



PROJECT REPORT :

predicting the energy output of wind turbine based on weather condition



PredictiNg the Energy Output of wind turbine based oN weather coNditioN

|  |
| --- |
| **PROJECT REPORT ON**  **“**Predicting the energy output of wind turbine based on weather condition**”** |

**SUBMITTED BY:**

**Name:** Shevya Solanki

**BRANCH:** Computer Science & Engineering

**PROJECT ID:**  SPS\_PRO\_700

**Category:** Machine Learning

**uNder The GUidaNce of:**

**SmartiNterNz**



|  |
| --- |
| **ACKNOwledgemeNt** |

* The, **IBM Hack Challenge 2020** is all about coding for a cause. It’s about finding solutions to problems plaguing society today, the opportunity that I got with **SmartBridge** was a great chance for learning and professional development. Therefore, I consider myself as a very lucky individual as I was provided with an opportunity to be a part of it. I am also extremely grateful for having a supportive mentors that led me through this Hack challenge and helped me a lot to achieve my goal. I have gained the knowledge and the actual practical application of that in a real world problem, that would definitely going to help in my future career.
* I perceive as this opportunity as a big milestone in my career development & I really consider myself lucky for being the part of this Initiative. I will strive to use gained skills and knowledge in the best possible way, and I will continue to work on their improvement, in order to attain desired career objectives. Hope to continue cooperation with all of you in the future.

Thank You.

Shevya Solanki

|  |
| --- |
| **ABSTRACT** |

* Wind energy plays an increasing role in the supply of energy world-wide. The energy output of a wind farm is highly dependent on the wind conditions present at its site. If the output can be predicted more accurately, energy suppliers can coordinate the collaborative production of different energy sources more efficiently to avoid costly overproduction.
* Wind energy plays an increasing role in the supply of energy world-wide. The energy output of a wind farm is highly dependent on the weather conditions present at its site. If the output can be predicted more accurately, energy suppliers can coordinate the collaborative production of different energy sources more efficiently to avoid costly overproduction. In this paper, we take a computer science perspective on energy prediction based on weather data and analyze the important parameters as well as their correlation on the energy output. To deal with the interaction of the different parameters, we use symbolic regression based on the genetic programming tool Data Modeler.
* Our studies are carried out on publicly available weather and energy data for a wind farm in Australia. We report on the correlation of the different variables for the energy output. The model obtained for energy prediction gives a very reliable prediction of the energy output for newly supplied weather data renewable energy such as wind and solar energy plays an increasing role in the supply of energy world-wide. This trend will continue because the global energy demand is increasing and the use of nuclear power and traditional sources of energy such as coal and oil is either considered as non-safe or leads to a large amount of CO mission.
* Wind energy is a key-player in the ﬁeld of renewable energy. The capacity of wind energy production was increased drastically during the last years. In Europe for example, the capacity of wind energy production has doubled since 2005. However, the production of wind energy is hard to predict as it relies on the rather unstable weather conditions present at the wind farm. In particular, the wind speed is crucial for energy production based on wind and the wind speed may vary drastically during diﬀerent periods of time. Energy suppliers are interested in accurate predictions, as they can avoid overproductions by coordinating the collaborative production of traditional power plants and weather dependent energy source.

|  |
| --- |
| **CONTENT Table** |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **INT 1. INTRODUCTION** | | | | |  | | |
|  |  | | 1.1 Overview | | | |  | | |
|  |  | | 1.2 Purpose | | | |  | | |
| **LITER 2. LITRATURE SURVEY** | | | | |  | | |
|  |  | | 2.1 Existing problem | | | | | | |
|  |  | | 2.2 Proposed solution | | | | | | |
| **3.THEORITICAL ANALYSIS** | | | | |  | | |
|  |  | | 3.1 Block diagram | | | |  | | |
|  |  | | 3.2 Hardware / Software designing | | | | | | |
|  | | | | | | | |
| **4. FLOWCHART** | | | | |  | | |
| **5. RESULT** | | | |  |  | | |
| **6. ADVANTAGES & DISADVANTAGES** | | | | | | | | |
| **7. APPLICATIONS** | | | | | |  | | |
| **8. CONCLUSION** | | | | | |  | | |
| **9. FUTURE SCOPE** | | | | | | **`** | | |
| **10. REFERENCE** | | | | | |  | | |
|  |  |  | | | | |  | | |

|  |
| --- |
| **1. Introduction** |

**1.1 OVERVIEW**

* ***Machine learning*** is a sub-domain of computer science which evolved from the study of pattern recognition in data, and also from the computational learning theory inartificial intelligence. It is the first-class ticket to most interesting careers in data analytics today. As data sources proliferate along with the computing power to process them, going straight to the data is one of the most straightforward ways to quickly gain insights and make predictions. Machine Learning can be thought of as the study of a list of sub-problems, viz: **decision making, clustering, classification, forecasting, deep-learning, inductive logic programming, support vector machines, reinforcement learning, similarity and metric learning, genetic algorithms, sparse dictionary learning,** etc.

