Forecasting the Generation of Wind Power in the Western and Southern Regions of India: Comparative Approach

MSc Research Project
Data Analytics

Saptarshi Das
x18127355

School of Computing
National College of Ireland

Supervisor: Dr Muhammad Iqbal

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Saptarshi Das
Student ID:	x18127355
Programme:	Data Analytics
Year:	2019
Module:	MSc Research Project
Supervisor:	Dr Muhammad Iqbal
Submission Due Date:	12/12/2019
Project Title:	Forecasting the Generation of Wind Power in the Western and Southern Regions of India: Comparative Approach
Word Count:	11084
Page Count:	28

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	
Date:	27th January 2020

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	☐
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	☐
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	☐

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Forecasting the Generation of Wind Power in the Western and Southern Regions of India: Comparative Approach

Saptarshi Das
x18127355
MSc Research Project in Data Analytics
National College of Ireland

27th January 2020

Abstract

With the increasing power demand, driven by the rising population and the push towards cleaner energy, India has grown interested in the production of wind-energy due to its renewable nature and high inhouse generation rate. Currently, the states of Tamil Nadu and Maharashtra have been producing the highest amounts of wind-energy. However, to make an informed decision regarding the maximum power output, the generated wind power needs to be predicted beforehand. Intriguingly, wind power forecasting can be generally viewed as a univariate time series analysis problem requiring the usage of complex variable like wind speed, keeping the turbine constant into consideration. Therefore, it can be challenging to model wind speed patterns using standard time series models like ARIMA. In this context, machine learning and neural network-based algorithms have been shown to successfully overcome these limitations. Still, the practicability of simpler yet powerful parsimonious models including Facebook, Inc's Prophet has not been tested in the purview of this specific domain. In this work, five timeseries forecasting models (ARIMA, Dynamic Harmonic Regression, Neural Network, Prophet, Simple Exponential Smoothing) have been implemented for predicting the wind speed (later converted to wind power) in both Tamil Nadu and Maharashtra. The evaluation results indicate that the neural network approach worked best for the given datasets. Nevertheless, the forecasts from the Prophet model were also promising and can be improved as part of future work. Moreover, these research findings can be incorporated into projects involving the site-suitability analysis for developing wind farms across different Indian states.

1 Introduction

Majority of the nations worldwide have grown conscious since the last decade regarding the adverse effects of greenhouse gasses on the global environment [1], and accordingly, the scientists have been exploring various sorts of renewable means to generate an adequate amount of power so as to meet ever-increasing overall requirement. Among the different sustainable power sources, windpower carries extreme emphasis with its high production rate and eco-friendly nature. The accompanying section provides a clear manifestation of the significance of this variety of power and demonstrates the inspiration driving the examination to appraise the generation of it. Furthermore, the questions that this research target to answer is documented in this section comprising the objectives to meet with a clear roadmap on the organization of the remaining literature.

1.1 Research Background and Motivation

With the struggle to minimise the gap between the demand and produced energy, many countries have indulged in importing fossil fuels, combustion which introduces several harmful pollutants to the environment and increased import rate leaves an impact on the nation's economy as well. For a developing country like India, this has become a problem with elevated importance. Due to the fast-growing population, the demand keeps on increasing and at the present scenario, India could produce only a quarter part of the total requirement of energy by renewable means [2]. The geographical location of India provides the country with suitable means to harness various renewable energy sources.. Amongst the renewable ones, wind power has the highest estimated potential in India placing the country on the fifth position among the wind power producers globally [3]. The monthly trend of the year 2018[1] shown in Figure 1, portrays constant fluctuations in the generated wind electricity, also indicating that the southern and the western parts of India to be the dominant producers. Besides having the highest allocated power capacity[2] the states of Tamil Nadu and Maharashtra have the highest power requirement as well from the two regions of south and west of India[3]. Taking this information into account, this research focusses to predict the generated wind power from these two states of India from Nov 2019 to May 2020 in order to meet the power requirement of the country.
The produced power from wind turbines strongly depends on the speed of wind which highly fluctuates every second [4]. So it becomes inevitable to accurately predict the speed of wind in order to robustly estimate the generated power from the wind turbines of any specific location. With the help of machine learning models and analytical techniques, a pattern can be drawn of the fluctuating speed of wind can be set to a deterministic value. The state having a higher forecasted wind speed will be consequentially having the higher capability of producing wind power and hence installing more wind turbines at that location would produce increased amount of power for the country which in turn will help in diminishing the energy deficit.
One of the reliable statistical ways for forecasting from historical data is Autoregressive Integrated Moving Average (ARIMA) model. But in the recent years, many models like Dynamic Harmonic Regression(DHR), Simple Exponential Smoothing(SES) and Neural Netwrok(NN) have been performing very well in different domains for doing time series analysis. So it has become inevitable to use more than one machine learning model to acquire the forecasted knowledge. Facebook recently designed a new forecasting model which has already been tested by several researchers to perform better than the contemporary models [5]. So puting that model to test with the wind speed data might unfold unexpected results.

[1] cea.nic.in/reports/monthly/renewable/2019/renewable-03.pdf
[2] web.archive.org/web/20160304051607/http://www.cea.nic.in/reports/monthly/installedcapacity/2015/installed_capacity-11.pdf
[3] cea.nic.in/reports/annual/lgbr/lgbr-2018.pdf

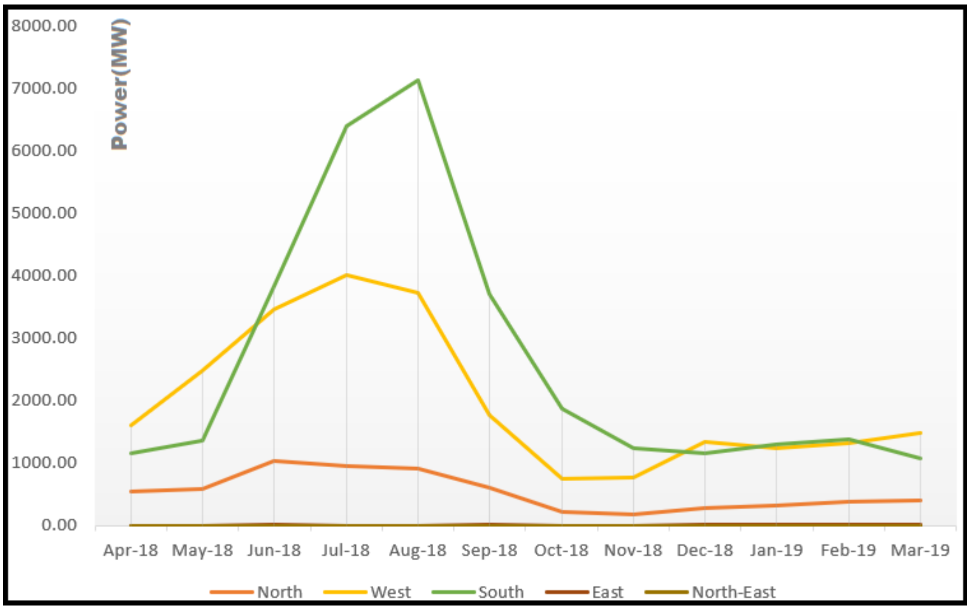

Figure 1: Region-wise Generated Wind Electricity in India

1.2 Research Question

RQ1: *"Which state among Tamil Nadu and Maharashtra can be estimated to potentially dominate the production of wind power in India and will be a stronger candidate to diminish the energy catastrophe of the Country?"*
RQ2: *"How does the business time series handling machine learning model Prophet perform against other standard machine learning models (ARIMA, NN, SES, DHR) for time series forecasting on the meteorological data of Tamil Nadu and Maharashtra?"*
For narrowing the distinction in the midst of generation and necessity of power in India, new and better ways should be acquainted with the framework for more accurately predicting and future-proofing the nation.

1.3 Research Objective

To tackle the previously mentioned research questions, the below-mentioned goals need to be accomplished.
Objective 1: *Forecast the generated wind power from the states of Tamil Nadu and Maharashtra by using the best time series forecasting model on the historical wind speed data (Jan 2016 - Nov 2019) and conclude the state with greater potential of production.*
Objective 2: *To conduct an evaluation process (comparing the errors by means of Root Mean Squared Error and Mean Absolute Error) of the implemented forecasting models and conclude the best performing model for forecasting wind speed for the states of Tamil Nadu and Maharashtra.*

The rest of the article is organized in a manner where the following segment explores the similar sorts of the investigations that have been done to date on the field of wind power generation and time series analysis. The third segment refers to the methodology on which this research is based and every segment of the methodology is briefly described in it. The fourth segment abridges the models which were utilised to predict the speed of wind of the two states and section 5 cites how the executed models performed and discuss how their performance is evaluated. Section 6 concludes the state probable to produce greater wind power in the future.

2 An Inspection of Researches Conducted on Forecasting with Time Series Data, Wind Power & Power Production in India (2009 - 2019)

Over the years, several reports from Central Electricity Authority (CEA) indicated the increasing crisis of power in India in accordance with which the study of S. Rao and S. Ghosh [6] back in 2012 laid down several facts which estimated the demand of electricity in different regions of India, and with help of those facts it could be determined that the western parts will continue to face the highest energy deficit. In support of this study, R. Kale [7] analysed the demands of the state of Maharashtra which pointed the sudden industrial growth in the state since the year 2010 to be the cause for the increased rate of load shedding in the state. Even after seven years of these studies, not much could have been done to fight against this crisis and in the year 2019, still the energy deficit prevails [2] causing the metropolitan cities like Chennai, Mumbai & Bangaluru to suffer from shortfall of energy. This can be put to diminish to a huge extent with a fairly accurate estimation of power in the upcoming years and increasing the generation of power to meet that demand. While there had been much research focussing on estimating the demand, but with the recent reports from CEA[4], it can be confirmed that shortage of supply still remains. So delving into new ways to fight against this situation bears utmost importance. The approval of the Indian government to establish an Indo-Danish Centre of Excellence[5] with a focus on wind energy being signed in March 2019 indicates the urgency of the country to overcome the shortage by renewable means. Around the globe, many researchers have explored various ways to estimate this wind energy to get an idea of the amount of power that this technology can promise to generate [8]. Other than simply forecasting the generated wind power from historical data of a wind turbine, many ones have identified the parameters like wind speed and direction to directly influence the generation [9].

