it publicly available.

Regulatory changes:

There is a consensus today among the engineers, data scientists and workers who are today in the electricity generation companies that there are several regulatory changes that must occur, some are already happening and each one of them dialogues with each other. Although no proposals were outlined from the interviews, the discussion and definition of all the laws do manifest themselves as concerns within the industry.

Electricity distribution law:

This law is the most advanced and in general faces some of the main challenges that the industry is facing today. Its impact with respect to the intensive use of data has to do mainly with the lowering of barriers for free customers, which allows generators to enter commercialization and therefore makes the energy sale activity more complex, in addition to making price models more complex and demands of the system in general, therefore an intensive use of data.

Climate change law:

This law does not yet exist, it is an issue that is culturally very installed as a global public policy problem. However, for now there is only a framework agreement between the large power generation companies. In the event that this law is promoted and comes to fruition, its greatest impact will be the intensive entry of NCRE and the closure of coal-fired thermoelectric plants that today are the ones that give the most “stability” to the electrical system. The entry of this type of energy would bring uncertainty within the system, which would make it necessary to enable data systems to have greater certainty of generation with NCRE. Regarding the agreement, see some more details in the previous item.

Law Regulation for the data market:

Today a data market law is in progress, however there are some minimum issues that are essential to address from an ethical perspective. The first thing to address is that the current legislation is deficient in the protection of personal data. If we are in search of a real protection of the interests involved and real protection of the data of individuals, matters absent in Law 19,628 (Law for the Protection of Private Life) must be deepened. In the second instance, the current law correctly defines basic concepts regarding the processing of personal data. It differentiates what should be understood by personal data, expired data, sensitive data and what is a database, however, it fails to attempt a precise definition of data source accessible to the public. The structuring of the principles enshrined in Law 19,628 is also deficient. Regarding the principle of freedom to process personal data, the legislation does not provide sufficient guarantees to establish a balance between this and the right to privacy and intimacy of individuals. In relation to the second principle recognized by the law, and under the same prism, the breadth of the exceptions to the consent requirement is excessive. This is very necessary in the case of energy data since there will be a very detailed X-ray of the energy consumption of each household and what will these data be used for? Given the latter, the proposal is to have an autonomous, administrative, specialized Data Protection Agency with its own assets, to be able to carry out their oversight work without external political pressure and whose decisions respond to professional standards.

To fully carry out its work, an agency with these characteristics must also have the necessary resources and powers. It is not enough that the entity can monitor compliance with the law by public and private organizations, nor that it can sanction them with a fine. It is also necessary that it have the power to carry out educational campaigns, keep records of existing databases, inspect those responsible for databases, establish precautionary measures and coordinate international cooperation in the area. Finally, the agency requires regulatory powers, to generate administrative jurisprudence on the scope and interpretation of the law; through opinions or resolutions, either of a general nature or specific pronouncements (Abd El-Aziz, 2022).

**Chapter 3 Methodology and Design**

3.1 Artificial Intelligence Methods

Artificial intelligence methods are computational processes that aim to provide computer systems of a behavior which tends towards the capacity of intelligence of the humans. These calculation methods are inspired by the calculation methods of the human brain (Forootan, 2022).

3.1.1 Neural networks

Artificial neural networks are algorithms, used in several fields including optimization, which are directly inspired by the functioning of biological neurons. Like most optimization algorithms, they need learning methods to adjust the weights of the algorithm. The learning methods used for artificial neural networks is generally of the probabilistic type, in particular the Bayesian type. They are considered from the family of statistical applications. But more specifically they are from the family of artificial intelligence methods.

3.1.2 Wavelet networks.

Like neural networks, wavelet networks compute a linear combination as a function of adjustable parameters, of non-linear functions whose form derives adjustable parameters (translations, dilations for the parameters of the wavelets in the hidden layers and network weights). However, the particularities in relation to the networks of neurons is that the probelets are rapidly decreasing and tending towards zero unlike the backbone functions of neural networks, also the wavelet shape is related to two structural parameters.

3.1.3 Support vector machines (SVM)

A "support vector machine" (SVM) is a learning algorithm that is often referred to as a wide-margin splitter. SVMs are originally defined for discrimination. SVMs have good generalization properties.

3.1.4 Form of learning

In the literature we find several forms of learning for optimization algorithms to adjust their parameters. The choice of the learning method depends on the application, the nature of the data and the resources available for the implementation of the system.

3.1.5 Supervised learning

The learning of an optimization algorithm is monitored, during the training of the parameters of the algorithm, when it is forced to converge to a final state at the same time as it receives data. In this approach, parameter adjustment is done while the algorithm is running.

3.1.6 Unsupervised learning

Unlike supervised learning, the optimization algorithm is left free i.e. if the error deviates from zero, the parameters will not be automatically adjusted while the algorithm is running.

3.1.7 Model performance

There are several prediction models in the literature. We have the persistence method, the physical, statistical, hybrid etc. approach. For short and very short term prediction, the statistical approach is the one that we find the most appropriate, if we rely on the literature. Neural networks are among the most used methods (Feed-Forecasting, Reccurent, Perceptron ADALINE, etc.), as well as time series (ARX, ARMA, ARIMA, Gray Predictors, Smoothing etc.), wavelet networks , and Support Vector Machines SVM. Hybrid methods also give good results in terms of prediction at short and at very short term. But these methods are generally more complex and expensive to perform. Some examples of hybrid methods are NWP (Numerical weather prediction) + NN (neural networks), ANN (artificial neural networks) + adaptive network-based fuzzy inference system (ANFIS) etc.

**3.2. Selection of Operating Parameters of the Model**

3.2.1. Database

Within the existing databases, it was decided to use datasets from NASA. The reasons for working with one of the two are based on the fact that their access is free and free of charge, in addition to meeting the requirements for their use. Although NASA presents ranges of lower precision since in the data collection by the satellite image method, the presence of external factors such as polluting particles and aerosols are taken into account, that alter the performance of the model by including indices such as AOD (Aerosol Optical Depth) which affect the magnitude of the measured values of GHI and DNI. Moreover, NASA on its web portal has updated information to date, providing data with minimal time delays that are ready within approximately four days of real time, thanks to the use of multiple data sources that provide more complete information that benefits the model so that it makes predictions that take into account current climate changes. NASA has an approach towards generation from renewable sources, giving greater attention to applications for the production of energy through non-conventional sources, therefore the acquisition of the required information is simple and fast in this database, giving it robustness and allowing the pre-processing to be efficient (Ahmed, 2020).

3.3. Model Development

Taking into account the previous considerations, the techniques that will be part of the final model are chosen. For this process, the Python® programming language is used, in the development environment Spyder (Scientific Python Development Environment) which is open source, designed for scientists, engineers and data analysts. Among its most relevant features, it presents advanced functionalities in editing, analysis and debugging for data inspection with an interactive execution. It begins with the treatment of the input information, which is key to feeding the classification algorithm that is going to be trained. In the classification stage, according to the theoretical framework, techniques such as fuzzy logic, Bayesian classifiers and support vector machines were considered. Initially, it was decided to test the performance of each of the techniques to be certain of which one is optimal for the application of the study carried out. The first technique evaluated is the Fuzzy Logic (Fuzzy Logic), which presents as strengths the management of multiple inputs, optimal performance in the classification against data with high variability, ease of implementation and is highly flexible due to the fact that it can be modified in any way, depending on the solution of the problem that is required. However, this technique is strongly linked to the expertise of the person who designs it, in addition to the fact that its operation is based on “if-then” type rules that depend on the number of variables and logical connectors used for its development. Therefore, the greater the number of variables taken at the input, this translates into an exponential growth for the number of rules that the technique takes into account, and likewise, it leads to the consumption of computational resources required by the algorithm to generate an output is larger (Nam, 2020).

Regarding the creation of the rules, there are alternatives that allow their development, and one of them is the use of the expertise provided by the investigation. However, when you don't have that prior knowledge, you can compensate with optimization techniques. For this study, techniques such as particle swarm optimization and genetic algorithms were used, with the purpose of being able to generate the set of ideal rules, adjusting the value of the weights of the outputs obtained for each rule. In works such as the use of these techniques in the creation of rules is shown in greater detail. Another of the algorithms that are tested are the Bayesian classifiers (Naive Bayes), of which, among its most relevant characteristics, is an easy implementation due to its simplicity and good handling of large amounts of data, in addition to predicting in data sets of high dimensionality. However, this technique is impaired when in the testing stage it finds data that is not in the training. Therefore, its performance against the data set for the study is evaluated and other algorithms are also reviewed, such as Support Vector Machines (SVM) that also adapt to the characteristics of the input set.