**I**n case of Predicting the energy output of wind turbine based on weather condition, although there have been lot of studies undertaken in the past on factors affecting **Date/Time** , LV **ActivePower (kW):** The power generated by the turbine for that moment, **Wind Speed (m/s):** The wind speed at the hub height of the turbine (the wind speed that turbine use for electricity generation), **Theoretical*Power*Curve (KWh):** The theoretical power values that the turbine generates with that wind speed which is given by the turbine manufacturer, **Wind Direction (°):** The wind direction at the hub height of the turbine (wind turbines turn to this direction automaticly)

* It was found that affect of immunization and human development index was not taken into account in the past. Also, some of the past research was done considering multiple linear regression based on data set of one year for all the countries. Hence, this gives motivation to resolve both the factors stated previously by formulating a Regression Model based on mixed effects model and **Multiple Linear Regression** while considering data from a period of 2000 to 2015 for all the countries. Important immunization like Hepatitis B, Polio and Diphtheria will also be considered. In a nutshell, this study will focus on immunization factors, mortality factors, economic factors, social factors and other health related factors as well. Since the observations this dataset are based on different countries, it will be easier for a country to determine the predicting factor which is contributing to lower value of life expectancy. This will help in suggesting a country which area should be given importance in order to efficiently improve the life expectancy of its population.
* n Wind Turbines, Scada Systems measure and save data's like wind speed, wind direction, generated power etc. for 10 minutes intervals. This file was taken from a wind turbine's scada system that is working and generating power in Turkey.
* Linear Regression  
  Ridge Regression  
  Lasso Regression  
  Elastic Net Regression  
  Decision Tree Regression  
  Random Forest Regression
* Wind Speed Prediction Models Wind forecasting models can be broadly classified into the following three categories: (i) physical model; (ii) statistical and computational model; and (iii) hybrid model, The artificial intelligence approach belongs to the statistical approach. The essence of the artificial intelligence approach is to establish the relationship between input and output by artificial intelligence methods, rather than using the analytical method. The model described in this form is usually a non-linear model. Many methods of artificial intelligence are better than conventional methods and have a good perspective of development . Energies 2017, 10, 1976 4 of 27 2.4. Statistical Models and Artificial Neural Networks Statistical models are easy to use and cheaper to develop compared to other models. Basically, statistical methods use the previous history of wind data to perform a forecast over the next few hours, they are good for short periods of time. The disadvantage of this method is that prediction error increases as time forecasting increases, i.e., statistical time series and methods of neural networks are primarily intended for short-term predictions . According to the sub classification of this approach can be defined as: models based on time series and methods based on neural networks. These forecasting methods are generally used for small forecast horizons (mesoscale), because, in these horizons, the correlation between the velocities of the winds, and consequently the generation, are greater. The statistical models most disseminated by researchers include: auto regressive (AR), auto regressive moving average (ARMA), and auto regressive integrated moving average (ARIMA). Statistical methods, in many predictions, use the difference between predicted and actual wind speeds to adjust model parameters. The advantage of Artificial Neural Networks (ANNs) is to know the relation between inputs and outputs by a non-statistical approach. According to, neural networks can easily learn from the input and output mapping during the training phase. This allows the neural models to perform well, even without the researchers’ knowledge of the problem. The advantage related to the other methods is to provide relatively inexpensive models of statistical projection that do not require any data other than historical wind power generation data. However, the prediction accuracy for these models falls significantly when the time horizon is extended. In it is presented a new statistical method based on the AR model and analysis of independent components. Based on the results obtained, the proposed method obviously gives a greater precision compared to direct predictions.

**1.2 PURPOSE OF PROJECT**

* Our aim is to map weather data to energy production. We want to show that even data that

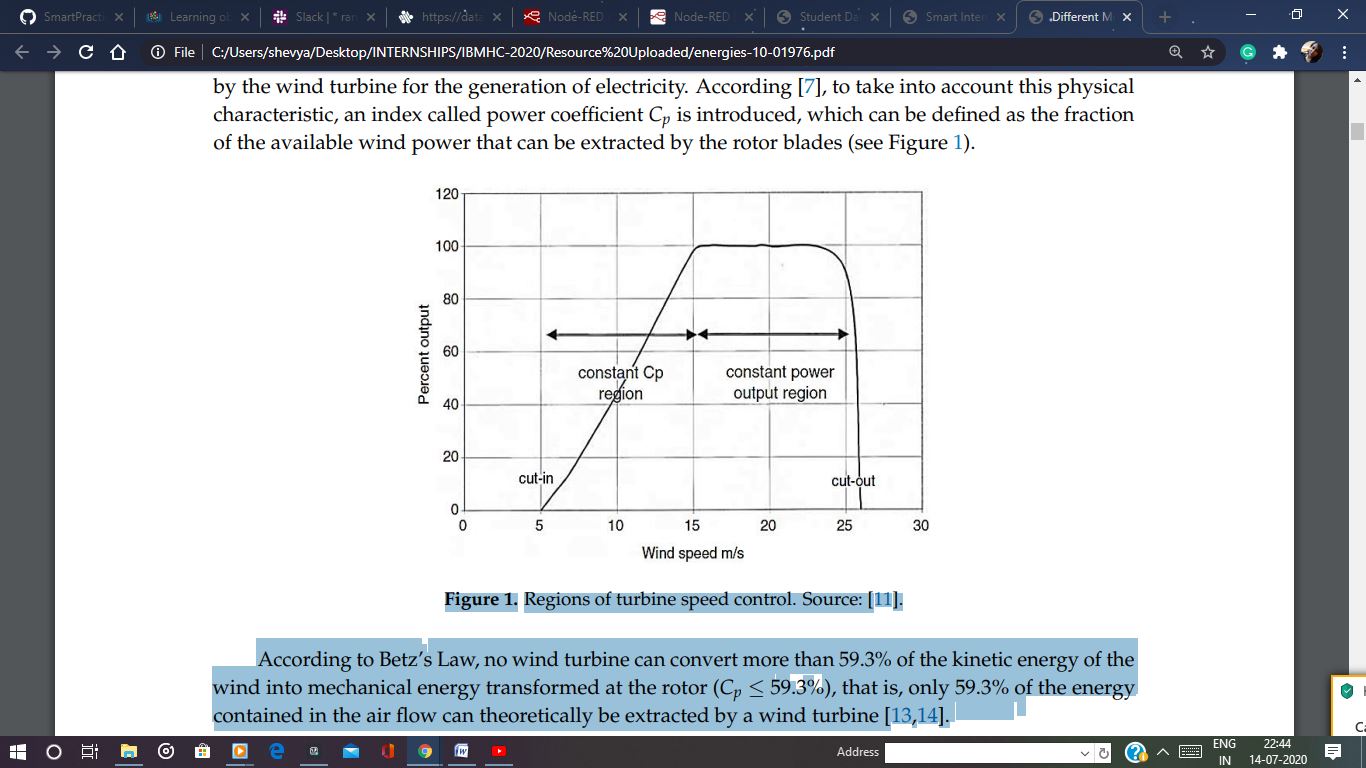
is publicly available for weather stations close to wind farms can be used to give a good prediction of the energy output. Furthermore, we examine the impact of diﬀerent weather conditions on the energy output of wind farms. We are, in particular, interested in the correlation of diﬀerent components that characterize the weather conditions such as wind speed, pressure, and temperature .A good overview on the diﬀerent methods that were recently applied in forecasting of wind power generation can be found in .Statistical approaches use historical data to predict the wind speed on an hourly basis or to predict energy output directly. On the other hand, short term prediction is often done based on meteorological data and learning approaches are applied. Kusiak, Zheng, and Song [8] have shown how wind speed data may be used to predict the power output of a wind farm based on times series prediction modeling.