2.1 Power Generation Dependence on Wind Speed

Many multivariate and univariate forecasting kinds of research have been conducted on the production of wind power. Among several parameters, wind speed has been the only uncontrollable parameter which solely can decide the output of a wind farm [10]. Among various approaches to estimate the future of generated wind power, the work of A. Kusiak et al. [4] deserves special mention, where he did not directly forecast the wind power. Instead, he used forecasting models like Support Vector Regression (SVR), Multilayer Perceptron (MLP), Reduced Error Pruning tree (REPT) and bagging tree (BT) on the historical wind speed data acquired from Supervisory Control and Data Acquisition (SCADA) to find short term forecast of wind speed and used K-Nearest Neighbour (kNN) to predict the corresponding wind power. This research [4] made predictions on

[4]`cea.nic.in/reports/annual/lgbr/lgbr-2018.pdf`

[5]`pib.gov.in/PressReleaseIframePage.aspx?PRID=1570597`

two separate ways, one on 10 min interval jumps for 6 times and the other of 4 hourly jumps. As the research opened up so many ways of comparing the results of the forecasting model, they carry much statistical confidence, though making predictions on a greater term might have been more helpful for business decision making and getting more knowledge out of the research. In the experiment SVR could perform the best for making 10 min gapped predictions, but it was outperformed by MLP while forecasting on an hourly manner [4]. He did not stop with introducing so many possibilities in the domain of wind power generation prediction, in the following year, he focussed on using 5 different clustering approaches to predict the future wind speed and hence calculate wind power. In this research he showed that forecasting the wind speed with time series (ts) techniques gave better results than estimating the speed with 10-second prior knowledge from exact previous data. In this research, he determined the importance of the contributing parameter by using Neural Network (NN), Boosting tree algorithm (BTA) and Random Forest (RF) and prepared the five clusters using k-Means algorithm with the parameters. From the results of Mean Absolute Error (MAE), Mean Relative Error (MRE) and standard deviation of both, it could be seen that the NN performed the best for two times [11]. Regarding the clustering techniques in wind power forecasting, a novel T.S.B k-Mean clustering was introduced by R. Azmi et al. [12] later in the year 2016 which ordered information into independent arrangements and helps in better learning for NN by recognizing irregularities and inconsistencies. In the same year 2010 with Z. Zhang, A. Kusiak [13] also published literature with same interval data and almost similar machine learning models but introducing the Double exponential smoothing (DES) algorithm for time series forecasting. Among about 120 parameters present in the dataset, he used boosting tree algorithm to reduce the parameters with low dependency and using Markov chain, he determined the speed of wind after 60secs. Although the forecasting models were almost the same, the approach he took in this research was completely different. Besides comparing the machine learning models to forecast the wind speed, where NN turned out to be the best, he even compared three different approaches to determine the wind power. This experiment of A. Kusiak determined that predicting the generated wind power with the administrable criterion like generator torque and the angle of blade pitch along with wind speed gives the best results which added high importance to this field for further studies [13]. This research even clarified that due to the uncontrollable parameter of wind speed, directly using time series forecasting on historical wind power records cannot produce much accurate results. In all of his researches, it was clearly shown that further the prediction window gets extended, the accuracy of prediction gets decreased as well. Similarly, Markov Chain was also used by Miao He et al. in his analysis for distribution forecast where it outperformed two variations of Autoregressive model [14]. Besides the uncontrollable wind speed, the unpredictable fault of the turbines due to breakdown of machinery can fluctuate the production rate of power as well. Even this was taken into consideration by A. Kusiak where he categorised the different states of a wind turbine into seventeen types and the machine learning models like NN, Support Vector Machine (SVM), RF, Boosting tree General chi-square automatic interaction detector were used to predict the future state of a turbine. For this experiment, he could find that the RF performed the best [10].

The prediction window of A. Kusiak [4] was maxed at 4 hours, which was further increased by K. Bhaskar et al. [15] predicting 30 hours ahead, to benefit business decision making in a better way. As the production of wind power fluctuates highly, so making a very short term prediction might not be very helpful to decide which location will be definite to produce more power in the long term. For this research [15], both historical data of wind speed and power were used to feed the machine learning models Feed Foreward Neural Network (FFNN) and Adaptive Wavelet Neural Network (AWNN) which used back-propagation gradient descent algo for modeling. The research was divided into two phases, where both the models were first used to predict the wind speed and then the forecasted data is utilised to get the wind power by Non-Linear Input-output mapping with FFNN. The forecasted power tested against the acquired wind power data and it could be seen that records obtained by AWNN gave better predictions and the proposed model even surpassed the predictions of Persistence model and New Reference model which were used for benchmarking. Although new two models were introduced in this field of forecasting wind power, the available wind power data could have been used to prove that the direct prediction of wind power cannot be as accurate as calculating it with the forecasted wind speed and hence provide a reason for choosing this path of forecasting [15].

Looking at the prediction of wind power forecasting from a multivariate perspective, researchers like Arthur Bossavy et al. [16] and M. Ozkan [9] considered the wind direction parameter to be another contributing parameter towards estimation of wind power. While A. Bossavy et al. [16] used the same SCADA data as was used by A. Kusiak et al. [4], but the goal of his research was very different. In his study besides forecasting wind power, he focussed on predicting the ramps in the generation which are the sudden drop or increase of generated power due to any kind of anomaly. In this study [16] he aggregated 10 Min interval records to hourly data points which were used on a random forest algorithm to get the forecasted power. D. Barbosa et al.'s [17] case study also pointed out that the accuracy of the forecasting models reaches the highest point with hourly data. But directly concluding to the RF algorithm for making the forecasts, without comparing it against the other contemporary forecasting models leaves off a high possibility of getting better results [16]. This was covered by the other research using multivariate forecasting by M. Ozkan and P. Karagoz [9] who compared their proposed Statistical Hybrid Wind Power forecast technique (SHWIP) against SVM and Artificial Neural Network (ANN) which were used as regression type along with a physical forecasting machine. The main contribution of this research was that with merely a month's data the SHWIP can produce forecasting results with increased accuracy. This was a major breakthrough, as other machine learning models cannot perform well enough with such a small base of training data. For different wind power plant data, the results varied, but mostly with the short span of data, the proposed statistical model performed best among the other three models [9].

At times, the records from the Numerical Weather Prediction (NWP) produces abrupt figures, setting off the whole prediction system for wind power. This anomaly in NWP data used by the majority of the researchers of wind power forecasting might lead them to inaccurate outcomes. Qianyao Xu et al. [18] in the year 2015 identified this issue and conducted research to identify these abnormalities using the Bayes information criterion. The examination even proved that the performance of ANN got improved with the combination of this identification technique [18].

2.2 Analysis of the Wind Farm Power Production

Besides forecasting the generated power from the wind turbines, analysing the data of produced power can even future-proof the financial sectors against loss. Research con-

cerning this sector was done in the year 2013 by M. Schlechtingen et al. [19] where his focus was on power curve monitoring. Though the research field was a bit different, even in his analysis, they figured out that wind speed is the most contributing factor in a process followed by wind direction and ambient temperature. In their research the abnormalities in the production of wind power could be forecasted by the monitoring process, using the machine learning technique. Among the tested models, Adaptive Neuro-fuzzy Interference system (ANFIS) performed the best followed by NN in case of predicting the fault, as well as forecasting the power output [19]. Similar results were also seen by H.M.I. Pousinho et al's [20] research where ANFIS was ensembled with Particle Swarm Optimization (PSO) algorithm and it again outperformed NN along with ARIMA and some more machine learning models incorporating wavelet theory on forecasting generated wind power [20]. Just a year before Meik Schlechtingen et al.'s research on power curve analysis, A. Marvuglia and A. Messineo [21] applied a self-supervised version of NN to monitor the power curve and their proposed model could not forecast the abnormalities in the power production as good as MLP with 10 neurons in many instances of the test . The MLP could not,however, beat the performance of SVR for estimating the ramp rates in the research of H. Zheng & A. Kusiak. In this experiment, SVR could even beat the performance of RF, classification & regression tree (CRT) and pace regression algorithm [22]. S. Kim and I. Seo previously figured out that combining clustering approach with the SVR model hypes the wind power forecasting accuracy by 10% [23]. For predicting the ramps in the generation of wind power, all researchers did forecasting of the wind power first. But as an alternate, H. Zareipour et al. produced a way to directly predict the ramps with Support Vector Machine classification model [24]. This research saved a lot of work by excluding the process of power production to figure out the ramps, but the researcher did the testing with only one model without considering the performance of other classification techniques [24]. Also, the performance of simple exponential smoothing for univariate forecasting on wind power was proved to be better than the forecasting of Auto-Regressive Integrated Moving Average (ARIMA) [25]. Another model named Dynamic Harmonic regression was once tested on wind power data, but its performance was never tested against the other benchmark models. The focus of the research was to make real-time predictions of wind power [26].
Not many researches had been done to handle the chaotic (random) nature of the time series data of wind power. Some of the works with data of this kind must include Xueli An et al's [27] research on 2012 and N. Safari et al's [28] research in 2018. Xueli An compared and proved that the prediction quality improves by handling the chaotic nature of the data, instead of directly feeding the randomised data to the machine [27]. With the same chaotic theory approach, N. Safari's novel decomposition technique could further enhance the prediction accuracy of the wind power [28]. Among all the approaches discussed above, the most important ones which help this research in the best possible way are summarised below in Table 1.

From the above summarization, it can be figured out that univariate way of forecasting has been the most popular way in determining the generated wind power and RMSE and MAE are very common evaluation techniques used by the researchers of this field.

Table 1: Overview Of The Models Used Over The Years.

Model	**Forecasting Type**	**Evaluation**	**Core Idea**	**Ref**
SVR, MLP, REPT, BT	Univariate	MAE, Std AE, MRE, Std RE	2 Fold Prediction of wind speed and then wind power	[4]
SHWIP, SVM, ANN	Multivariate	NRMSE, NMAE, IAE	Forecasting windpower with less than a month's database	[9]
NN, BTA, RF	Univariate	MAE, Std AE, MRE, Std RE	Clustering approach with kMeans used to predict wind speed	[11]
DES, NN, SVM, kNN, RF	Univariate	MAE, MAPE, Std MAE, Std MAPE	Considering blade pitch and generator torque enhances the prediction accuracy of wind power	[13]
AWNN, FFNN	Univariate	MAE, BIAS, RMSE	Forecasted wind speed and predicted wind power with non-linear i/o mapping	[15]
RF	Multivariate	N/A	Forecasted the ramps in windpower production by forecasting wind power	[16]
ANFIS, NN, Cluster Center Fuzzy logic, KNN	Univariate	MAPE, RMSE, MAE	Monitoring the power curve and predicting faults	[19]
SVM	NA	MPCE	Classification way to predict the ramps in wind power generation without even forecasting the wind power	[24]

2.3 Application of Facebook's Prophet Model for Time Series Analysis

A newly designed forecasting model by Facebook which was primarily developed for business ts data are now getting tested in different sectors. The Prophet is already renowned to handle the ts data in the areas where holidays need to be considered and different temporal patterns with random fluctuations are present. The predictions of the Prophet could even outperform the standard ARIMA model in forecasting the Bitcoin prices [29]. The outcome did not change even while making pollution forecasts. The Prophet model integrated with log transformation could perform way better than Box-Jenkins based ARIMA model on making forecasts on several pollutants [5]. In another research to predict the fluctuations in the concentration of ambient fine particles, Prophet was utilized, which turned out to be a success as well [30]. Though Prophet performed well in all the researches, none of them compared its performance with the variety of forecasting models currently present in the contemporary field.