SVMs are highly generalizable and resistant to overfitting problems; furthermore, it performs well with high dimensional features. It also has low memory consumption because they use a subset of training points in the decision phase. However, this depends on the database used, since if there are scattered values in the data set, the algorithm requires more resources to be able to establish the classification. This type of model has a solid theoretical base that, making use of statistical learning theory, guarantees a good level of generalization from the input data. For the evaluation of the performance of the different classifiers that are used, it is decided to use the correlation metrics, MSE and MAE, which are metrics frequently used in regression stages, in order to have a reference that reflects the improvement of these metrics when the regression stage is added. In order to carry out this, the normalization process of the discrete values obtained by the algorithms is carried out in order to be able to use the metrics mentioned initially. Both Support Vector Machine and Naive Bayes techniques present reasonable metrics, where the SVM classifier turns out to be the one with the lowest error values (MSE, MAE) and the highest in terms of correlation (Pearson) and (Coefficient of determination). However, the execution time was almost double for Fuzzy and much higher compared to Naive Bayes. In addition, it is contemplated that the results of the metrics for Naive Bayes, although they are below SVM, their values are still within the acceptable ranges for the good performance of the model. For the choice of the classifier that best suits the needs of the study, the execution time is an item that becomes important since a general and light model is sought so that it can be portable without requiring considerable amounts of resources for its execution. Therefore, in this classification stage the algorithm selected is Naive Bayes. Regarding the regression stage, use is made of Neural Networks and SVR (Support Vectors Regression), which is the modification of the SVM focused on regression, which are the techniques most used in primary resource prediction tasks.

**3.4. Model Structure**

The dataset was divided into different blocks. Each block represents a separate country or an area that comes under the radar of data analytics. Our data set contains analytics from 10 different blocks, namely, Africa, Asia Pacific, CIS, China, Europe, France, Germany, Italy, Mexico, Portugal, United States, World. Moreover, data available regarding these countries is Biofuels Production - TWh - Total, Wind Capacity, Hydro (% sub energy), Electricity from hydro (TWh)\_x, Geothermal Capacity, Solar Capacity, Wind Generation - TWh, Solar Generation - TWh, Geo Biomass Other - TWh, Hydro Generation - TWh, Electricity from wind (TWh)\_x, Electricity from hydro (TWh)\_y, Electricity from solar (TWh)\_x, Electricity from other renewables including bioenergy (TWh), Renewables (% sub energy), Hydro (% electricity), Renewables (% electricity), Solar (% electricity), Wind (% electricity), Electricity from solar (TWh)\_y, Solar (% sub energy), Electricity from wind (TWh)\_y, Wind (% sub energy).

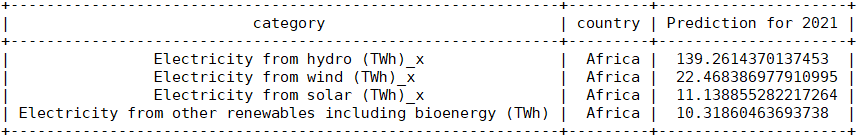
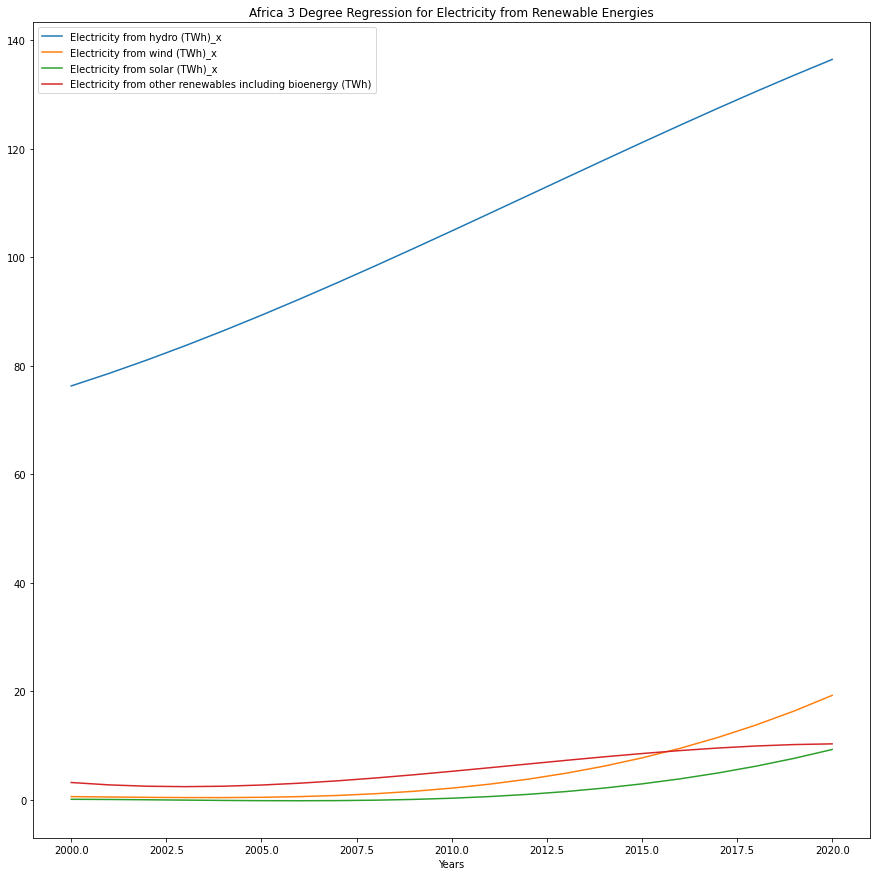
Thereafter, most of this data was found to be redundant and some of this data was not useful for our machine learning applications. Therefore, what we needed to do was to figure out the trend of individual renewable energies and see if any of them correlate with each other. This way for practical purposes a trend of investment can be figured out for these industries proving beneficial for both the investors and the consumers of these renewable energies.

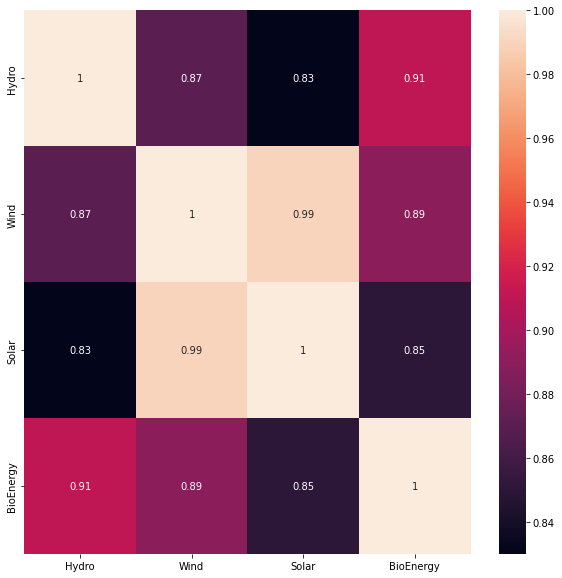
In this regard, we first applied 3-degree polynomial regression on each block, taking only four of their features. These features include Electricity from hydro (TWh), Electricity from wind (TWh), Electricity from solar (TWh), Electricity from other renewables including bioenergy (TWh). For the input feature we used year that goes from 2000-2020, and all the other 4 features were the output. This way we performed multi-output polynomial regression on the dataset. This regression analysis allowed us to calculate the prediction for the next year for that block. We give these prediction values in the table below the regression results. After calculating future prediction, we calculate the correlation that exists between different renewable energies in different regions. What this allows us is that we can see which countries are favoring which renewable energy sector, or if some country is putting extra effort in some sector at the expense of the other.

**Chapter 4 Results and Discussion**

With the advances generated in technological matters, the diversification of the energy matrix worldwide has become quite relevant by including generation methods whose essential raw material is non-conventional sources of energy such as the wind and the sun mainly, in addition to others sources such as tidal, biomass, geothermal, etc. This set of sources, when implemented for the production of electrical energy and also for its integration into existing networks, must have aspects such as in-depth knowledge of the physical resource, its availability and its possible contribution to meet the future energy needs of the region. By including these types of generation in the matrix (solar and wind generation), the benefits they bring are quite attractive for current governments since they reduce dependence on conventional sources and obtaining the fuel necessary for operation in the case of thermal plants. In addition to providing the benefits that are already known at the environmental level, they reduce greenhouse gas emissions, improve the quality of the air that is breathed in the world and reduce water consumption per unit of electricity produced. The model developed in this research in this regard to analyse the implications of predicting the utilization of renewable energy resources showed the following results for different countries:

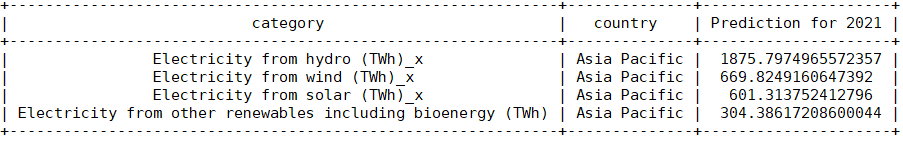
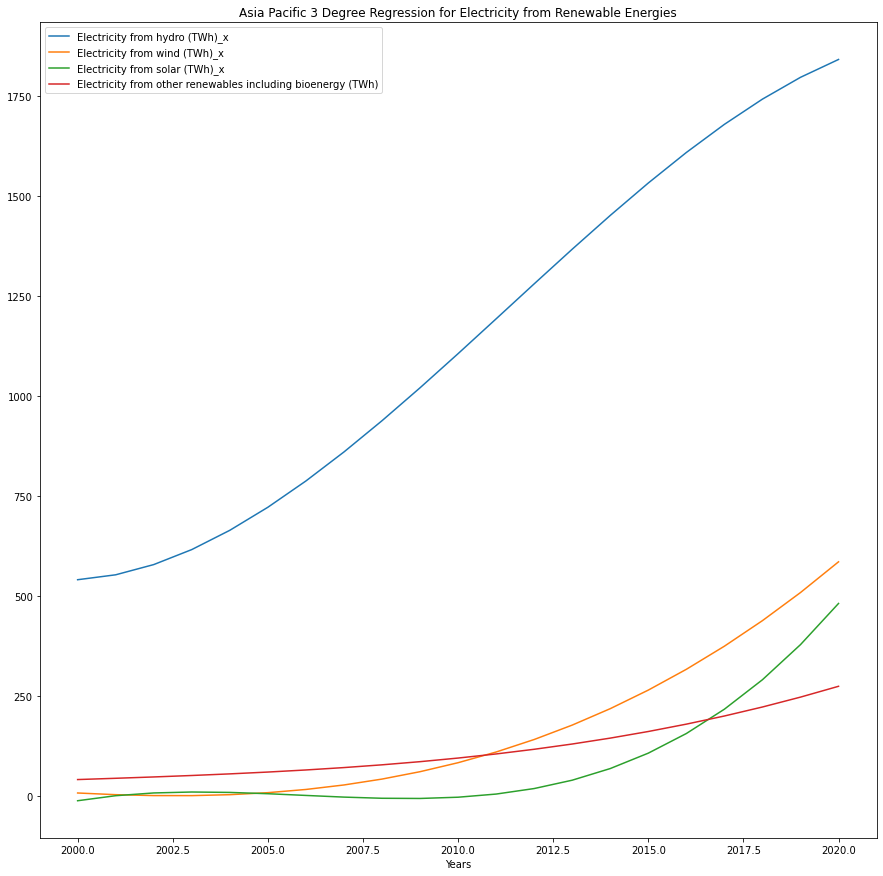
# Africa:

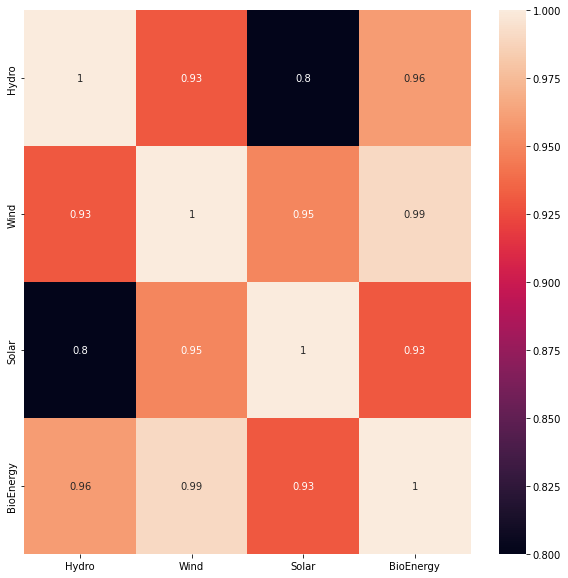




A correlation value of over 0.85 in the heatmap indicates a strong correlation exists between them meaning that if one sector is invested in heavily, the other gets heavy investment too, which in turn generates more energy from that sector.

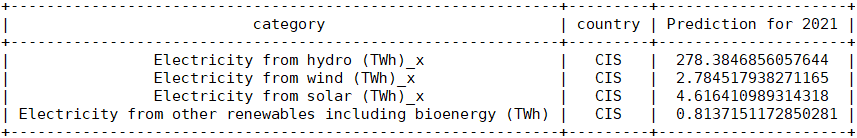
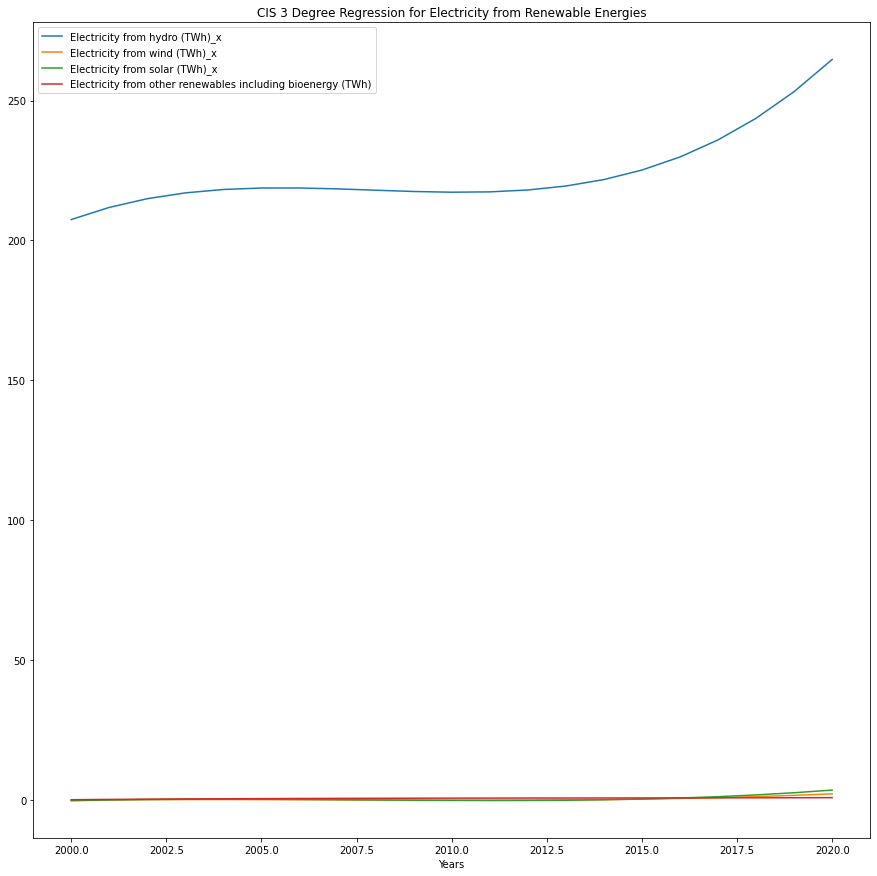
# Asia Pacific:

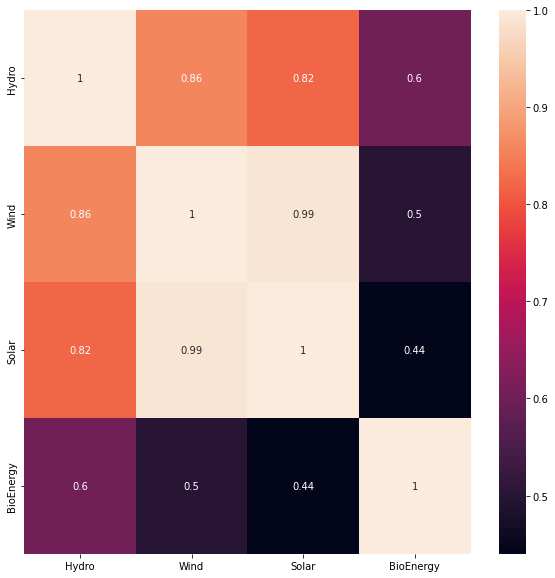




A correlation value of over 0.85 in the heatmap indicates a strong correlation exists between them meaning that if one sector is invested in heavily, the other gets heavy investment too, which in turn generates more energy from that sector.

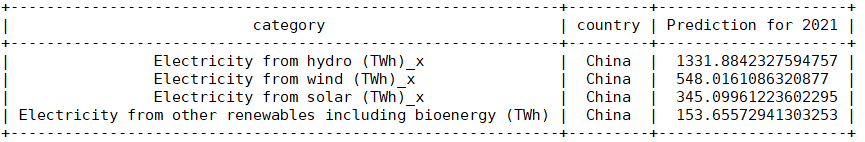
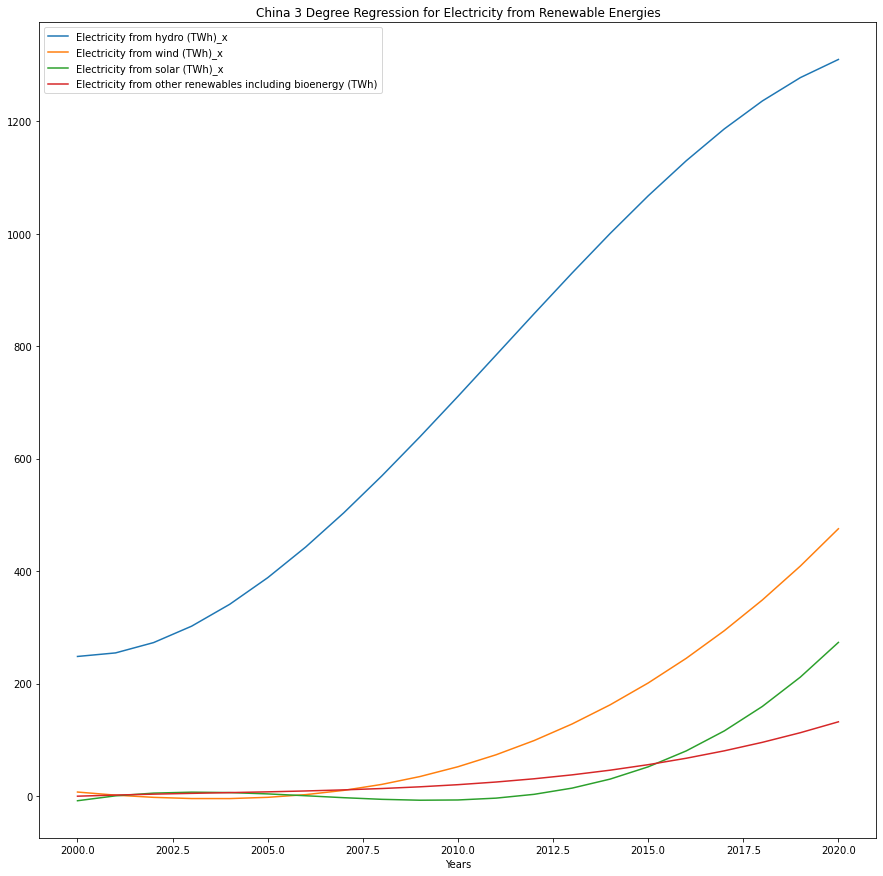
# CIS:

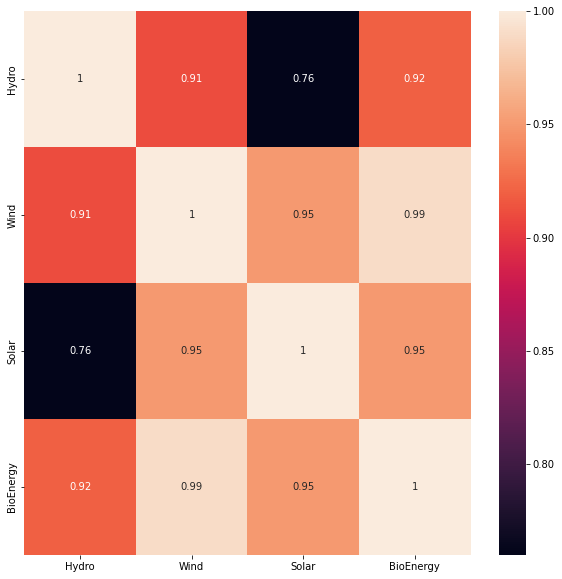




A correlation value of over 0.85 in the heatmap indicates a strong correlation exists between them meaning that if one sector is invested in heavily, the other gets heavy investment too, which in turn generates more energy from that sector.

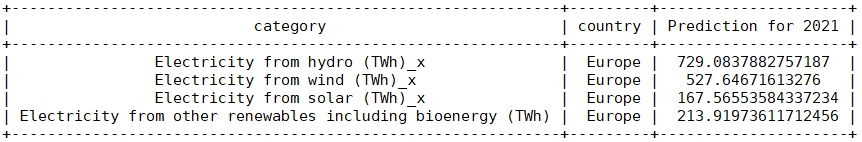
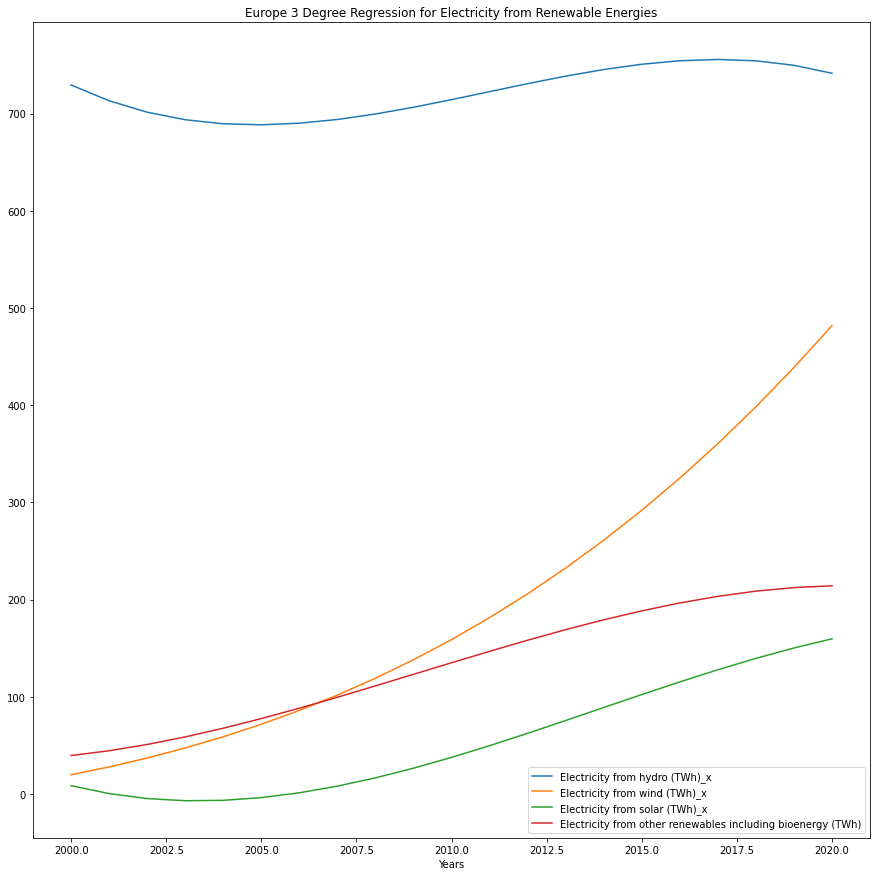
# China:

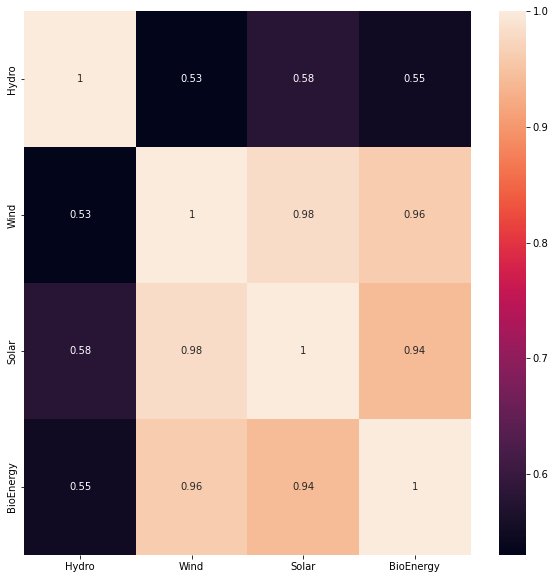




A correlation value of over 0.85 in the heatmap indicates a strong correlation exists between them meaning that if one sector is invested in heavily, the other gets heavy investment too, which in turn generates more energy from that sector.

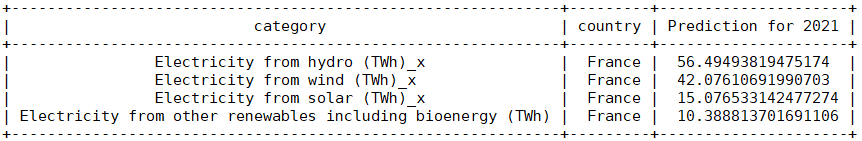
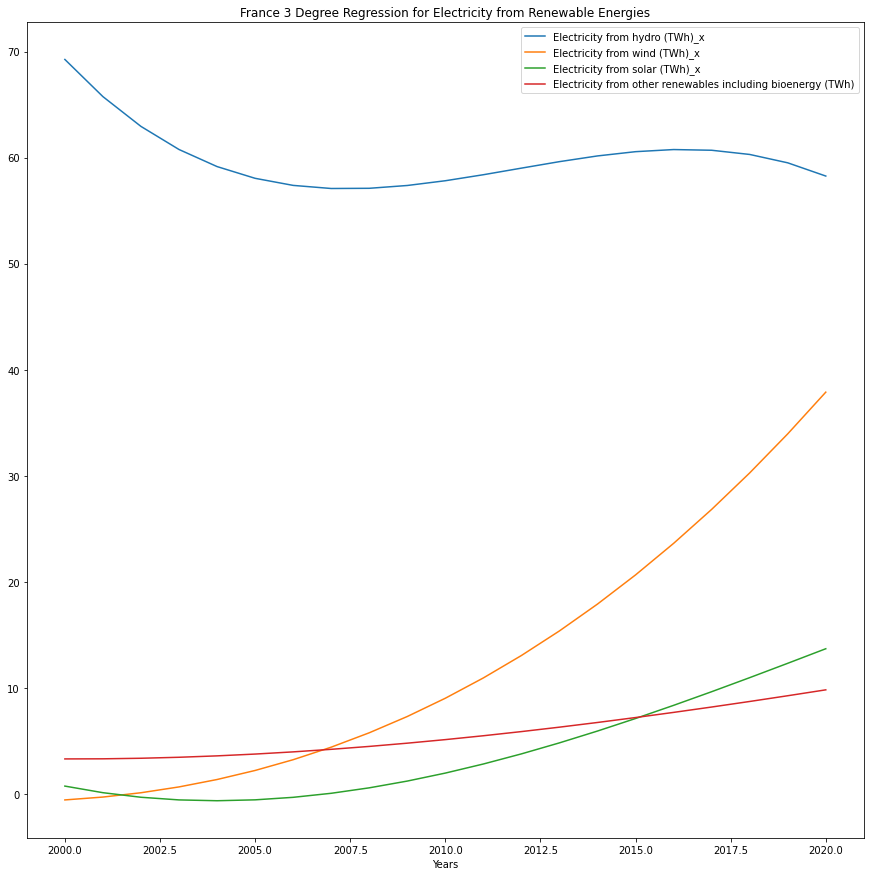
# Europe:

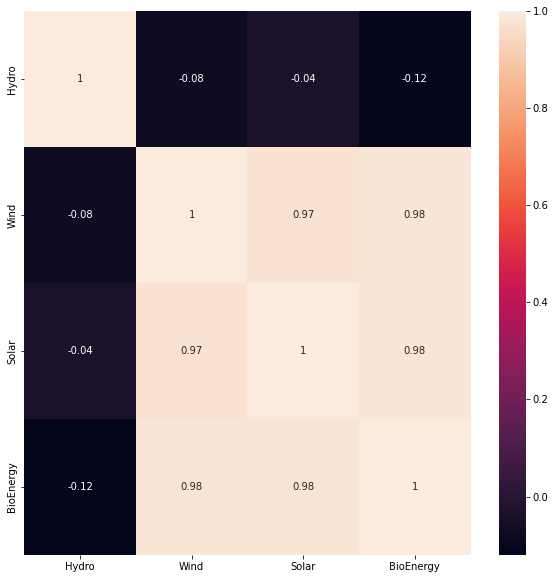




A correlation value of over 0.85 in the heatmap indicates a strong correlation exists between them meaning that if one sector is invested in heavily, the other gets heavy investment too, which in turn generates more energy from that sector.

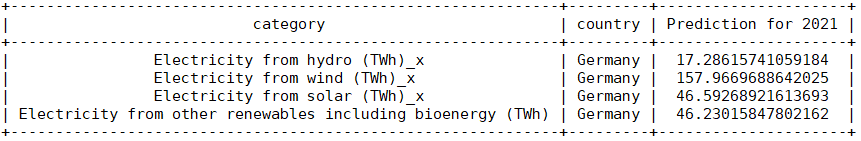
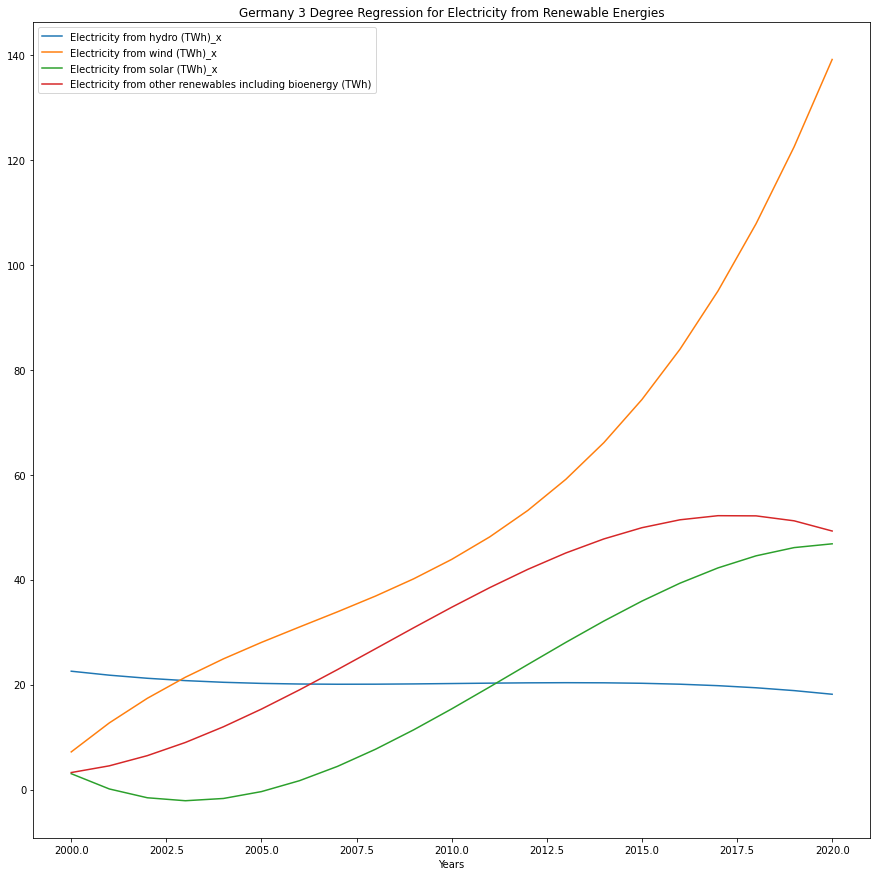
# France:

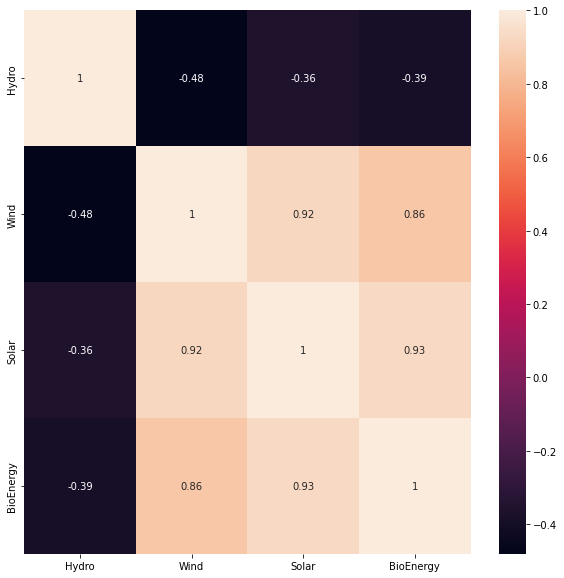




A correlation value of over 0.85 in the heatmap indicates a strong correlation exists between them meaning that if one sector is invested in heavily, the other gets heavy investment too, which in turn generates more energy from that sector.

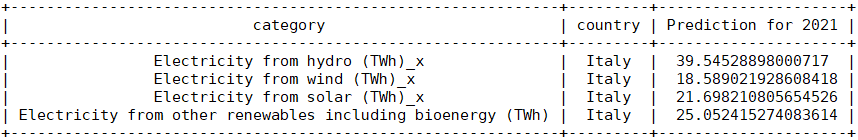
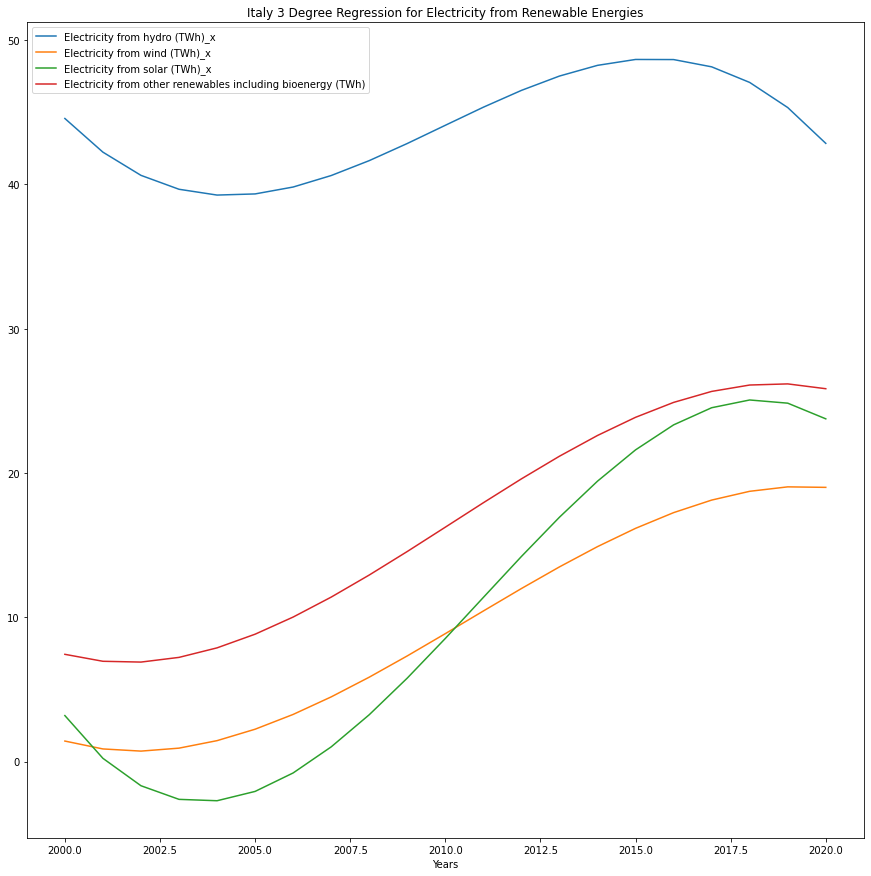
# Germany:

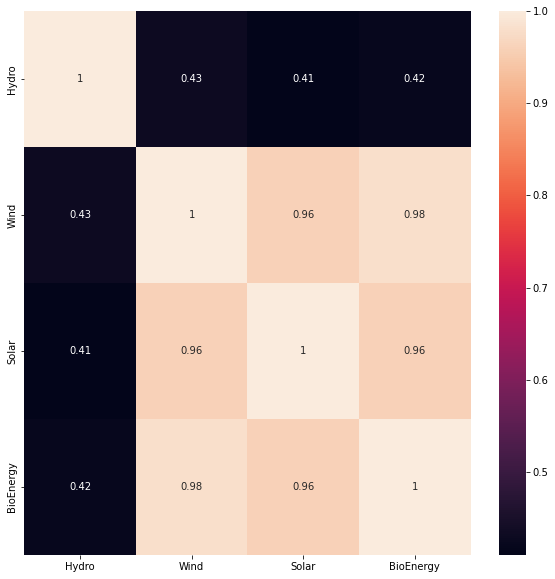




A correlation value of over 0.85 in the heatmap indicates a strong correlation exists between them meaning that if one sector is invested in heavily, the other gets heavy investment too, which in turn generates more energy from that sector.