* Neural networks are a very popular learning approach for wind power forecasting based on given time series. They provide an implicit model of the function that maps the given weather data to an energy output. Jursa and Rohrig have used particle swarm optimization and diﬀerential evolution to minimize the prediction error of neural networks for short-term wind power forecasting. Kramer and Gieseke used support vector regression for short term energy forecast and kernel methods and neural networks to analyze wind energy time series . These studies are all based on wind data and do not take other weather conditions into account. Furthermore, neural networks have the disadvantage that they give an implicit model of the function predicting the output, and these models are rarely accessible to a human expert. Usually, one is also interested in the function itself and the impact of the diﬀerent variables that determine the output. We want to study the impact of diﬀerent variables on the energy output of the wind farm. Surely, the wind speed available at the wind farm is a crucial parameter. Other parameters that inﬂuence the energy output are for example air pressure, temperature and humidity. Our goal is to study the impact and correlation of these parameters with respect to the energy output.

.

|  |
| --- |
| **2. LITERATURE SURVEY** |

* 1. **EXISTINg PROBLEM**
* Wind Power Generation Potential The potential of electric energy produced from wind generation is obtained through the kinetic energy of the wind, which is converted into mechanical energy by a process that turns the wind force into a torque that acts on the rotor blades. The amount of energy generated by winds is a function of their speed (v) and mass (m) and is given by the kinetic energy equation thus, it is very important to make a good wind speed prediction. The power available in the wind, however, cannot be fully utilized by the wind turbine for the generation of electricity. According to take into account this physical characteristic, an index called power coefficient Cp is introduced, which can be defined as the fraction of the available wind power that can be extracted by the rotor blades .

****

* According to Betz’s Law, no wind turbine can convert more than 59.3% of the kinetic energy of the wind into mechanical energy transformed at the rotor (Cp ≤ 59.3%), that is, only 59.3% of the energy contained in the air flow can theoretically be extracted by a wind turbine. Figure 1. Regions of turbine speed control. Source. According to Betz’s Law, no wind turbine can convert more than 59.3% of the kinetic energy of the wind into mechanical energy transformed at the rotor (Cp ≤ 59.3%), that is, only 59.3% of the energy contained in the air flow can theoretically be extracted by a wind turbine .

The potential of electric energy produced from wind generation is obtained through the kinetic energy of the winds, which is converted into mechanical energy by a process that turns the wind into torque acting on the rotor blades. According to the power curve regions can be described as follows:

• Optimum constant Cp region, where increasing power with increasing wind speed;

• Limited power region, generating a constant power, even in higher winds, by decreasing the Cp rotor efficiency; and

• Region of power shutdown, where power generation is decelerated to zero, and wind speed approaches the cut-out limit. It is emphasized that wind speed prediction plays a vital role in the management, planning and integration of the energy system. In previous studies, most forecasting models have focused on improving the accuracy or stability of wind speed prediction. However, for an effective forecast model, considering only one criterion (precision or stability) is insufficient. A new design methodology to predict wind farm energy production by means of a spiking neural network-based system is developed. Authors established that the calculation of the flow around wind turbines is a very complicated issue. They affirmed that turbine wakes are responsible for important power losses in wind farms and that randomness of wind speed and unexpected variations of wind speed may increase operating costs of the electricity grid as well as set potential threats to the reliability of electricity supply. A synergetic neural network (SNN)-based model for the prediction of wind farm energy production is developed. This model performs a prediction of the energy produced by one wind turbine of the wind farm, by using the wind speed and direction data coming from the three anemometric towers, during the whole day. This is a very useful model for analyzing turbines inside of a farm. Authors demonstrate that this model could accurately predict the energy produced by each wind turbine of the wind farm by means of testes conducted with experimental data that show low values of median absolute deviation (MAD), mean absolute percentage error (MAPE) and root mean square error (RMSE).

* 1. **Proposed solution**
* Four algorithms have been used:

Linear Regression  
Ridge Regression  
Lasso Regression  
Decision Tree Regression  
Random Forest Regression

The data-set aims to answer the following key questions:

* We collected both the weather and energy production data for the time window September 2010 till July 2011. The output of the farm is available with a rate of one measurement every ﬁve minutes, and the weather data with a rate of one measurement every 30 minutes.The wind farm’s production capacity is split into two sites, which complicated the generatio of models. The site ”Studland Bay” has a maximum output of 75 MW, and ”Bluﬀ Point” has a maximum output of 65 MW and is located 50km south of the ﬁrst site. The weather station is located on the ﬁrst site. For wind coming from west (which is the prevailing wind direction), the diﬀerence in location is negligible. But if wind comes from north, there will be an energy and wind increase right away, plus another energy increase 1-2 hours later (the time delay depends on the actual wind speed). Similarly, if wind comes from south, there will be an increase in the energy production (although no wind is indicated by the weather station) and then 1-2 hours later an energy increase accompanied by a measured wind speed increase

|  |
| --- |
| **3. THEORETICAL Analysis** |

**3. 1 PROJECT REQUIREMENT :**

* + 1. **FUNCTIONAL REQUIREMENT :**
* Predicting the energy output of wind turbine based on weather condition

**3.1.2 TECHNICAL REQUIREMENTS :**

* Python, IBM Cloud, IBM Watson, Github , Node Red in IBM Watson, Jupyter Notebook Watson.