3 Methodology

The research targets to forecast wind power that could be generated from the states of Tamil Nadu and Maharashtra from the historical wind speed data and would help the country to extract a higher amount of power by setting up more wind turbines in the state as to assess which has the potential to produce more power in the future. The whole research can be described in the process flow in Figure 2 which resembles CRISP-DM. This methodology is widely accepted by the business industry with a more thorough architecture [31]. The segments of the methodological flow are further described in the following literature.

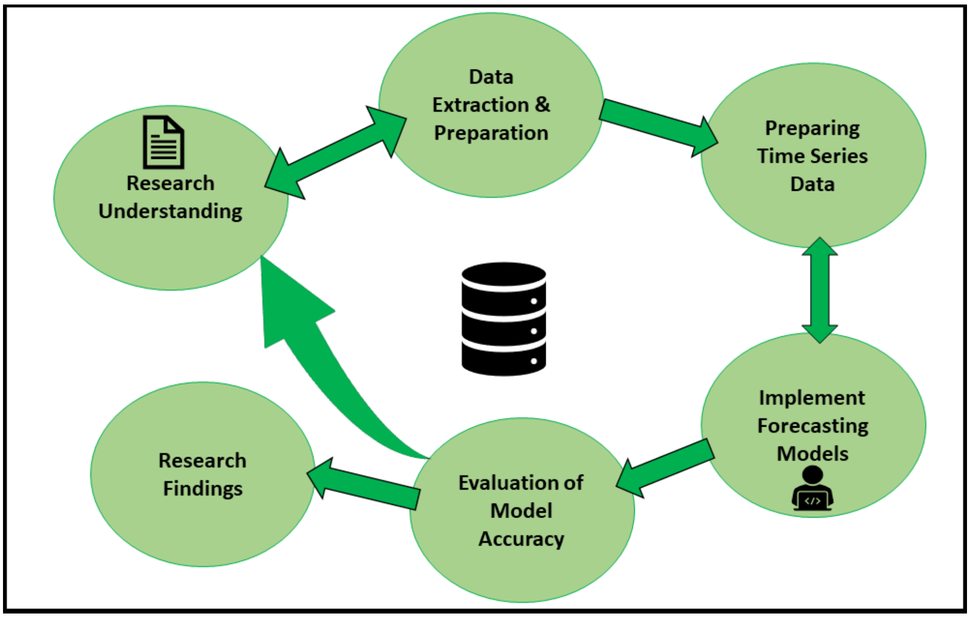

Figure 2: Methodological Structure (CRISP-DM)

3.1 Data Assemblage and Preparation with Research Understanding

From the knowledge gained by analysing several pieces of research on the field of wind power forecasting, it could be concluded that wind speed is the most contributing parameter, its historical records can be enough to predict the potential of a region to generate wind power. So the required wind speed data was extracted along with other meteorological parameters recorded hourly at the Alandur and Karve station of Tamil Nadu and Maharashtra respectively by the Central Pollution Control Board (CPCB)[6]. For consistent availability of records, the time span between 2016 and 2019 has been selected for this research and two individual datasets were downloaded in .xlsx format. The raw datasets contained several columns like wind direction (WD), relative humidity (RH), temperature (Temp) and bar pressure (BP) data, along with information on wind speed (WS). For making univariate forecasting of wind speed, as it has been considered to be the most contributing parameter by several researchers [4] and is also mandatory information for calculating the generated wind power as specified by Asis Sarkar et al. [32], all other parameters of the dataset are dropped for no contribution towards the current research objective. The skimmed dataset was imported to RStudio using xlsx package [33] and some (15 records) negative values of wind speed found in the extracted data were replaced with locally smoothed values, as they were instrumental errors caused by the automatic monitoring machines which was confirmed by the site engineers of CPCB[7]. The description of the univariate datasets which hold a single attribute of wind speed has been shown in Figure 3 and Figure 4 where the dataset of Tamil Nadu had 5546 null records and Maharashtra had 5060 null values out of the initial 33961 records in both datasets. Similar to the negative records, these null records along with the outliers were replaced by the locally smoothed values by the use of tsclean() function. Instead

[6] `https://app.cpcbccr.com/ccr/#/caaqm-dashboard-all/caaqm-landing/data/%7B%22state%22:%22Rajasthan%22,%22city%22:%22Jaipur%22,%22station%22:%22site_134%22%7D`

[7] Email conversation

of replacing the negative or null values with the mean or median of the whole univariate data series, the locally smoothed values preserved the integrity of the dataset in a better way. In the plot of Tamil Nadu in Figure 4, the wind speed remains low during the 3rd quarter of every year whereas, for Maharashtra data, the wind speed of the 2nd quarter remains highest for all the years which is shown in Figure 3. Both the plots can be senn show a vague seasonality. In order to utilize the forecasted results for making business decisions, a mid-term forecasting path was chosen by which half-year (180 days) of future values were forecasted. Making a shorter-term prediction might not be able to help the governing bodies of the nation in making decisions of where to set up more wind turbines in order to generate the most amount of power for the country. In order to make this happen, hourly data were converted to daily records. The datasets were having a "To Data" column which specified the date and exact hour of recording the data. In order to prepare the daily records, the "To Date" column was feature engineered to "Date" column which contained 24 hourly entries of each date and the aggregation is done on the basis of redundant dates. The information of time was irrelevant for the current work, so was not considered for further use. After aggregation, the datasets previously holding 33961 records were left with 1416 days of wind speed data. The measuring unit of wind speed as specified at the source was meter per second. A similar approach was taken by A. Bossavy et al. [16] while he aggregated 10 min intervals of data into hourly entries.

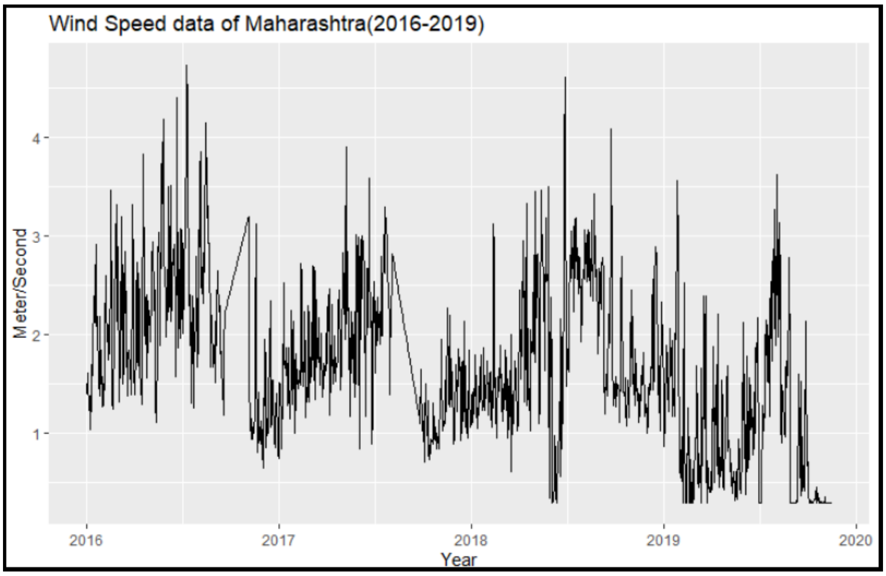

Figure 3: Temporal Wind Speed Data of Maharashtra (2016-2019)

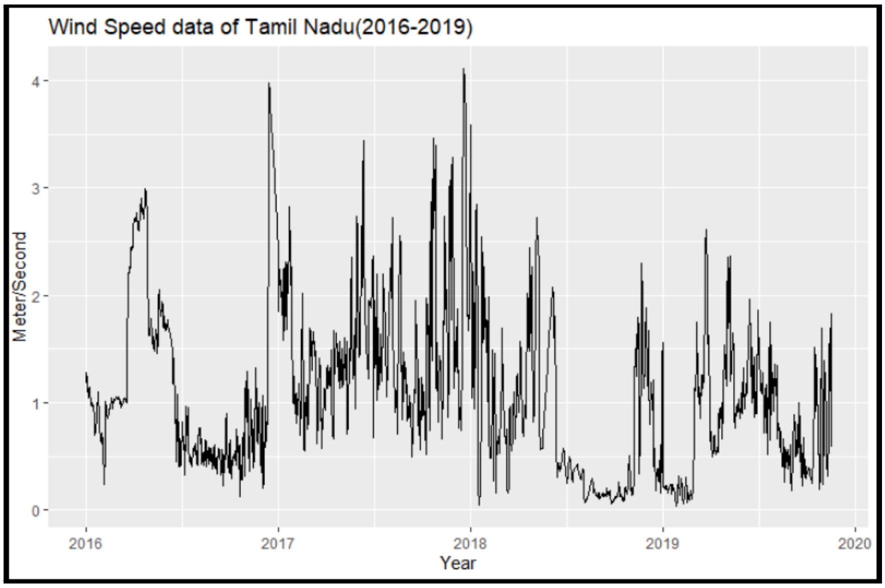

Figure 4: Temporal Wind Speed Data of Tamil Nadu (2016-2019)

The datasets were tested that they were not random white noise and that they carry knowledge along by Ljung-Box test. Also by Dickey-Fuller test, the datasets were found trend stationary. From the Q-Q plot as shown in Figure 5, it can be seen that the data is not very normally distributed, but to be sure, a Shapiro Wilk test[8] is conducted on the datasets and the observed p-value was very less than 0.05, so doing Box-Cox transformation becomes unnecessary. These datasets helped in solving the first research question earlier specified, by which the state having higher wind speed can be determined to generate higher wind power.