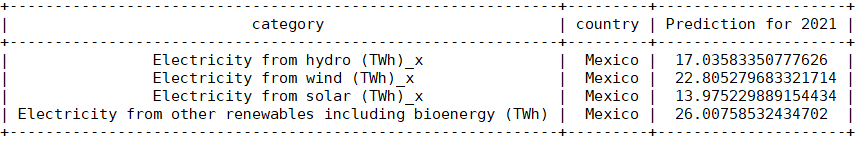
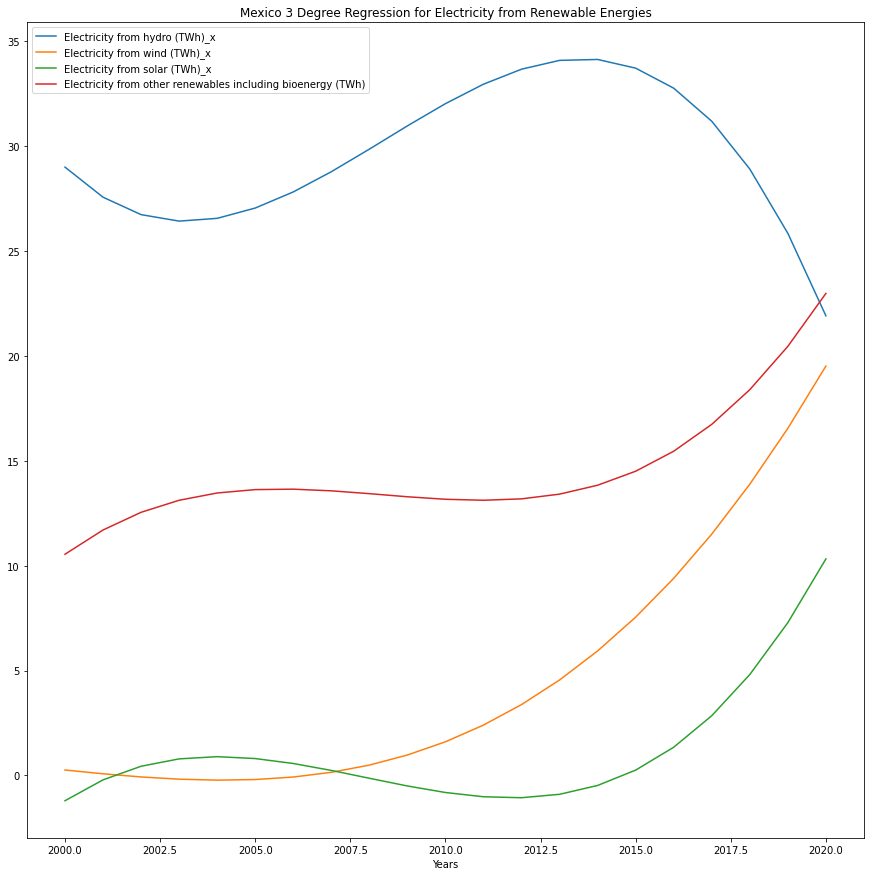
# Italy:

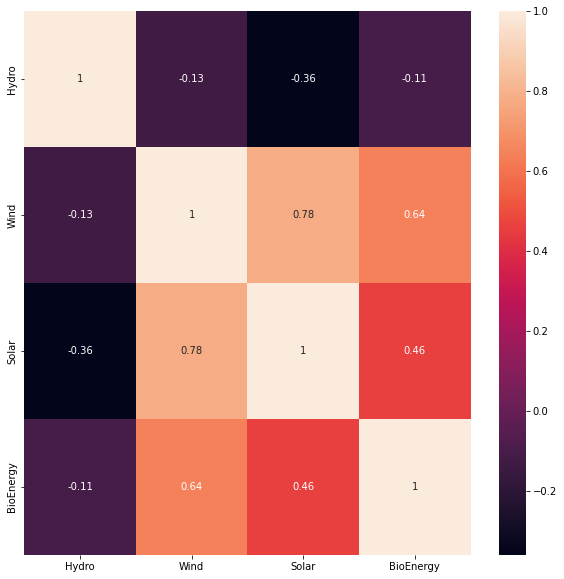




A correlation value of over 0.85 in the heatmap indicates a strong correlation exists between them meaning that if one sector is invested in heavily, the other gets heavy investment too, which in turn generates more energy from that sector.

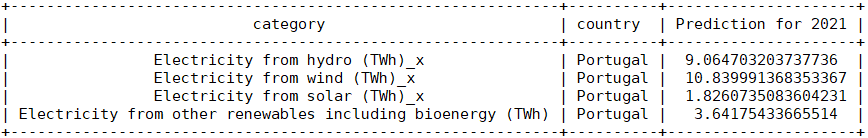
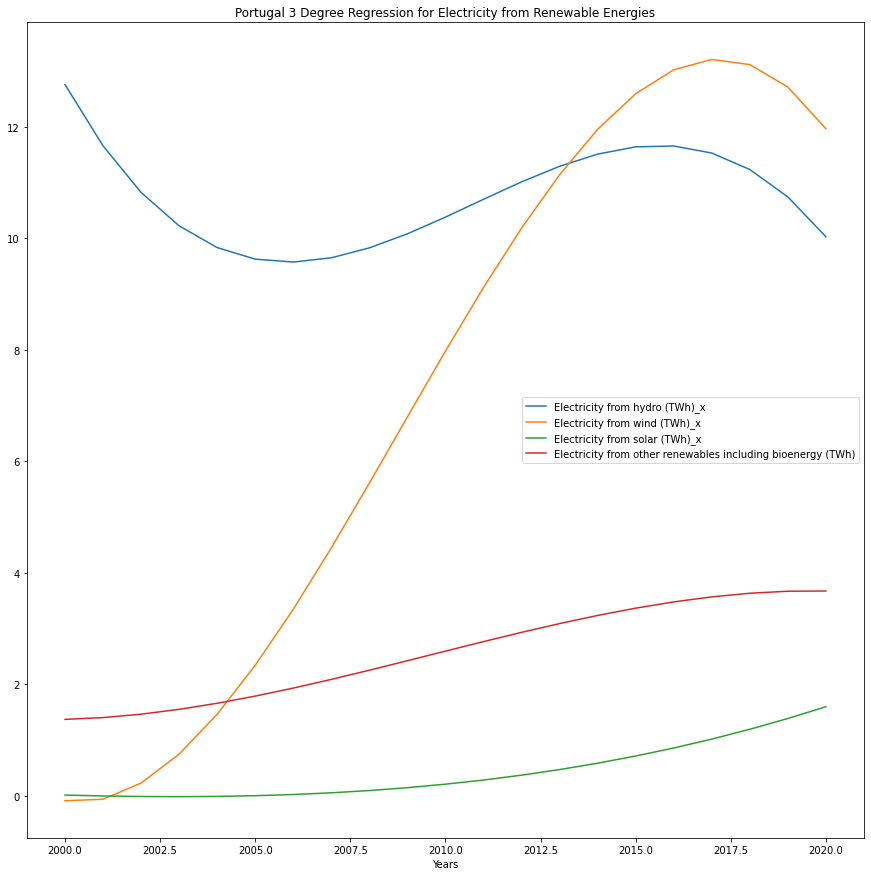
# Mexico:

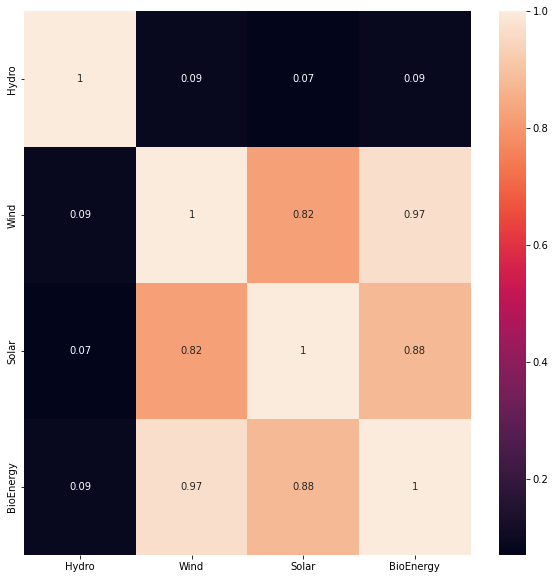




A correlation value of over 0.85 in the heatmap indicates a strong correlation exists between them meaning that if one sector is invested in heavily, the other gets heavy investment too, which in turn generates more energy from that sector.