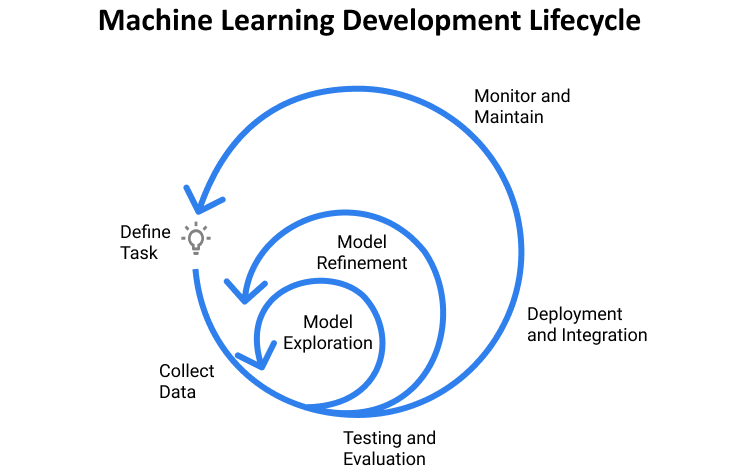
**3.2 HARDWARE  REQUIREMENT :**

*Processor -*i5 7th Gen

*Speed -*2GHz or more

*Hard Disc -*10 Gb or more

*3.*3 HARDWARE SOFTWARE DESIGN



1. [**Planning and project setup**](https://www.jeremyjordan.me/ml-projects-guide/#planning)

a. Define the task and scope out requirements.

b. Determine project feasibility.

c. Setting up project codebase.

**TOOLS**- IBM Watson Studio, Node-Red workflows, IBM cloud.

2. [**Data collection and labeling**](https://www.jeremyjordan.me/ml-projects-guide/#data)

<https://www.kaggle.com/berkerisen/wind-turbine-scada-dataset>

a. Define ground truth (create labeling documentation)

b. Build data model

c. Validate quality of data

**TOOLS -**Excel SpreadSheets , CSV files.

3. [**Model exploration**](https://www.jeremyjordan.me/ml-projects-guide/#exploration)

a. Establish baselines for model performance.

b. Overfit simple model to training data.

c. Using various Libraries to explore data.

**TOOLS-**Matplotlib, Sklearn, Seaborn, Keras, Scikit-Learn.

4. [**Model refinement**](https://www.jeremyjordan.me/ml-projects-guide/#refinement)

a. Perform model-specific optimizations (Hyper parameter tuning)

b. Iteratively debug model as complexity is added

5. [**Testing and evaluation**](https://www.jeremyjordan.me/ml-projects-guide/#testing)

a. Evaluate model on test distribution; understand differences between train and test set distributions (how is “data in the wild” different than what you trained on)

b. Revisit model evaluation metric; ensure that this metric drives

6. [**Model deployment**](https://www.jeremyjordan.me/ml-projects-guide/#deployment)

a. Expose model via a REST API, IBM Watson, Node-Red Kit.

b. Deploy new model to small subset of users to ensure everything goes smoothly, then roll out to all users

c. Monitor live data and model prediction distributions

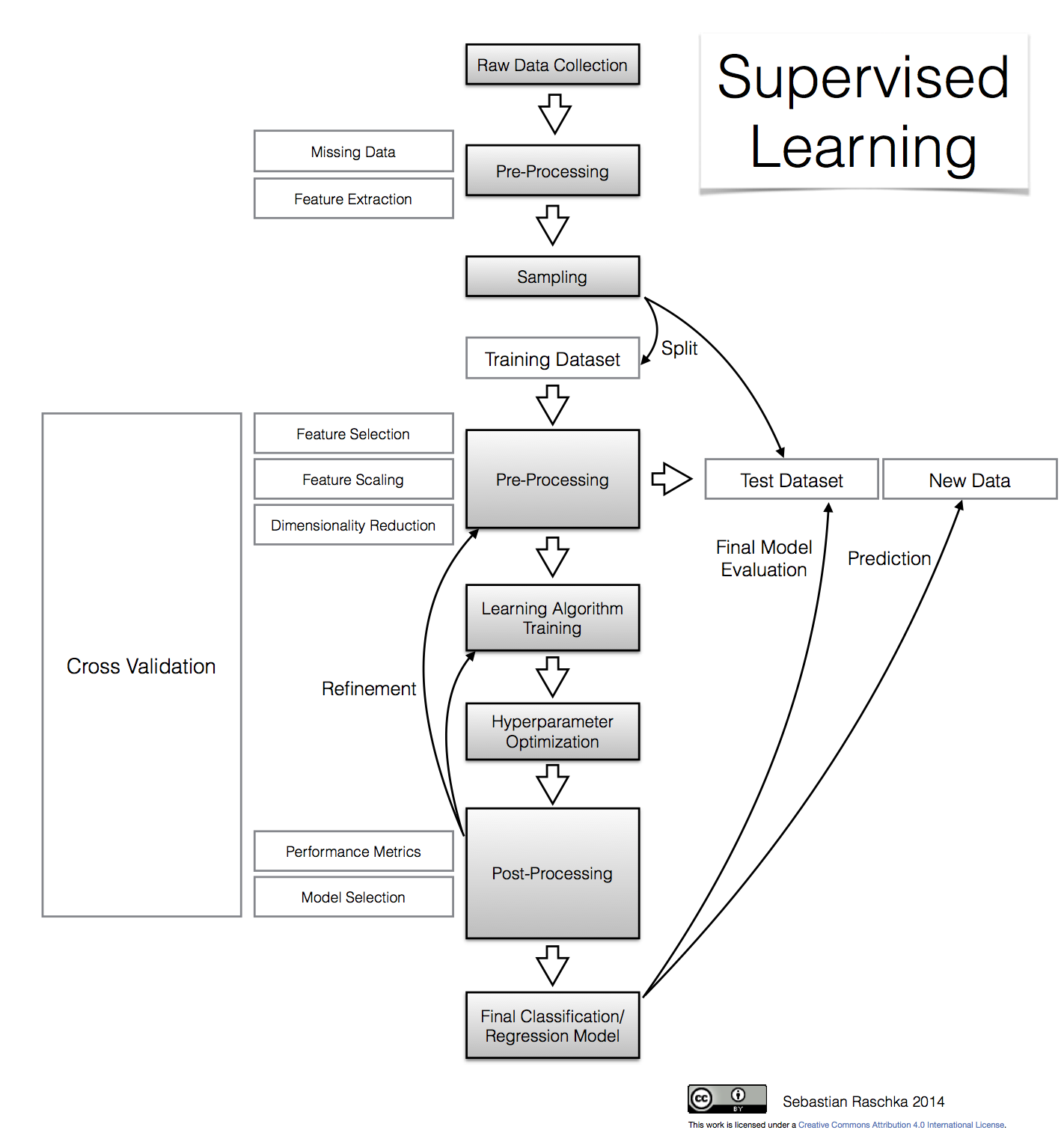
7. [**Ongoing model maintenance**](https://www.jeremyjordan.me/ml-projects-guide/#maintenance)