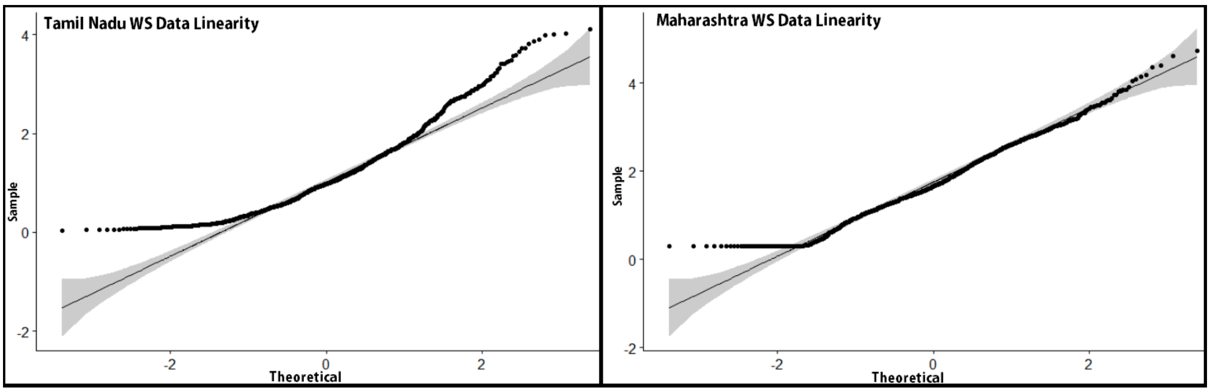

Figure 5: Q-QPlot to Check Data Linearity

3.2 Modelling and Evaluation

Four machine learning model including NN, SES, ARIMA, DHR were implemented against the new Prophet model. ARIMA, SES and DHR are time series analytical models which have been quite popular in handling temporal data. NN has become famous in the last few years for making predictions with high accuracy, While Prophet model has been recently developed in 2017 by Facebook's data science team. This was initially developed to handle temporal business records with the ability to handle irregular data

[8] www.statisticshowto.datasciencecentral.com/shapiro-wilk-test/

with the holiday effect taken into consideration. A detailed description of all the models are discussed in the following section 4.
All applied models are then evaluated through two error assessment indices of Root Mean Squared Error (RMSE)[9] and Mean Absolute Error (MAE)[10]. A description of each of these strategies along with the results of how the models performed on these benchmarks is detailed in section 5 which target to achieve the Objective 2 specified in this research.

3.3 Research Design

To depict a better comprehension of the current research to the fellow scientists, a neat structure of the arranged project has been presented in Figure 6. As the raw data which will be procured from the automatic monitoring machines of CPCB, it will be preprocessed and organized in accordance with the machine learning models' necessity. Additionally, to confine the customers from the data processing layer to keep away from any kind of bias to the result, a 3 level structural plan has been embraced.

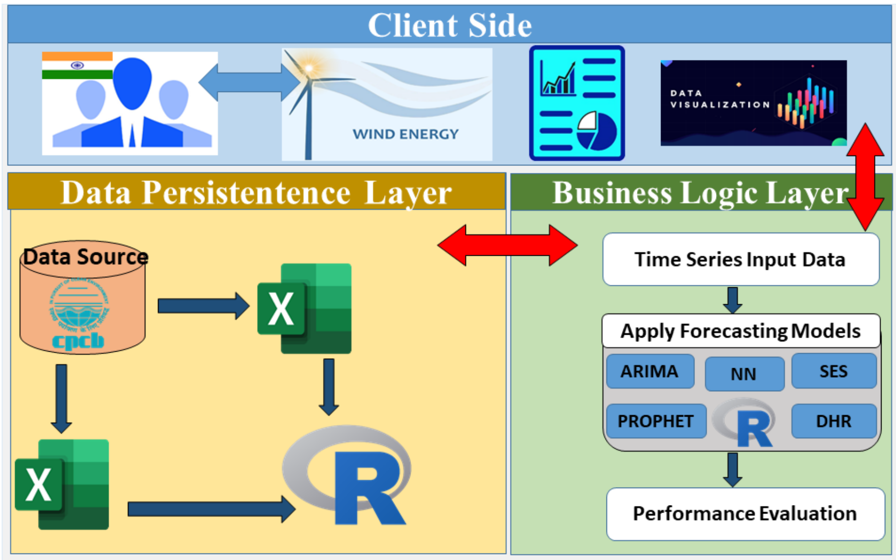

Figure 6: Project Design

The above design specifies how the client who in this research is the governing body of India will be only accessing the visualized end results and make the ultimate decision without intervening into the complexities of data. The data preprocessing steps discussed in the earlier section will be limited to the data persistence layer and all the modelling work along with evaluating them to produce the results to visualize to the clients will be done in Business logic section.

3.4 Tools and Specification

Along with the methodology applied to achieve the goal of this research, the technologies that helped achieve it are breifly describeded below.

[9]www.statisticshowto.datasciencecentral.com/rmse/
[10]medium.com/@ewuramaminka/mean-absolute-error-mae-sample-calculation-6eed6743838a

• Microsoft Excel 2019 was used for part of preprocessing the data which saved the coding time and also preliminary analysis which is one of the basic benefits of using Excel [34]. The irrelevant column and row drop along with initial checking of the dataset of whether it holds all records or not was done by this tool.
• RStudio Integrated Development Environment (IDE) was used for coding in R language for further manipulation of the datasets and applying them to the machine learning models. R has many forecasting packages for easy time series data handling [35], so it has been chosen to increase the efficiency of the pace of the research. The models were coded in R language as well as visualizing the datasets on the preliminary level. Some of the visualizations in R studio helped to get an idea of the components of time series data like seasonality and trend.
• Tableau 2018.2 was used for visualizing the end results for better presentable graphics. Tableau provides a user-friendly interface to visualize a wide spread of data for better analysis [36], so incorporating this software to this research makes the final results more comprehensive.
• All the above mentioned softwares have been running on Windows 10 machine.
A detailed description of the tools and their usage can be found in the configuration manual provided.

4 Model Implementation

After the pre-processed data were converted to time series format using zoo and tseries packages of RStudio as guided by Robert Lund [37], 75% of it was split using splitTrainTest() function from CombMSC package to train the models which is a good practice according to A. Metcalfe et al. [38]. Only the Prophet model has a requirement of having a temporal column named as "ds" and a data column named "y"[11]. So a copy of the dataset was prepared separately in that fashion. All the models were implemented with the help of R base packages, along with some extra ones which are specified in the Configuration Manual provided with the report. The execution of this section accomplishes the first objective of this research. A schematic flow of the implemented data throughout the research has been described in Figure 7. The models that the wind speed dataset got implemented to, are discussed in details in the upcoming subsections.

[11] `https://facebook.github.io/prophet/`

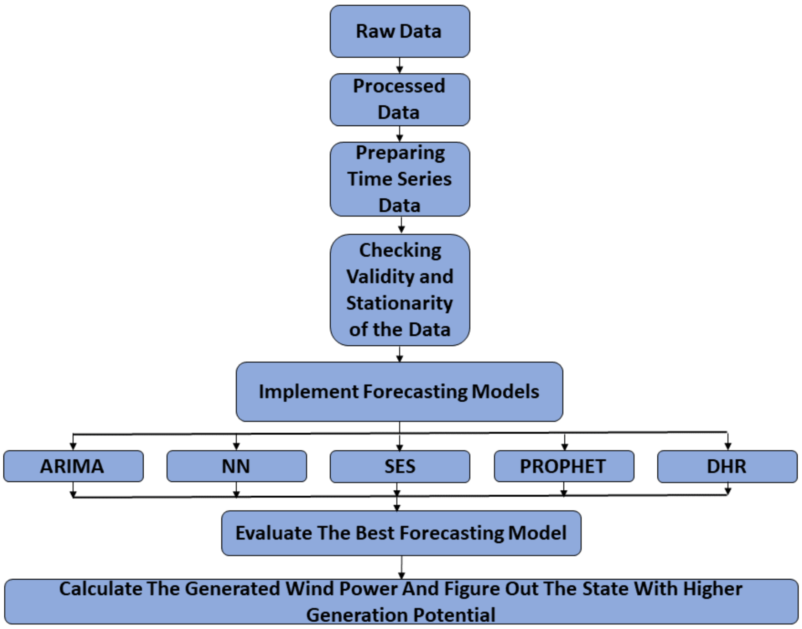

Figure 7: Data Flow Diagram

4.1 ARIMA(p,d,q) Model

The auto.arima() function from forecast package [39] chose the ARIMA(3,0,2) model for Tamil Nadu wind speed dataset and ARIMA(2,0,2) model for Maharashtra wind speed dataset. This auto.arima() function searches through all the possible models for a dataset and assigns the best suited model with lowest AICc value. AICc is basically the parameter measured to compare two models. The model with a lower value of this parameter is always preferred. But the AICc value of two different forecasting methods cannot be compared, as their measure AICc values are on completely different scales. So the variation of the forecasted values of a model from the actual data has been widely used make a comparison across different forecasting techniques. Both the datasets were tested to be stationary by Dickey-Fuller test, so the search has been bounded to stationary models for a faster run. But as the applied dataset size was not too huge, to search through all the available models would not have made much of a difference to the computational time, so the approximation parameter was turned down.

```
############################## ARIMA #########################################

MHarima <-auto.arima(Maharashtra$train,D=0,lambda=NULL, stationary = T,
                    stepwise=F, trace=F, approximation=F, biasadj=F)
checkresiduals(MHarima)
```

Figure 8: R Code of ARIMA with used Hyperparameters

A snippet of the code is shown in Figure 8 where it can be seen that the model has not incorporated Box-Cox lambda transformation, as the data was tested to be normalized. The model chosen can be depicted in the below form of notation (1)-

$$\textbf{ARIMA(p,d,q)} \tag{1}$$

Though the datasets were stationary, but ARIMA model has been quite popular for its ability to perform well with non-stationarity as well. In the above representation, p stands for the order of autoregressive terms, q depicts moving agerage part of the model and d denotes the number of times the data needed to be differenced non-seasonally. As from the Shapiro-Wilk test, it was found that the data were normally distributed, so both the datasets needed no differencing and that is why both the selected models by auto.arima() function had 0 as the value for d of notation (1)

4.2 Neural Network Model

The nnetar() function from the same forecast package [39] has been used to develop the neural network for the time series data. As this function feeds the neural network with lagged data of the input time series, similar to linear regression, so the neural network it creates can also be called NNAR. The model can be best represented in the form of the notation (2) shown below

$$\textbf{NNAR(p,P,k)} \tag{2}$$

Here p specifies the number of lagged observations in the input layer, P tells about the seasonal lagged inputs, k holds the information about the number of nodes in the hidden layer, m depicts the frequency of the data. By hyperparameter tuning, the best model selected for the datasets is NNAR(41,2,19)[365] producing forecasts nearest to the actual records. This portrays a scenario somewhat relatable to Figure 9 which has 1 input layer and 1 hidden layer where the input layer holds (41+2) 43 node viz. Y(t-1), Y(t-2), Y(t-3)... Y(t-41), Y(t-365), Y(t-730). The split train data holds just 2 seasons, so maximum 2 seasonal lags could be assigned to the model. The hidden layer has been assigned with 19 nodes which produced the highest accuracy of prediction. The nnetar() function is so designed that if the number of hidden layer nodes was not assigned, then by default it would have got 21 (by (41+2)/2) hidden nodes. The known daily frequency is also shown in the notational representation.