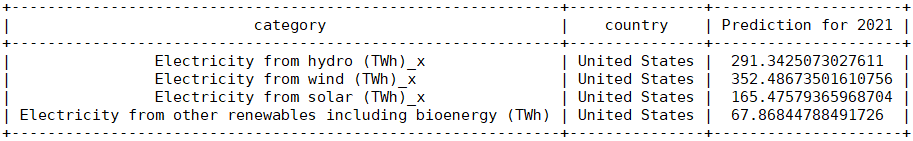
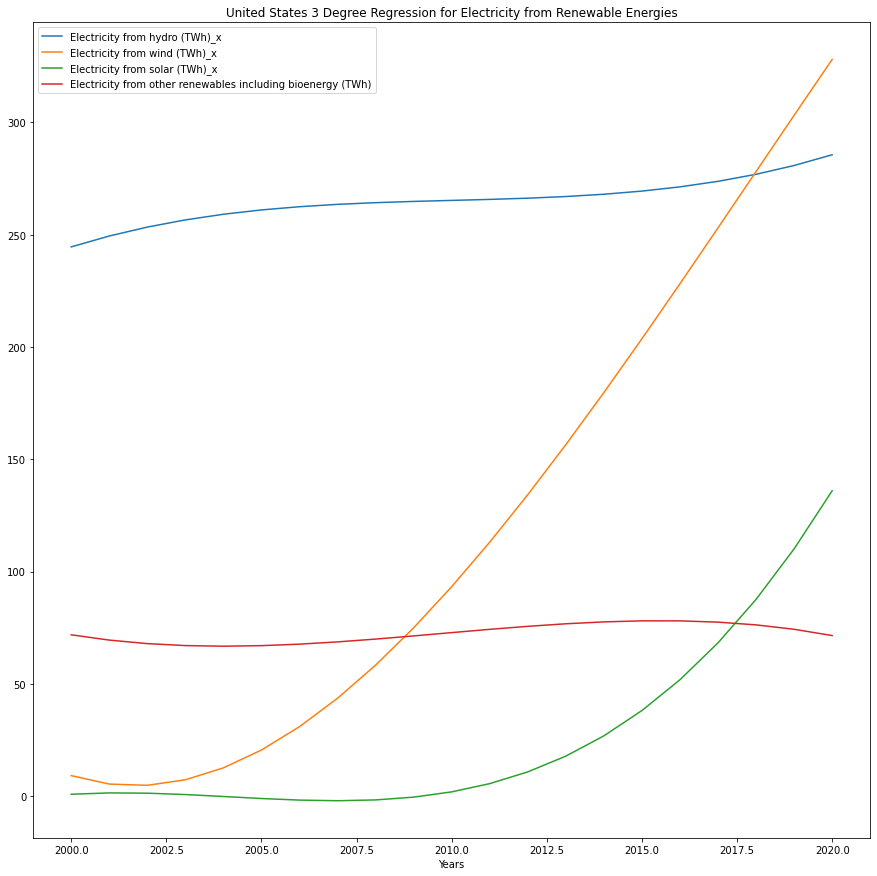
# Portugal:

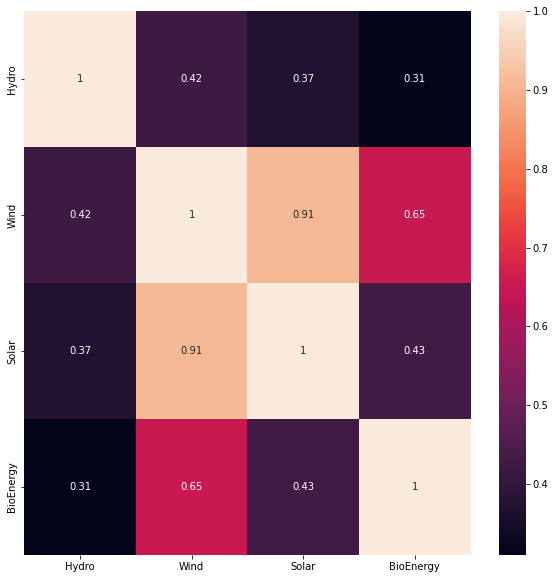




A correlation value of over 0.85 in the heatmap indicates a strong correlation exists between them meaning that if one sector is invested in heavily, the other gets heavy investment too, which in turn generates more energy from that sector.

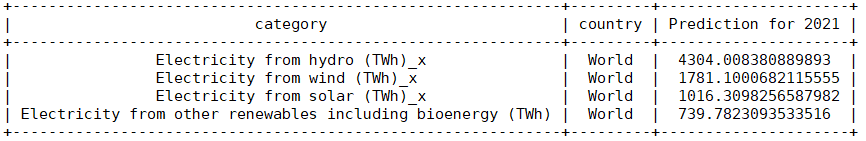
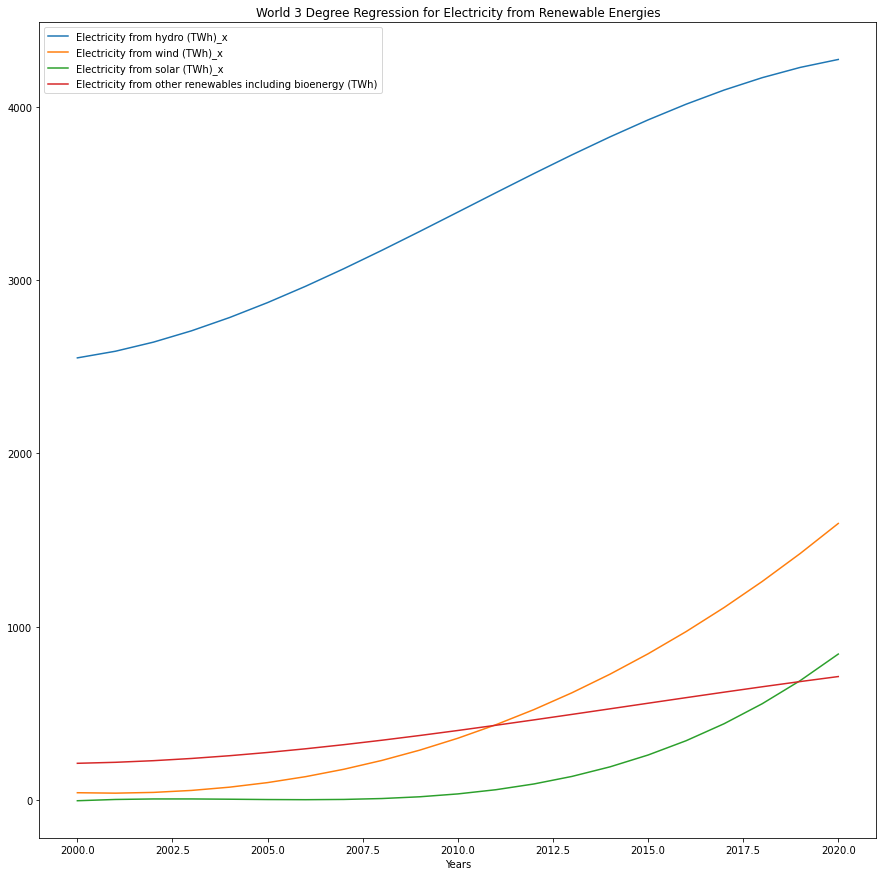
# United States:

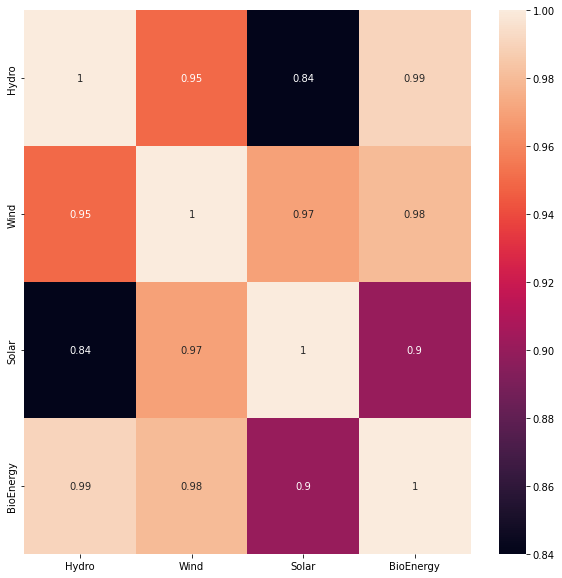




A correlation value of over 0.85 in the heatmap indicates a strong correlation exists between them meaning that if one sector is invested in heavily, the other gets heavy investment too, which in turn generates more energy from that sector.

# World:





A correlation value of over 0.85 in the heatmap indicates a strong correlation exists between them meaning that if one sector is invested in heavily, the other gets heavy investment too, which in turn generates more energy from that sector.