a. Understand that changes can affect the system in unexpected ways

b. Periodically retrain model to prevent model staleness

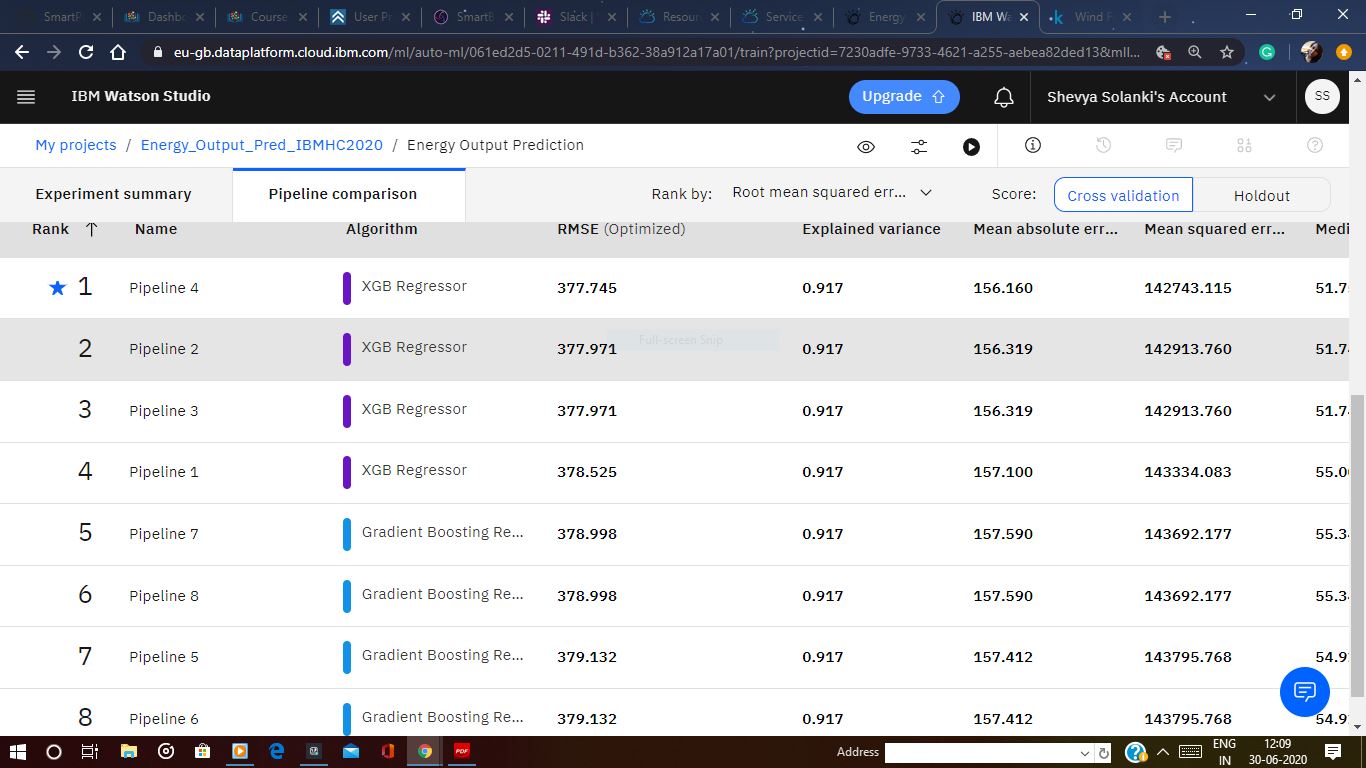
c. If there is a transfer in model ownership, educate the new team