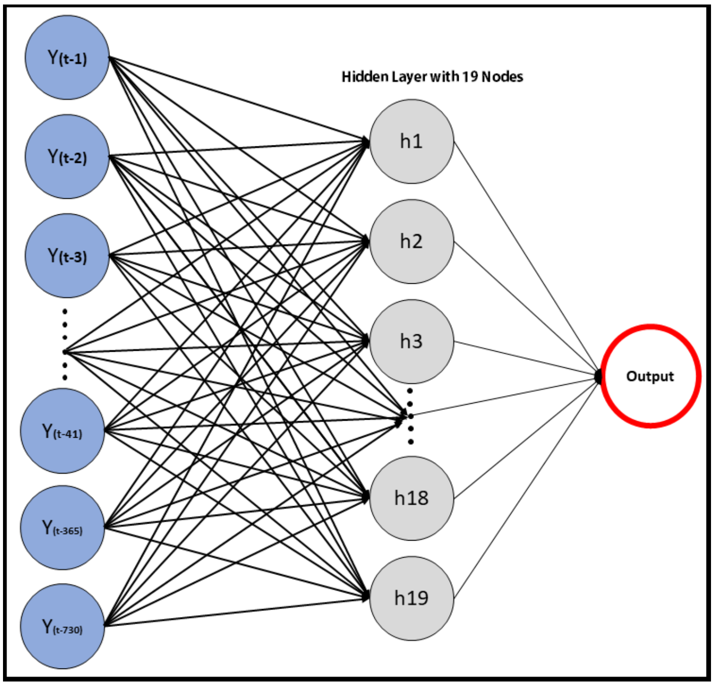

Figure 9: NNFOR(41,2,19)[365] Architecture

A snippet of the code in R language of the use of nnetar() function has been shown in the below Figure 10. Neural networks randomly select inputs and every time accuracy may differ, so a mean of 50 repeated runs will be considered for this research. The repeat parameter in the function has been specified accordingly and as specified before, the data are normalized, so no lambda transformation or scaling of the input data is required.

```
##################################### Neural Net ############################################
set.seed(18127355)
TNnn=nnetar(Tamil$train,p=41,P=2,size=19, repeats=50, lambda = NULL, scale.inputs=F)
```

Figure 10: R Code for NNETAR

4.3 SES Model

Unlike ARIMA model which has the ability to handle non-stationary data, Simple Exponential Smoothing performs the best with non-seasonal and stationary univariate data [39]. The ses() function from forecast package has been used to implement the model. For this function, the forecasting needed to be set fitting the model itself. A snippet of the code has been shown in Figure 11 where the initial parameter is set to optimal which specifies that the initial inputs to the model will be optimized by ets model. As the data is normalized, so lambda transformation and alpha smoothing parameter have not been utilized. As the mean forecasts are not necessarily of benefit, so biasadj has been kept false.

```
########################## Simple Exponential Smoothening ################################

TNses <- ses(Tamil$train, h = 354, lambda = NULL, initial="optimal", biasadj=F)
```

Figure 11: R Code for Simple Exponential Smoothing

4.4 Facebook's Prophet Model

The application of Facebook's Prophet model to the wind speed data to predict the wind power has not been done till date. So exploring and testing the performance of the model against other models which has been already in use in this domain comprises the novelty of this research. The quick and responsive model designed by the Data Science team of Facebook in 2017 is mostly praised due to its ability to handle irregular data. The parameter to handle the holiday gap is mostly utilized to forecast business time series. But for this wind speed data, no particular gap was found, so the holiday parameter remained unused. Another benefit of this model is its highly accurate predictions in very less time, due to the simple parameters of the model. The main 3 components of this model are trend represented by g, seasonality represented by s and holidays represented by h which makes it a decomposable model [40]. The below equation (3) clarifies represents the working strategy of the model.

$$y(t) = g(t) + s(t) + h(t) + \epsilon_t \tag{3}$$

Other than the three components previously mentioned, ε relates to the error term which unlike ARIMA or other models used in this project is not a white noise. Rather it is the error due to any unusual changes accomodated by the model during fitting. The error is a normally distributed like error of ARIMA model. The model specifically needs a temporal series of data in the form of a data-frame with univariate data column named to be "y" and datestamp column named as "ds" [41]. The model uses the prophet() function from the opensource prophet package of Facebook.

```
TN.prophet<-prophet(TN.temporal[1:1062,], changepoints =c("2016-12-12"),
                    seasonality.mode = 'additive', daily.seasonality=F, fit = T)
TNfuture <- make_future_dataframe(TN.prophet, periods = 354)
TN.prophetforecast <- predict(TN.prophet, TNfuture, type="response")
```

Figure 12: R Code for Prophet Forecasting Model

From the above snippet in Figure 12, it can be observed that the changepoints parameter has been nullified as from the wind speed data of Maharashtra observed during the preliminary data analysis in Excel, there were no changing points observed. But for Tamil Nadu data series, one change point was observed at "2016/12/12" which was specified while modelling for that data. As the data did not show any daily seasonality, so it has been turned off and to fit the model instead of just initializing and testing, fit parameter has been marked true. For the preliminary data analysis, additive seasonality was observed form the datasets, so it has been specified as well instead of multiplicative. For making predictions of the future by this model, a blank frame of only dates need to be created, which is done and can be seen in the snippet.

4.5 Dynamic Harmonic Regression Model

While the ARIMA model performs better with shorter seasonal periods, DHR has the ability to produce better outcomes irrespective of the seasonal period. The Fourier term[12] of DHR has the ability to handle the seasonal pattern of any time series and it can be dynamically adjusted for smoother forecasts. This model takes error from ARIMA process as an input as well. The below equation [42] demonstrates the working principle of this model.

$$y_t = T_t + S_t + e_t \tag{4}$$

Here e is the ARIMA error, y represents the ts data and T is the trend component which in this wind speed data is not required and S stands as the periodic component. The below snippet in Figure 13 shows the code for DHR where the auto.arima() function from forecast package has been used.

```
###################Dynamic Harmonic Regression##################

TNdhr<- auto.arima(Tamil$train, xreg= fourier(Tamil$train, K=3), lambda = 0,
                   stepwise=F, trace=F, approximation=F, biasadj=F)
```

Figure 13: R Code for Dynamic Harmonic Regression Model

The value of K in the above function controls the degree of seasonal smoothing and it would be decided solely on the AICc value. The minimum AICc value was observed at K=3. As the data size is not too huge, so searching through all the models to get the best one, will not take much time, so stepwise and approximation parameter is turned negative.

[12] content.pivotal.io/blog/forecasting-time-series-data-with-multiple-seasonal-periods

5 Result Evaluation

In this section, how the models performed will be discussed in terms of errors and also the state producing the most wind power will be evaluated. By the end of this segment, the 2nd objective of this research would be achieved. RMSE and MAE are the error measurement parameters by which the predictions of the applied models are assessed. To be sure about investing more on the wind farms of a state, the predicted wind power needs to be as close as possible to the reality, so the evaluated best model will be chosen from these error test results.

5.1 Model Evaluation

5.1.1 ARIMA Forecast

The selected ARIMA(3,0,2) and ARIMA(2,0,2) models were used to forecast the 354 days ahead which was the testing part of the data left out previously. Though the auto.arima function automatically selects the best model of ARIMA for the dataset, but to ensure that, the residuals were checked of both the models and they showed nothing but white noise. If a model is perfect fit, then the residuals must be left with no usable knowledge in the data, which is as good as white noise [43]. The forests of the testing counterparts by these models are shown in Figure 14.

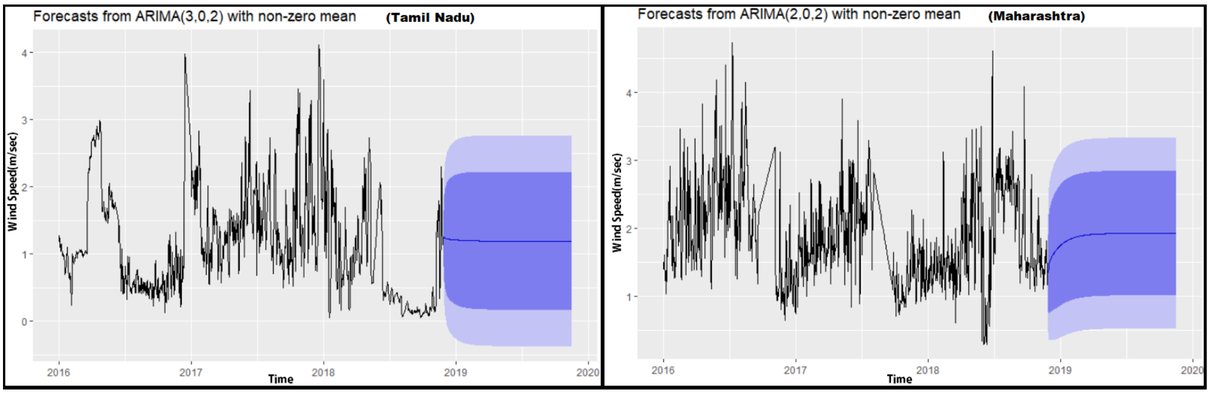

Figure 14: ARIMA Model Forecast Plot

The forecasted results are plotted by autoplot function which can only plot time series data and in the plot, the blue line shows the prediction and the light blue area resembles the prediction interval.

5.1.2 Neural Network Forecast

Due to the higher execution time of the neural network, only the point estimates are forecasted and plotted. The prediction interval was avoided for this model. The autoplot output of the forecasts can be seen in Figure 15 where the blue line represents the predictions of the model for 354 days.

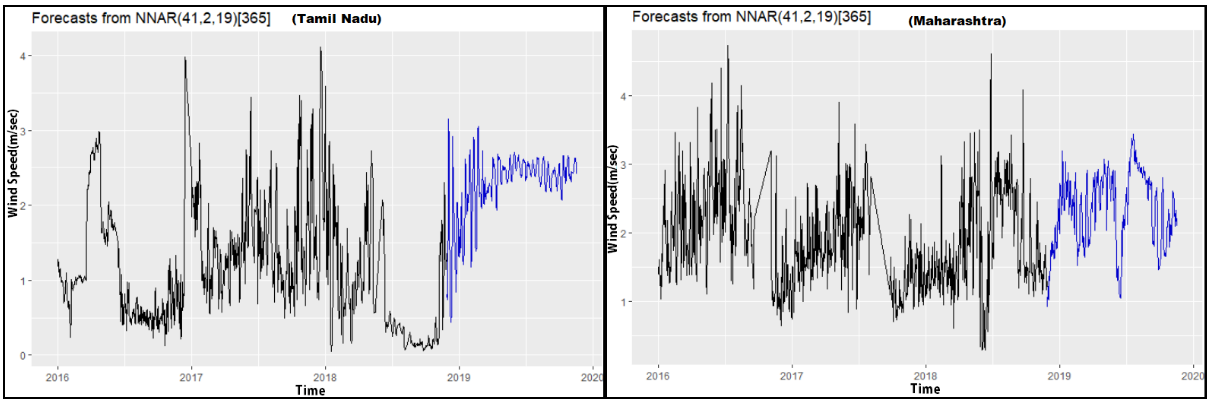

Figure 15: Neural Network Forecast Plot

The forecast of Tamil Nadu wind speed can be seen to considerably increase in the forecasted period as compared to the Maharashtra wind speed.