According to the model obtained, the chosen techniques that are part of the model make the results obtained from the work perform efficiently with respect to the times for its execution, in addition to the values adequately representing the radiation profiles, according to the areas where it was evaluated. With this, it is stated that the model meets the need for which it was created, and allows estimations of the solar potential with great precision and accuracy. Likewise, with the development of the four stages, the performance improves, since, without the use of any of them, the way in which the model allows generating trends and adjustments is impaired, which leads to low performance, taking away its robustness and versatility in the face of data entry with different characteristics. By using the model as a computational tool that helps decision making in different fields of the energy sector, it is intended to be light, fast and practical to be used in any area of development that is interested in primary resource forecasts. Likewise, the model is supported by free access and open source software such as Python®, which has great potential focused on ML and artificial intelligence globally, so that it can be used on any computer and at no cost. In the energy sector, the use of these tools becomes essential when it comes to optimizing conventional processes within the main stages of the electrical energy chain, and with proper use, they can represent great benefits to the participants who, in one way or another, form interact directly or indirectly in this field. From generation to distribution and commercialization, extensive studies are carried out to achieve ever greater benefits, accompanied by the inclusion of new technologies that are more intuitive and efficient than, if incorporated with the branches of science with autonomous improvement capabilities and that take into account safety in all aspects, allow significant advances with shorter times.

With these tools, added to new sources of information, in the generation part, and under the recent model of DG, alternatives are presented for decision making in order to guarantee a precise potential and with low uncertainty that can be generated in certain periods, giving the participants the ability to prepare for possible events that can occur with this type of variable resources and also to be able to establish projections that allow them to know the appropriate moments where, depending on the actor, they can provide their service or not. Taking into account the constant development of technologies and processes in the treatment of information, the possibility arises that these models are fed with data that are more reliable in a range of times and horizons that allow these models to be focused on studies with different purposes. The use of information collectors and receivers in a global way increases the possibilities of finding predictions closer to the real values in smaller instants of time, which allow further characterizing the variability of the resource. At this point, the main challenge comes hand in hand, not only with technological development but with the correct treatment and management that should be given to the information, which also depends on the instruments that help collect the data to feed the processes of the prediction methods.

**Chapter 5 Conclusions**

**5.1. Conclusions**

A hybrid model based on Machine Learning techniques that allows to quantify/estimate the daily use of renewable resources through the Python® programming language, which has a strong behaviour against the high variability of the resource, with the management of multiple inputs and with a good performance for the different study areas. Within the design process, a database was selected according to the needs of the study (NASA), which presents greater robustness and practicality with respect to the others, adapting to the approach proposed for its use in this type of applications, and from this information, the response of the model to the various weather conditions according to each study area was evaluated. Results with low uncertainty and dispersion of the data were obtained, and the model has acceptable precisions that validate the performance in the regions of the world.

The potential of Machine Learning in the electrical sector, and specifically in the generation of electrical energy, was evidenced. Its advantages and benefits contribute to the improvement of practices carried out throughout the electrical energy chain, which provides tools that allow increasingly efficient solutions, which facilitate decision-making regarding the generation capacity of PV systems, in addition to knowing the appropriate times for said system to connect or not to the network, improving the reliability of the electrical system. According to the results obtained, the model provides validated, precise information with low uncertainty, key for generation with variable renewable sources, which is subsequently an essential component to be able to apply it in electricity generation projections.

**5.2. Future Works**

As possible works that can be developed following the purpose of the present, it is proposed to allocate them in different lines of application, which contemplate different types of sampling, representative climates of the different world territories, optics with financial and technical points of view for the systems that use this type of tools, and with the use of information sources that are suitable for the purposes of the studies. In the main stages of the model, such as pre-processing of the data to be worked on, the use of a sampling of the databases that provide an ideal set of data that characterizes the area to be worked on is recommended, to optimize later stages of training and testing that allow to reduce the consumption of computational resources and bring greater performance.

In order to have efficient training and with better results, it is suggested to identify cities or areas that contemplate the majority of the representative climates, so that they can be used in this stage and allow the model to perform adequately in any environment, capturing its radiation profile with great precision, optimizing response and execution times. The use of prediction models not only involves technical aspects, but on the contrary, it presents multiple areas that may be of interest. It is proposed for this case to evaluate the financial aspect from the results obtained by the model, since by being certain of the available potential it is possible to avoid oversizing for the operation of the systems, which translates into cost reduction for its planning and execution.

The development of mobile or portable applications is recommended so that the use of these models is available at any time or place, and to have easy access to the data, giving it versatility and practicality in the dynamics of obtaining the radiation values to take them to the desired applications.

# **References**

Abd El-Aziz, R. M. (2022). Renewable power source energy consumption by hybrid machine learning model. *Alexandria Engineering Journal*, *61*(12), 9447-9455.

Abualigah, L., Zitar, R. A., Almotairi, K. H., Hussein, A. M., Abd Elaziz, M., Nikoo, M. R., & Gandomi, A. H. (2022). Wind, Solar, and Photovoltaic Renewable Energy Systems with and without Energy Storage Optimization: A Survey of Advanced Machine Learning and Deep Learning Techniques. *Energies*, *15*(2), 578. https://doi.org/10.3390/en15020578

Ahmed, W., Ansari, H., Khan, B., Ullah, Z., Ali, S. M., Mehmood, C. A. A., ... & Nawaz, R. (2020). Machine learning based energy management model for smart grid and renewable energy districts. *IEEE Access*, *8*, 185059-185078.

Donti, P. L., & Kolter, J. Z. (2021). Machine Learning for Sustainable Energy Systems. *Annual Review of Environment and Resources*, *46*(1). https://doi.org/10.1146/annurev-environ-020220-061831

Forootan, M. M., Larki, I., Zahedi, R., & Ahmadi, A. (2022). Machine Learning and Deep Learning in Energy Systems: A Review. *Sustainability*, *14*(8), 4832. https://doi.org/10.3390/su14084832

Gu, G. H., Noh, J., Kim, I., & Jung, Y. (2019). Machine learning for renewable energy materials. *Journal of Materials Chemistry A*, *7*(29), 17096-17117.

Hon, B. S. (2020). *A machine learning approach to evaluating renewable energy technology: An alternative LACE study on Solar Photo-Voltaic (PV).* Dspace.mit.edu. https://dspace.mit.edu/bitstream/handle/1721.1/127172/1191626239-MIT.pdf?sequence=1&isAllowed=y

Khan, P. W., Byun, Y. C., Lee, S. J., Kang, D. H., Kang, J. Y., & Park, H. S. (2020). Machine learning-based approach to predict energy consumption of renewable and nonrenewable power sources. *Energies*, *13*(18), 4870.

Khan, P. W., Byun, Y.-C., Lee, S.-J., Kang, D.-H., Kang, J.-Y., & Park, H.-S. (2020). Machine Learning-Based Approach to Predict Energy Consumption of Renewable and Nonrenewable Power Sources. *Energies*, *13*(18), 4870. https://doi.org/10.3390/en13184870

Lai, J. P., Chang, Y. M., Chen, C. H., & Pai, P. F. (2020). A survey of machine learning models in renewable energy predictions. *Applied Sciences*, *10*(17), 5975.

Lanio, K. (2018, November 13). *How Machine Learning, Big Data & AI are Changing Energy*. RapidMiner. https://rapidminer.com/blog/machine-learning-big-data-ai-energy/

Nam, K., Hwangbo, S., & Yoo, C. (2020). A deep learning-based forecasting model for renewable energy scenarios to guide sustainable energy policy: A case study of Korea. *Renewable and Sustainable Energy Reviews*, *122*, 109725.

Perera, K. S., Aung, Z., & Woon, W. L. (2014). Machine Learning Techniques for Supporting Renewable Energy Generation and Integration: A Survey. *Data Analytics for Renewable Energy Integration*, 81–96. https://doi.org/10.1007/978-3-319-13290-7\_7

Rangel-Martinez, D., Nigam, K. D. P., & Ricardez-Sandoval, L. A. (2021). Machine learning on sustainable energy: A review and outlook on renewable energy systems, catalysis, smart grid and energy storage. *Chemical Engineering Research and Design*, *174*, 414-441.

Strubell, E., Ganesh, A., & McCallum, A. (2020, April). Energy and policy considerations for modern deep learning research. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 34, No. 09, pp. 13693-13696).

United Nations. (2021). *What is renewable energy?* United Nations. https://www.un.org/en/climatechange/what-is-renewable-energy

Vennila, C., Titus, A., Sudha, T. S., Sreenivasulu, U., Reddy, N. P. R., Jamal, K., Lakshmaiah, D., Jagadeesh, P., & Belay, A. (2022). Forecasting Solar Energy Production Using Machine Learning. *International Journal of Photoenergy*, *2022*, e7797488. https://doi.org/10.1155/2022/7797488

Wang, H. (2021). Fault Diagnosis in Hybrid Renewable Energy Sources with Machine Learning Approach. *Journal of Trends in Computer Science and Smart technology (TCSST)*, *3*(03), 222-237.

Wang, H., Lei, Z., Zhang, X., Zhou, B., & Peng, J. (2019). A review of deep learning for renewable energy forecasting. *Energy Conversion and Management*, *198*, 111799.

Zitnick, C. L., Chanussot, L., Das, A., Goyal, S., Heras-Domingo, J., Ho, C., ... & Ulissi, Z. (2020). An introduction to electrocatalyst design using machine learning for renewable energy storage. *arXiv preprint arXiv:2010.09435*.