|  |
| --- |
| **4. FLOWCHART** |

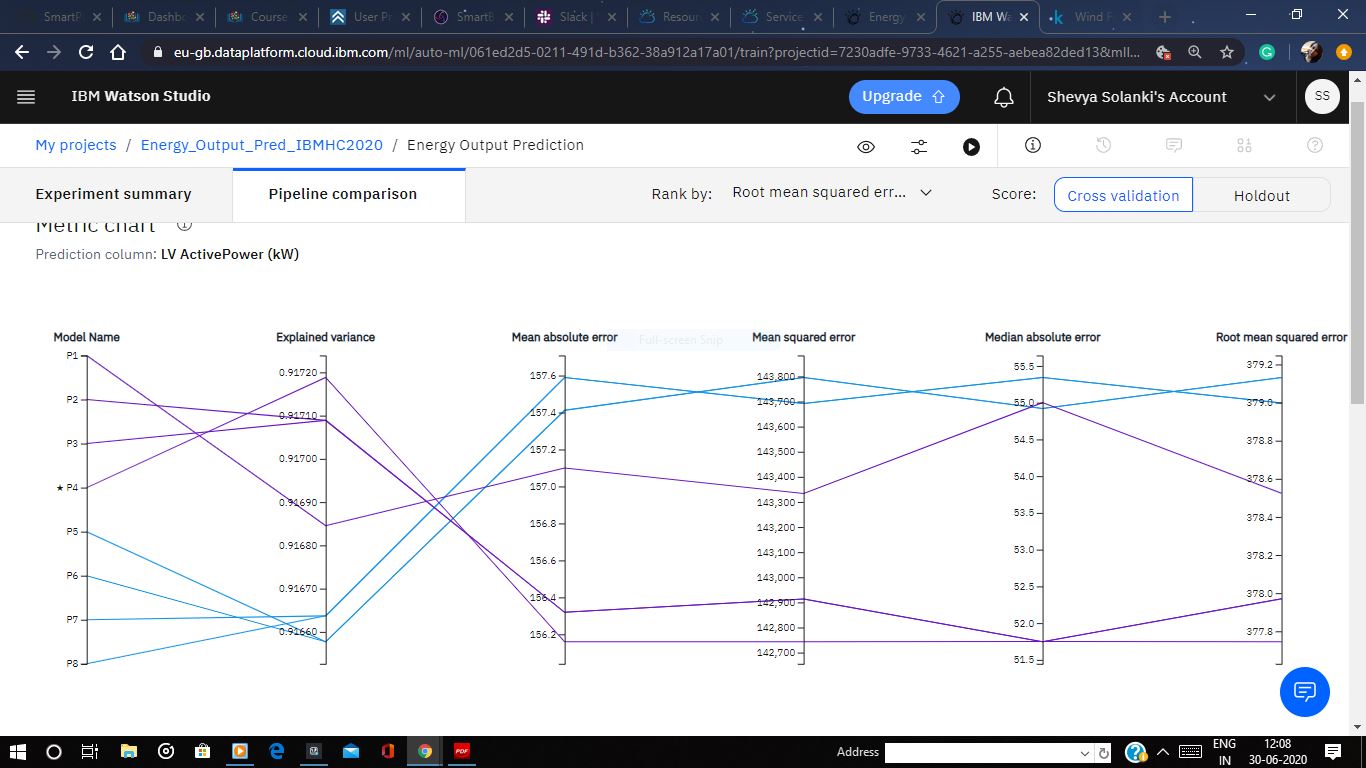
****

|  |
| --- |
| **5. Result** |

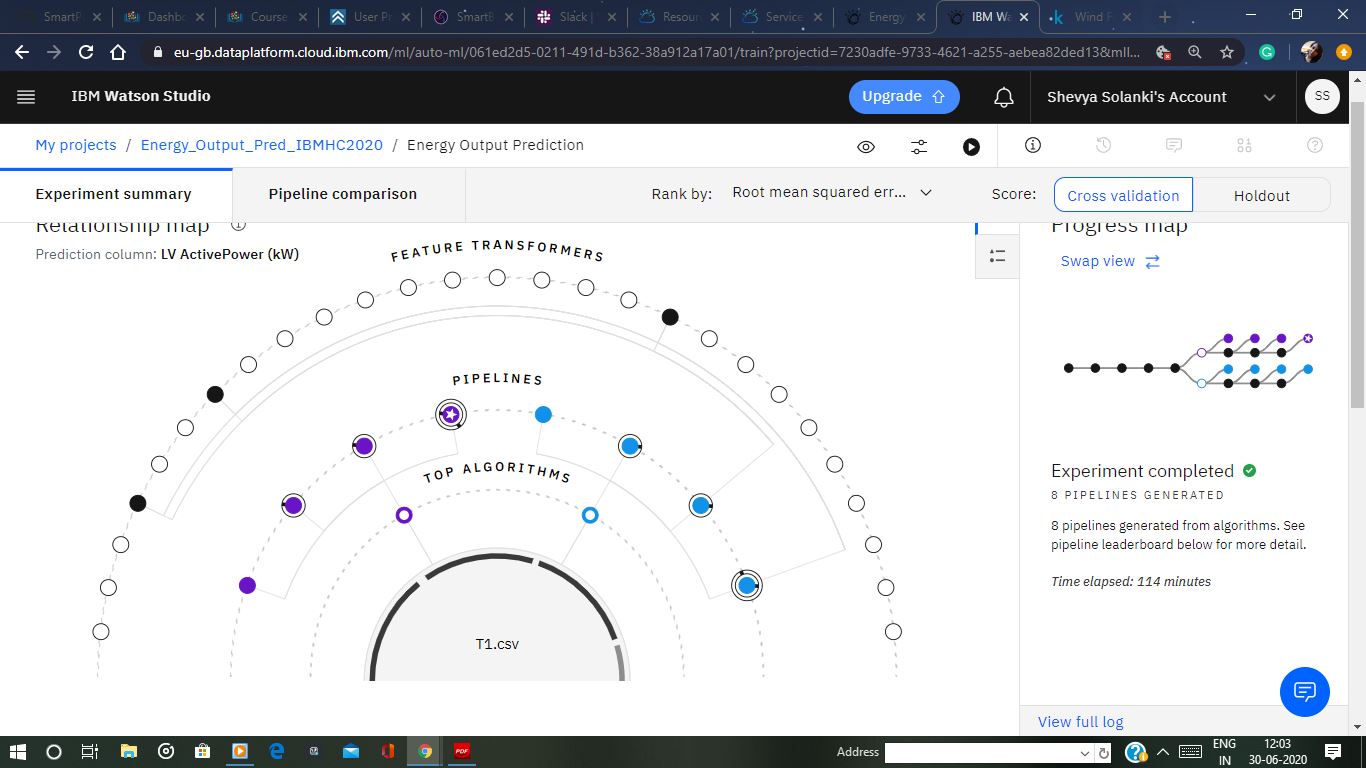
Below given is the Pipelines used on the WHO data set to predict the outcome of the Life- Expectancy data collected. Out of all models Extra Tree Regressor was the pipeline.