5.1.3 Simple Exponential Smoothing Forecast

For SES, the fitted model is plotted on top of the original data to check how thoroughly the prediction reciprocates the original series. Figure 16 shows that the fitted model forecasts following the original series quite closely. The red line represents the fitted model output and the black line shows the original data. The fitted line of SES can be seen to closely follow the original data points. This is be because SES uses weighted sum of the recent past values.

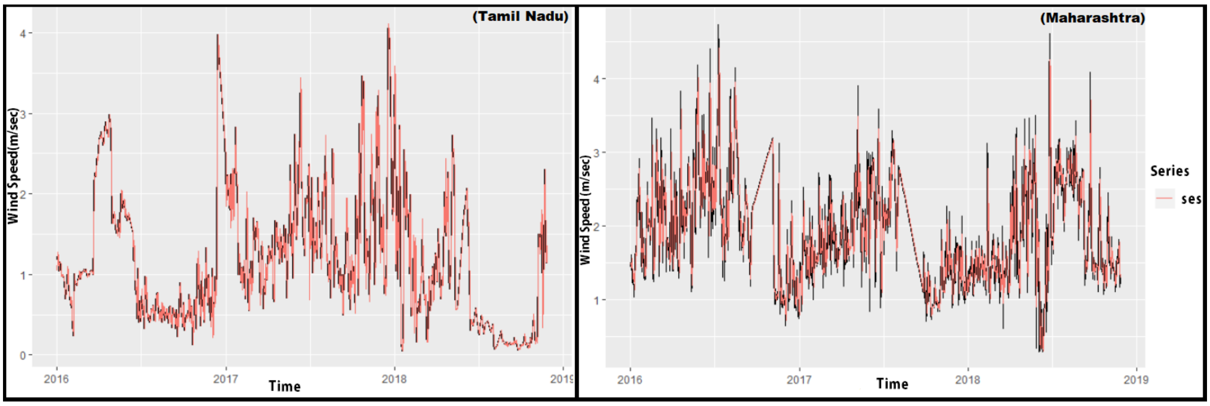

Figure 16: Simple Exponential Smoothing Forecast Plot

A detailed visualization is attached in the attached configuration manual where the closeness of the fitted line with the original data plot can be observed more clearly.

5.1.4 Prophet Forecast

The dyplot.prophet() function was used to picture the forecasts of the fitted Prophet model. The output produced by this function was an interactive visualization which is shown in Figure 17.

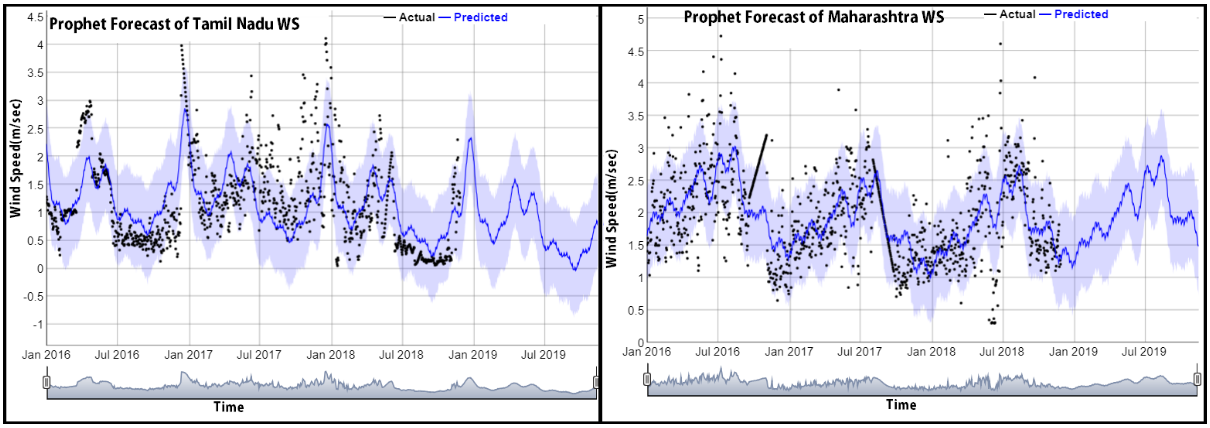

Figure 17: Prophet Forecast Plot

The black dots seen in the plot are the actual data points and the blue line is the fitted model point estimates along with the lighter blue region showing the predicting interval. The blue line can be seen to follow the actual datapoints quite closely and due to the interactive nature of the visual piece, browsing the cursor along the time stamp could neatly show the actual target value and predicted value at that particular date.

5.1.5 Dynamic Harmonic Regression Forecast

Much like the plotted visualization of ARIMA, the forecasts of the DHR model show a similar curve, but a bit smoother, due to the Fourier term. Figure 18 shows the forecasted output of the fitted model.

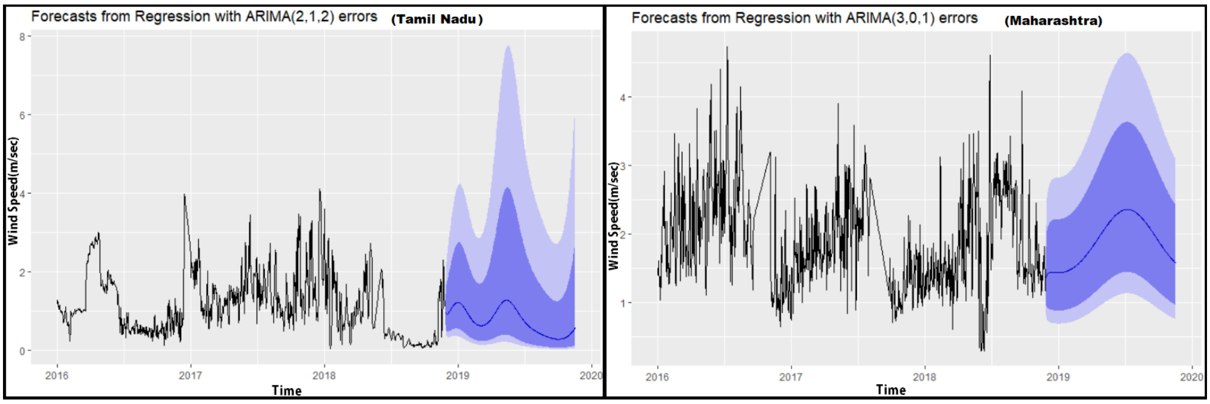

Figure 18: Dynamic Harmonic Regression Forecast Plot

The plotted forecasted values of DHR to some extent contradict the results of Neural Network, as DHR forecasts more higher wind speed values for Maharashtra.

5.2 Results

Measuring the errors of the five implemented models on two different scales helped in concluding the best forecasts of wind speed for the two states. Both RMSE and MAE and basically different metrics to check how much the residuals of a model are normally spread over the fitted line. Though there had been many differences in opinion about choosing one technique over another, but the study of T. Chai and R. R. Draxler [44] clarified that both techniques have benefits depending on the chosen models to evaluate and depending on a single projection of a technique to evaluate machine learning models is not accurate. So the results from both the metrics are considered for the current study. Below in Table 2, the results of both the techniques have been plotted.

Table 2: Evaluation Matrix

	Tamil Nadu		Maharashtra	
Models	RMSE(m/sec)	MAE(m/sec)	RMSE(m/sec)	MAE(m/sec)
ARIMA	0.32179887	0.2052467	0.43926421	0.31914015
NN	0.02025666	0.0128214	0.05275985	0.03290602
SES	0.33379503	0.2031907	0.46603376	0.32747497
Prophet	0.726561	0.5901662	1.17976898	1.03512878
DHR	0.32980306	0.2105339	0.43779327	0.31764086

Though both the scales placed the models at different positions for Tamil Nadu, but NN consistently performed the best and the Prophet model could not perform well enough as compared to all other models in both scales. NN could handle the wind speed data of Maharashtra in the best possible way as well and from the performance scales, it could clearly be seen that with the Fourier term DHR performed slightly better than ARIMA followed by SES. DHR could have performed better if a longer seasonal period prevailed in the applied data. Similar outcome was observed about the performance of Facebook's Prophet model with the Maharashtra data as well. This result answers the second research question and determines the position of the Prophet model among the other forecasting models which have been performing generously well over the years. So, from both states results and across the two error measuring scales, NN can be chosen to get the best forecasts of wind speed for the next 180 days. Though the NN could perform the best as like the observation of A, Kusiak [13] in his research, but also by hyperparameter tuning, the NN in the current research could achieve greater height and score lower error rate on MAE scale with wind speed data. With the forecasted wind speed, the calculation of wind power will be described in the following sub-section.

5.3 Calculation of Generated Wind Power

The book of T. Burton et al. [45] guided with the below equation (5) to calculate the generated power from a wind turbine.

$$P_W = 1/2(C_p \rho A s^3) \tag{5}$$

Here the parameters can be depicted in the below form-
P_W- Wind Power(Watt)
C_p- Power Coefficient(unitless)
ρ- Air Density(kg/m^3)
A- Rotor Swept Area(m^2)
s- Wind Speed(m/sec)

The unit derivation of the above equation has been described in the appendix section. In order to calculate the potential of generating wind power from each state, other than forecasted wind speed, the values for the rest of the parameters of the equation are required. While researching on this domain of wind turbine power generation, it was observed that the highest reported power coefficient was 0.45 [46], so the current research is going to consider the same value. Over the years the diameter of the rotor has been steadily increasing for getting maximum power output and by a survey done by wind-monitor.de[13], the diameter of the rotor mostly used in the current year is 129 meters. The known density of air at the sea level[14] is 1.225 kg/m^3. Accumulating these constant with fluctuating wind speed, the generatable power of wind from a single wind turbine installed in any of these two states of India was calculated and portrayed in the line chart shown in Figure 19.

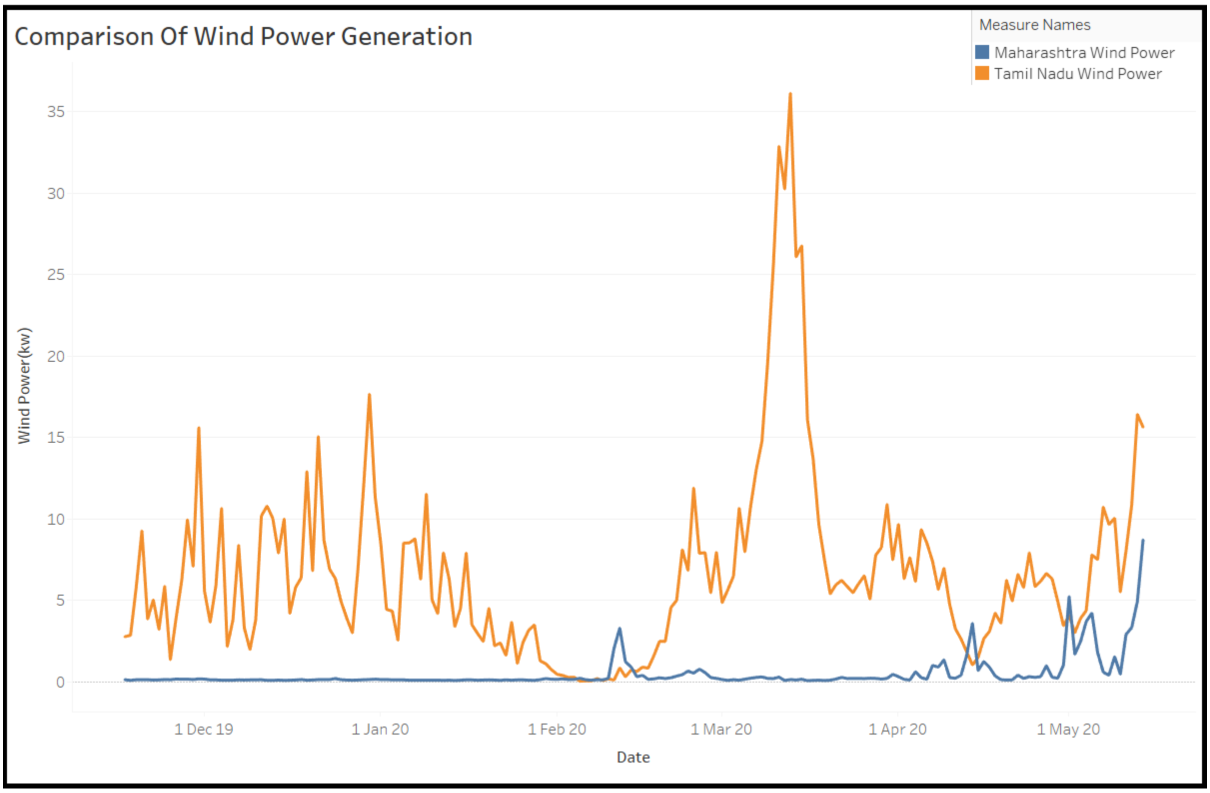

Figure 19: Produced Power of Wind in Tamil Nadu & Maharashtra

From the chart, it can be clearly seen that with the predicted wind speed of the Neural Network, a considerably higher power of wind is getting generated from the state of Tamil Nadu. A single turbine can generate as high as 36kw of power in a day and a cumulated power of about 1200kw in the 6 months (180 days). As the generated wind power line of Tamil Nadu of the chart can be seen consistently higher than the power generated in Maharashtra, so the knowledge of investing more for installing wind turbines in the state of Tamil Nadu will be more beneficial in combating against the deficit of power in the country. And finally, the first research question specified in this research work gets solved by interpreting this line chart in Figure 19.

5.4 Discussion

At the source of the data, the records were getting inconsistent further going into history. So the data of the research was considered for the limited 3.5 years. With the availability

[13] https://www.statista.com/statistics/263901/changes-in-the-size-of-wind-turbines/
[14] https://www.thoughtco.com/density-of-air-at-stp-607546

of more data, the machine learning models could be better fitted and also the range of confident prediction could have been increased. In this research, a single layer NN could outperform all other conventional models with a broad margin. It is thrilling to imagine with these results, that how close to reality predictions could have been made of such an unpredictable wind speed parameter with more complex neural networks. The Multilayer perceptron (MLP) neural network was even tried to implement in this research with the available range dataset, but instead of producing better results, its MAE scores came close to a single layer NN only. As the error results were measured across the same measuring scales which have been the deciding factor for the researches previously done in this same field, so the results carry high confidence in this field of wind power forecasting. Due to the unavailability of wind power data, a novel multivariate forecasting could not be conducted which could further enhance the value of this research in the domain of forecasting wind power generation.
The Prophet model could outperform ARIMA model in the research of Isil Yenidogan et al [29], but could not beat the ARIMA model in handling wind speed data, which might be because the selection of changepoints for the model needed to be more accurate. But as the model is very newly designed, not much work could be found on it and so the concept of changepoint is a bit ambiguous.
Due to the limitation of data availability, the research has been done on the wind speed data of these two states. But even a state is a huge landmass on the country. If, by any means, the wind speed data of every city of these two states were available, then by data comparison, the wind speed of the city with consistently high speed of wind would have been considered for the research. This would help in pinpointing the location more accurately for setting up wind turbines.

6 Future Scope & Research Conclusion

The prior aim of this research was to determine the best state for the country of India for setting up more wind turbines in order to produce the maximum wind power. With clear results, Tamil Nadu could be seen in the future to produce the maximum amount of wind power. In the process of determining the suitable state, the next important objective of testing Facebook's Prophet model was accomplished and could be seen that it could not perform as good as ARIMA, DHR, SES or Neural Network. Neural Network performing the best, was utilised to predict wind speed for the two states. Successfully both the research questions previously mentioned of figuring out the best machine learning model and the state with higher wind power generation potential were done by successfully achieving the prior stated objectives. Due to the success of this research, the acquired knowledge of generatable wind power in the next six months, if provided to the country, can benefit them in producing more inhouse power from wind turbines and that would help them reduce the prevailing energy gap. The Prophet model though could not outperform the Neural Network, but its application shed some light on the unknown corners of the current domain arising the possibility of enhancement. With appropriate data, the benefit of the holiday parameter of this model could have been utilized and with such inconsistent flow, this model might produce better results than other models which are sensitive to such data.
In future, this research can be further extended by widening the domain of renewables. Though currently, the government of India is most interested in the growth of wind power

production in the country, but the scenario was not the same a couple of years back. So forecasting and comparing the generation power from biomass, solar radiation and hydro plants with a broader base of data to predict beyond a couple of years would increase the confidence of the country on their reserve of power and slowly they could reduce the import non-renewable resources to the country.

Acknowledgment

This research is fully bolstered by the National College of Ireland and I would specifically convey my gratitude towards Dr Muhammad Iqbal for his thorough support & guidance for helping me in completing the research on time and for his significant proposals on expressing my work on paper with clarity. I am also highly grateful to my parents, Mr & Mrs Das for their constant support at every step.

References

[1] P. B. Duffy, C. B. Field, N. S. Diffenbaugh, S. C. Doney, Z. Dutton, S. Goodman, L. Heinzerling, S. Hsiang, D. B. Lobell, L. J. Mickley, S. Myers, S. M. Natali, C. Parmesan, S. Tierney, and A. P. Williams, "Strengthened scientific support for the Endangerment Finding for atmospheric greenhouse gases," *Science*, vol. 363, no. 6427, p. eaat5982, feb 2019. [Online]. Available: http://www.sciencemag.org/lookup/doi/10.1126/science.aat5982

[2] R. Varma and Sushil, "Bridging the electricity demand and supply gap using dynamic modeling in the Indian context," *Energy Policy*, vol. 132, pp. 515–535, sep 2019. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0301421519303842

[3] A. Kumar, K. Kumar, N. Kaushik, S. Sharma, and S. Mishra, "Renewable energy in India: Current status and future potentials," *Renewable and Sustainable Energy Reviews*, vol. 14, no. 8, pp. 2434–2442, oct 2010. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S1364032110001140

[4] A. Kusiak, Haiyang Zheng, and Zhe Song, "Short-Term Prediction of Wind Farm Power: A Data Mining Approach," *IEEE Transactions on Energy Conversion*, vol. 24, no. 1, pp. 125–136, mar 2009. [Online]. Available: http://ieeexplore.ieee.org/document/4749292/

[5] K. K. R. Samal, K. S. Babu, S. K. Das, and A. Acharaya, "Time Series based Air Pollution Forecasting using SARIMA and Prophet Model," in *Proceedings of the 2019 International Conference on Information Technology and Computer Communications - ITCC 2019*. New York, New York, USA: ACM Press, 2019, pp. 80–85. [Online]. Available: http://dl.acm.org/citation.cfm?doid=3355402.3355417

[6] S. R. Rallapalli and S. Ghosh, "Forecasting monthly peak demand of electricity in India—A critique," *Energy Policy*, vol. 45, pp. 516–520, jun 2012. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0301421512001814

[7] R. V. Kale and S. D. Pohekar, "Electricity demand supply analysis: Current status and future prospects for Maharashtra, India," *Renewable and Sustainable Energy Reviews*, vol. 16, no. 6, pp. 3960–3966, 2012. [Online]. Available: http://dx.doi.org/10.1016/j.rser.2012.03.008

[8] P. Enevoldsen and G. Xydis, "Examining the trends of 35 years growth of key wind turbine components," *Energy for Sustainable Development*, vol. 50, pp. 18–26, jun 2019. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S097308261831353X

[9] M. Ozkan and P. Karagoz, "A Novel Wind Power Forecast Model: Statistical Hybrid Wind Power Forecast Technique (SHWIP)," *IEEE Transactions on Industrial Informatics*, vol. 11, no. 2, pp. 1–1, 2015. [Online]. Available: http://ieeexplore.ieee.org/document/7018961/

[10] A. Kusiak and A. Verma, "A Data-Mining Approach to Monitoring Wind Turbines," *IEEE Transactions on Sustainable Energy*, vol. 3, no. 1, pp. 150–157, jan 2012. [Online]. Available: http://ieeexplore.ieee.org/document/5963730/

[11] A. Kusiak and W. Li, "Short-term prediction of wind power with a clustering approach," *Renewable Energy*, vol. 35, no. 10, pp. 2362–2369, oct 2010. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0960148110001448

[12] R. Azimi, M. Ghofrani, and M. Ghayekhloo, "A hybrid wind power forecasting model based on data mining and wavelets analysis," *Energy Conversion and Management*, vol. 127, pp. 208–225, nov 2016. [Online]. Available: http://dx.doi.org/10.1016/j.enconman.2016.09.002https://linkinghub.elsevier.com/retrieve/pii/S0196890416307804