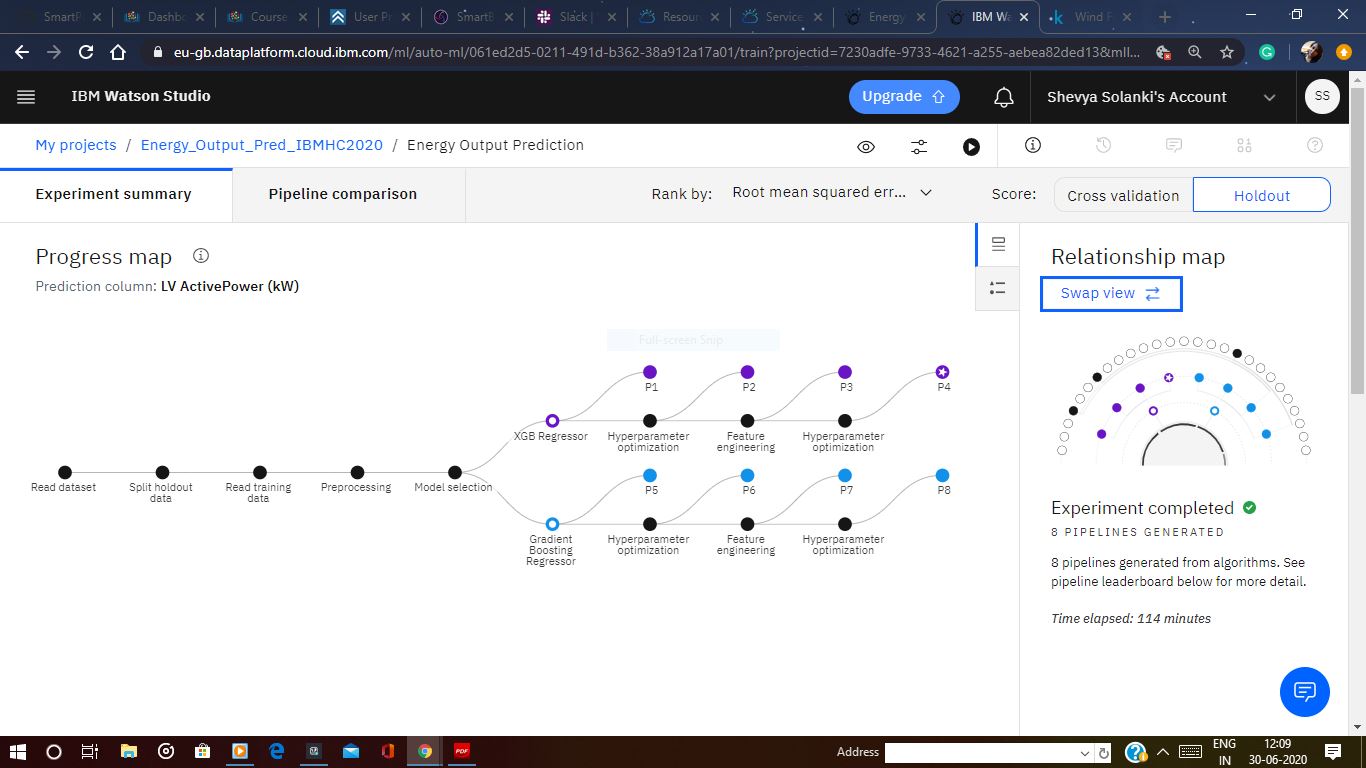
****

Below given is the comparison among different Algorithms used to reach out to conclusion.

****

Below is the predicted values of the following algorithm :

****

****

|  |
| --- |
| **6. Advantage & Disadvantage** |

**6.1 AdvaNtages**

* Wind energy plays an increasing role in the supply of energy world-wide. The energy

output of a wind farm is highly dependent on the weather condition present at the wind farm. If the output can be predicted more accurately, energy suppliers can coordinate the collaborative production of diﬀerent energy sources more eﬃciently to avoid costly overproductions. With this paper, we take a computer science perspective on energy prediction based on weather data and analyze the important parameters as well as their correlation on the energy output. To deal with the interaction of the diﬀerent parameters we use symbolic regression based on the genetic programming tool DataModeler. Our studies are carried out on publicly available weather and energy data for a wind farm in Australia. We reveal the correlation of the diﬀerent variables for the energy output. The model obtained for energy prediction gives a very reliable prediction of the energy output for newly given weather data.