[13] A. Kusiak and Z. Zhang, "Short-Horizon Prediction of Wind Power: A Data-Driven Approach," *IEEE Transactions on Energy Conversion*, vol. 25, no. 4, pp. 1112–1122, dec 2010. [Online]. Available: http://ieeexplore.ieee.org/document/5451084/

[14] M. He, L. Yang, J. Zhang, and V. Vittal, "A Spatio-Temporal Analysis Approach for Short-Term Forecast of Wind Farm Generation," *IEEE Transactions on Power Systems*, vol. 29, no. 4, pp. 1611–1622, jul 2014. [Online]. Available: http://ieeexplore.ieee.org/document/6727513/

[15] K. Bhaskar and S. N. Singh, "AWNN-Assisted Wind Power Forecasting Using Feed-Forward Neural Network," *IEEE Transactions on Sustainable Energy*, vol. 3, no. 2, pp. 306–315, apr 2012. [Online]. Available: http://ieeexplore.ieee.org/document/6170987/

[16] A. Bossavy, R. Girard, and G. Kariniotakis, "Forecasting ramps of wind power production with numerical weather prediction ensembles," *Wind Energy*, vol. 16, no. 1, pp. 51–63, jan 2013. [Online]. Available: http://doi.wiley.com/10.1002/we.526

[17] D. Barbosa de Alencar, C. de Mattos Affonso, R. Limão de Oliveira, J. Moya Rodríguez, J. Leite, and J. Reston Filho, "Different Models for Forecasting Wind Power Generation: Case Study," *Energies*, vol. 10, no. 12, p. 1976, nov 2017. [Online]. Available: http://www.mdpi.com/1996-1073/10/12/1976

[18] Q. Xu, D. He, N. Zhang, C. Kang, Q. Xia, J. Bai, and J. Huang, "A Short-Term Wind Power Forecasting Approach With Adjustment of Numerical Weather Prediction Input by Data Mining," *IEEE Transactions on Sustainable Energy*, vol. 6, no. 4, pp. 1283–1291, oct 2015. [Online]. Available: http://ieeexplore.ieee.org/document/7116614/

[19] M. Schlechtingen, I. F. Santos, and S. Achiche, "Using Data-Mining Approaches for Wind Turbine Power Curve Monitoring: A Comparative Study," *IEEE Transactions on Sustainable Energy*, vol. 4, no. 3, pp. 671–679, jul 2013. [Online]. Available: http://ieeexplore.ieee.org/document/6462006/

[20] H. Pousinho, V. Mendes, and J. Catalão, "A hybrid PSO–ANFIS approach for short-term wind power prediction in Portugal," *Energy Conversion and Management*, vol. 52, no. 1, pp. 397–402, jan 2011. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0196890410003080

[21] A. Marvuglia and A. Messineo, "Monitoring of wind farms' power curves using machine learning techniques," *Applied Energy*, vol. 98, pp. 574–583, oct 2012. [Online]. Available: http://dx.doi.org/10.1016/j.apenergy.2012.04.037https://linkinghub.elsevier.com/retrieve/pii/S0306261912003236

[22] H. Zheng and A. Kusiak, "Prediction of Wind Farm Power Ramp Rates: A Data-Mining Approach," *Journal of Solar Energy Engineering*, vol. 131, no. 3, pp. 0 310 111–0 310 118, aug 2009. [Online]. Available: https://asmedigitalcollection.asme.org/solarenergyengineering/article/doi/10.1115/1.3142727/419017/Prediction-of-Wind-Farm-Power-Ramp-Rates-A

[23] S.-J. Kim and I.-Y. Seo, "A Clustering Approach to Wind Power Prediction based on Support Vector Regression," *International Journal of Fuzzy Logic and Intelligent Systems*, vol. 12, no. 2, pp. 108–112, jun 2012. [Online]. Available: http://koreascience.or.kr/journal/view.jsp?kj=E1FLA5{&}py=2012{&}vnc=v12n2{&}sp=108

[24] H. Zareipour, D. Huang, and W. Rosehart, "Wind power ramp events classification and forecasting: A data mining approach," in *2011 IEEE Power and Energy Society General Meeting*. IEEE, jul 2011, pp. 1–3. [Online]. Available: https://ieeexplore.ieee.org/document/6039625/

[25] Mi Yeong Hwang, Cheng Hao Jin, Yang Koo Lee, Kwang Deuk Kim, Jung Hoon Shin, and Keun Ho Ryu, "Prediction of wind power generation and power ramp rate with time series analysis," in *2011 3rd International Conference on Awareness Science and Technology (iCAST)*. IEEE, sep 2011, pp. 512–515. [Online]. Available: http://ieeexplore.ieee.org/document/6163182/

[26] A. J. Zavala and A. R. Messina, "Dynamic harmonic regression approach to wind power generation forecasting," in *2016 IEEE PES Transmission & Distribution Conference and Exposition-Latin America (PES T&D-LA)*. IEEE, sep 2016, pp. 1–6. [Online]. Available: http://ieeexplore.ieee.org/document/7805684/

[27] X. An, D. Jiang, M. Zhao, and C. Liu, "Short-term prediction of wind power using EMD and chaotic theory," *Communications in Nonlinear Science and Numerical Simulation*, vol. 17, no. 2, pp. 1036–1042, feb

2012. [Online]. Available: http://dx.doi.org/10.1016/j.cnsns.2011.06.003https://linkinghub.elsevier.com/retrieve/pii/S1007570411003182

[28] N. Safari, C. Y. Chung, and G. C. D. Price, "Novel Multi-Step Short-Term Wind Power Prediction Framework Based on Chaotic Time Series Analysis and Singular Spectrum Analysis," *IEEE Transactions on Power Systems*, vol. 33, no. 1, pp. 590–601, jan 2018. [Online]. Available: http://ieeexplore.ieee.org/document/7902233/

[29] I. Yenidogan, A. Cayir, O. Kozan, T. Dag, and C. Arslan, "Bitcoin Forecasting Using ARIMA And PROPHET," in *3rd International Conference on Computer Science and Engineering*, vol. C, 2018, pp. 621–624.

[30] N. Zhao, Y. Liu, J. K. Vanos, and G. Cao, "Day-of-week and seasonal patterns of PM2.5 concentrations over the United States: Time-series analyses using the Prophet procedure," *Atmospheric Environment*, vol. 192, no. April, pp. 116–127, nov 2018. [Online]. Available: https://doi.org/10.1016/j.atmosenv.2018.08.050https://linkinghub.elsevier.com/retrieve/pii/S1352231018305715

[31] U. Shafique and H. Qaiser, "A Comparative Study of Data Mining Process Models (KDD , CRISP-DM and SEMMA)," *International Journal of Innovation and Scientific Research*, vol. 12, no. 1, pp. 217–222, 2014. [Online]. Available: http://www.ijisr.issr-journals.org/

[32] A. Sarkar and D. K. Behera, "Wind Turbine Blade Efficiency and Power Calculation with Electrical Analogy," *International Journal of Scientific and Research Publications*, vol. 2, no. 2, pp. 1–5, 2012.

[33] A. A. Dragulescu, "Package ' xlsx '," *Cell*, 2012.

[34] D. Z. Meyer and L. M. Avery, "Excel as a Qualitative Data Analysis Tool," *Field Methods*, vol. 21, no. 1, pp. 91–112, feb 2009. [Online]. Available: http://journals.sagepub.com/doi/10.1177/1525822X08323985

[35] J. Corbyn, "Time Series Analysis with Applications in R," *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, vol. 174, no. 2, pp. 507–507, apr 2011. [Online]. Available: http://doi.wiley.com/10.1111/j.1467-985X.2010.00681{_}4.x

[36] R. Wesley, M. Eldridge, and P. T. Terlecki, "An analytic data engine for visualization in tableau," in *Proceedings of the 2011 international conference on Management of data - SIGMOD '11*. New York, New York, USA: ACM Press, 2011, p. 1185. [Online]. Available: http://portal.acm.org/citation.cfm?doid=1989323.1989449

[37] R. Lund, "Time Series Analysis and Its Applications: With R Examples," *Journal of the American Statistical Association*, vol. 102, no. 479, pp. 1079–1079, sep 2007. [Online]. Available: http://www.tandfonline.com/doi/abs/10.1198/jasa.2007.s209

[38] A. V. Metcalfe and P. S. Cowpertwait, *Introductory Time Series with R*. New York, NY: Springer New York, 2009. [Online]. Available: http://link.springer.com/10.1007/978-0-387-88698-5

[39] R. J. Hyndman and Y. Khandakar, "Automatic Time Series Forecasting: The forecast Package for R," *Journal of Statistical Software*, vol. 27, no. 3, 2008. [Online]. Available: http://www.jstatsoft.org/v27/i03/

[40] A. C. Harvey and S. Peters, "Estimation procedures for structural time series models," *Journal of Forecasting*, vol. 9, no. 2, pp. 89–108, mar 1990. [Online]. Available: http://doi.wiley.com/10.1002/for.3980090203

[41] S. J. Taylor and B. Letham, "Forecasting at Scale," *The American Statistician*, vol. 72, no. 1, pp. 37–45, jan 2018. [Online]. Available: https://www.tandfonline.com/doi/full/10.1080/00031305.2017.1380080

[42] M. Bujosa, A. García-Ferrer, and P. C. Young, "Linear dynamic harmonic regression," *Computational Statistics & Data Analysis*, vol. 52, no. 2, pp. 999–1024, oct 2007. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0167947307002666

[43] R. J. Hyndman and G. Athanasopoulos, "Forecasting: Principles and Practice: Notes," 2014.

[44] T. Chai and R. R. Draxler, "Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature," *Geoscientific Model Development*, vol. 7, no. 3, pp. 1247–1250, jun 2014. [Online]. Available: https://www.geosci-model-dev.net/7/1247/2014/

[45] T. Burton, N. Jenkins, D. Sharpe, and E. Bossanyi, *Wind Energy Handbook, Second Edition*, 2011.

[46] S. Dixon and C. Hall, "Wind Turbines," in *Fluid Mechanics and Thermodynamics of Turbomachinery*. Elsevier, jan 2014, pp. 419–485. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780124159549000103https://linkinghub.elsevier.com/retrieve/pii/B9780124159549000103

Appendices

From the units of all the parameters mentioned in the parameter description of equation(5), the lowest derivation gives the unit of kg.m^2/sec^3 on the Right Hand Side(R.H.S) of the equation. Force can stand as the product of mass and acceleration due to gravity, so it can be expressed as kg.m/sec^2. Work can be represented in the form of the product of force and displacement whose unit representation would be kg.m^2/sec^2. When the work is measured in unit time, it can be termed as power and can be represented as kg.m^2/sec^3. So it can be proved that the aforementioned equation stands correct. The calculated wind power from this research will bear the same S.I. unit of Watt(W).