* Wind energy is a key-player in the ﬁeld of renewable energy. The capacity of wind energy production was increased drastically during the last years. In Europe for example, the capacity of wind energy production has doubled since 2005. However, the production of wind energy is hard to predict as it relies on the rather unstable weather conditions present at the wind farm. In particular, the wind speed is crucial for energy production based on wind and the wind speed may vary drastically during diﬀerent periods of time. Energy suppliers are interested in accurate predictions, as they can avoid overproductions by coordinating the collaborative production of traditional power plants and weather dependent energy sources

**6.2 DIADVANtages**

* The main disadvantage is that NO ONE can predict the future. No one knows when someone will die, who will get cancer or not, who will recover and who won't. A person who appears to have been sickly for many years can surprise everyone by outliving all of his peers. Likewise, a person known to be robustly healthy and in shape can succumb immediately to a stroke, heart attack, or other unexpected calamity.
* Insurance companies and others who make their living from using life expectancy as an indicator know they are basically gambling. They rely on statistics that may or may not, ultimately, turn out to be accurate for an individual. For example, they can say "Hypertension tends to be higher in African Americans," which appears to be a statistically accurate statement, in general. However, that doesn't always translate into accurate risk prediction for James, who is an African American male age 54 with high blood pressure, who turns out to live to 105 because he took his blood pressure medication and started practicing meditation and eating healthy to control his blood .Statistics work in generalities. Humans, however, do not.

|  |
| --- |
| **7. Applications** |

* Wind energy plays an increasing role in the supply of energy world-wide. The energy output of a wind farm is highly dependent on the weather condition present at the wind farm. If the output can be predicted more accurately, energy suppliers can coordinate the collaborative production of diﬀerent energy sources more eﬃciently to avoid costly overproductions. With this paper, we take a computer science perspective on energy prediction based on weather data and analyze the important parameters as well as their correlation on the energy output. To deal with the interaction of the diﬀerent parameters we use symbolic regression based on the genetic programming tool Data Modeler. Our studies are carried out on publicly available weather and energy data for a wind farm in Australia. We reveal the correlation of the diﬀerent variables for the energy output. The model obtained for energy prediction gives a very reliable prediction of the energy output for newly given weather data.
* Wind energy is a key-player in the ﬁeld of renewable energy. The capacity of wind energy production was increased drastically during the last years. In Europe for example, the capacity of wind energy production has doubled since 2005. However, the production of wind energy is hard to predict as it relies on the rather unstable weather conditions present at the wind farm. In particular, the wind speed is crucial for energy production based on wind and the wind speed may vary drastically during diﬀerent periods of time. Energy suppliers are interested in accurate predictions, as they can avoid overproductions by coordinating the collaborative production o traditional power plants and weather dependent energy sources.

|  |
| --- |
| **8. Conclusion** |

* In this study we showed that wind energy output can be predicted from publicly available

weather data with accuracy at best 80% R on the training range and at best 85,5% on the

unseen test data. We identiﬁed the smallest space of input variables (windGust2 and dewPoint),

where reported accuracy can be achieved, and provided clear trade-oﬀs of prediction accuracy

for decreasing the input space to the windGust2 variable. We demonstrated that an oﬀ-the-shelf

data modeling and variable selection tool can be used with mostly default settings to run the

symbolic regression experiments as well as variable importance, variable contribution analysis,

ensemble selection and validation.

We are looking forward to discuss the results with domain experts and check the applicability

|  |
| --- |
| **9. Future scope** |

1. Recently, ANNs have been extensively studied and used in the prediction of time series. According to [47], a neural network is a massively parallel distributed processor made up of simple processing units (neurons), which have the natural propensity to store experimental knowledge and make it available for use. It resembles the brain in two respects:

(1) knowledge is acquired by the network from its environment through a learning process; and

(2) connecting forces between neurons, known as synaptic weights, are used to store the acquired knowledge.

According , the arrangement of layered neurons and the pattern of binding between layers are called neural network architecture. The network architecture determines the number of connection weights and how the input signals are processed in the network. A neuron is an information processing unit that is fundamental to the functioning of a neural network. To achieve good performance, neural networks employ a massive interconnection of simple computational cells called “neurons” or “processing units”. The procedure used to carry out the learning process is called a learning algorithm;

|  |
| --- |
| **10. Reference** |

The following sources have been used for the project development :

* To collect the data set : <https://www.kaggle.com/berkerisen/wind-turbine-scada-dataset>.

<https://cognitiveclass.ai/courses/introduction-to-cloud>

* Introduction to Machine Learning with Python by Andreas C. Muller & Sarah Guido.
* Node-Red Labs of the course.
* Stack overflow.
* <https://bookdown.org/caoying4work/watsonstudio-workshop/jn.html#deploy-model-as-web-service>
* W3Schools. Com
* <https://github.com/>
* <https://bookdown.org/caoying4work/watsonstudio-workshop/jn.html#deploy-model-as-web-service>
* Nagababu, G.; Simha, R.R.; Naidu, N.K.; Kachhwaha, S.S.; Savsani, V. Application of OSCAT satellite data for offshore wind power potential assessment of India. Energy Procedia 2016, 90, 89–98. [CrossRef]
* Gao, X.; Yang, H.; Lu, L. Study on offshore wind power potential and wind farm optimization in Hong Kong. Appl. Energy 2014, 130, 519–531.
* Adaramola, M.S.; Agelin-Chaab, M.; Paul, S.S. Assessment of wind power generation along the coast of Ghana. Energy Convers. Manag. 2014, 77, 61–69. [CrossRef]
* Kaplan, Y.A. Overview of wind energy in the world and assessment of current wind energy policies in Turkey. Renew. Sustain. Energy Rev. 2015, 43, 562–568. [CrossRef]
* GWEC. Strong Outlook for Wind Power. 2017. Available online: http://www.gwec.net/global-figures/ graphs/ (accessed on 2 June 2017).
* Manwell, J.F.; McGowan, J.G.; Rogers, A.L. Wind Energy Explained: Theory, Design and Application; John Wiley & Sons: Hoboken, NJ, USA, 2010.
* Dutra, R. Energia Eólica–Princípios e Tecnologia; Centro de Referência para Energia Solar e Eólica Sérgio de Salvo Britto: Rio de Janeiro, Brazil, 2011.
* Fundação Estadual do Meio Ambiente (FEAM). Utilização da Energia Eólica no Estado de Minas Gerais: Aspectos Técnicos E O Meio Ambiente. 2013. Available online: http://www.feam.br/images/stories/ materia/comunicado\_tecnico2\_gemuc.pdf (accessed on 23 February 2017).