Wind Electricity Generation Forecasting using Time Series Analysis Techniques in India

MSc Research Project
Data Analytics

Sonali Gupta
x01527245

School of Computing
National College of Ireland

Supervisor: Dr.Catherine Mulwa

National College of Ireland
Project Submission Sheet – 2017/2018
School of Computing

Student Name:	Sonali Gupta
Student ID:	x01527245
Programme:	Data Analytics
Year:	2018
Module:	MSc Research Project
Lecturer:	Dr.Catherine Mulwa
Submission Due Date:	13/08/2018
Project Title:	Wind Electricity Generation Forecasting using Time Series Analysis Techniques in India
Word Count:	6860

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 encouraged to use the Harvard Referencing Standard supplied by the Library. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action. Students may be required to undergo a viva (oral examination) if there is suspicion about the validity of their submitted work.

Signature:	
Date:	15th September 2018

PLEASE READ THE FOLLOWING INSTRUCTIONS:

1. Please attach a completed copy of this sheet to each project (including multiple copies).
2. **You must ensure that you retain a HARD COPY of ALL projects**, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer. Please do not bind projects or place in covers unless specifically requested.
3. 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):	

Wind Electricity Generation Forecasting using Time Series Analysis Techniques in India

Sonali Gupta
x01527245
MSc Research Project in Data Analytics

15th September 2018

Abstract

Electricity crisis is a major concern for developing countries and this problem can be fulfilled by various technologies of renewable energy such as water energy, solar energy, wind energy and many more. The project presented the idea of forecasting electricity generation using wind speed in two different cities of India: Jodhpur and Bengaluru. This thesis investigates the use of time series models for wind electricity forecast for both the cities. A contrast for four time series models i.e. TBATS, ARIMA, SES, and HOLTS is conducted and results are presented.The result of this study shows that the TBATS model is outperformed all the other developed models. The key finding is that TBATS model able to deliver accurate forecast of wind speed with MASE of 0.76%, RMSE of 1.41 %, MAE of 1.19%, and MAPE of 0.34% for 8 months forecast horizon. The results of forecasted value of best performing TBATS model are used in the computation of electricity generation for both cities. finalized that on the same investment for setting up wind power plant, electricity produced by Jodhpur is more than the Bengaluru city. Hence, Jodhpur is a better location to set up the wind power plant.

keywords- *Time series, Electricity, Forecasting, Wind.*

1 Introduction

In Introduction section comprises the idea of project background of how wind electricity is helpful for an environment and highly significant forecasting. In further, discussed the problem for the project in terms of research question and some sub_objectives.

1.1 Project Background and Motivation

Wind is one of the rapidly developing energy sources because of ample, renewable, Eco-friendly and free of pollution and it is also alternative of electricity generated from non-renewable sources such as fossil fuel which hazardous for our environment. The wind, the sun, and flowing water are the sources of renewable energy no one denies that these sources heavily depend on location and climate-condition. on the other hand, oil, gas, coal, and petroleum which heavily depend on mines of fossil fuel. Mentis et al. (2016)

India mostly rely on non-renewable sources which produce greenhouse gas pollution and global warming. The total population of India passes 1.32 billion while 240 million citizens still need electricity supply, a point that might be in an upcoming year facing energy crisis due to growing population and possibly drop economic progress. So, wind energy has lots of benefits such as pollution free, cost-effective and low levels of greenhouse gas emission as perDaut et al. (2012). This research focuses on wind electricity generation forecasting in south and west-northern of India. Currently, there is power shortage countrywide and with a population of more than billion people and energy demand skyrocketing every year, this project would significantly have an impact in improving energy crisis. Additionally, India geographical goldmine has a number of the potential coastal zone around 76000 km which can be used in the generation of electricity with the help of wind energy Bakshi (2002). This research highlight on the subject forecasting wind electricity generation and used different time series model which exclusively based on wind power prediction like Autoregressive moving average (ARIMA), Simple Exponential Smoothing (SES), HOLT's and Trigonometric Regressors box-cox transformation Arma errors trend seasonality (TBATS) model. Not only need to forecasting also compare the forecasting model performance evaluation using Mean absolute percentage error (MAPE), Root mean square error (RMSE), Mean absolute scaled error (MASE) and Mean absolute error (MAE) according to Hyndman and Koehler (2006). Additionally, on the basis of error rates find the best-forecasted model to predict wind electricity generation of Jodhpur and Bengaluru city in India. It would be helpful to discover the better place for the set up the wind power plant.

1.2 Project Requirement Specification

1.2.1 Research Question

The Research question address the problem of energy crisis which affects the people living in North and South region of India due to lack of proper forecasting Mechanism in the Indian government.
RQ: *"Can we efficiently improve/enhance and reduce future energy crisis of India due to (Pollution, population, public demand) by forecasting wind electricity considering factors (wind speed, Relative humidity, Air temperature, Bar pressure, wind direction) using time series techniques(ARIMA ,SES, HOLT's TBATS)?"*
In order to solve the research question, the following research objectives (refer figure1) are specified and tackled.

Figure 1: Research Objectives

Objectives	Description
Objective1	A critique and investigation of wind electricity generation forecasting in India (2002-2018)
Objective2	Data prepossessing, Implementation, Evaluation and Results of wind electricity forecasting models in Jodhpur and Bangalore cities using time series techniques/algorithms.
Sub_Objective2 (i)	Central pollution control Board (CPCB) Dataset Prepossessing
Sub_Objective2(ii)	Implement, Evaluate and results of ARIMA model
Sub_Objective2(iii)	Implement, Evaluate and results of SES model
Sub_Objective2(iv)	Implement, Evaluate and results of HOLT's model
Sub_Objective2(v)	Implement, Evaluate and results of TBATS model
Objective3	Comparison of Developed Models (Sub_Objective (i) to Sub_Objective (v)
Objective4	Choose the best forecasting models in terms of Accuracy result perform/computation of Wind Electricity generation
Objective5	Comparison of Computed Results for Jodhpur v/s Bangalore to determine which city generate more electricity and provide suggestions on how the government administrators of two cities can set up power plant in future.

The rest of technical report is structure as followings. Chapter 2 presents a critical review of related work in forecasting wind electricity generation using forecasted time series techniques and some other prediction model. Chapter 3 presents the scientific methodology approach used and project design. In Chapter 4 presents the implementation of models Chapter 5 evaluations and Chapter 6 results of wind electricity forecasting models. Finally, chapter 7 presents a conclusion and recommended future work.

2 Literature Review of Wind Electricity Forecasting (2000-2018)

2.1 Introduction

Forecasting wind electricity generation is an imperative study of wind electricity. There are several types of models for forecasting wind speed and they are categorized as Time Series (univariate and multivariate), neural network, statistical, intelligent, hybrid and fuzzy-logic model. This section represents related work done by number of researchers related to forecasting wind electricity with different machine learning techniques. In this section, thorough study of techniques applied by researchers so get idea about which model we will used in my thesis. All the sources, papers, journals and articles are available on Google scholar, National College of Ireland digital library that support my thesis study.

2.2 A Review of Input Variables used in Forecasting Wind Electricity

Forecasting model need some input parameter for forecasting wind electricity generation. This section examined some input parameter for forecasting wind power. Abdullah et al. (2014) used wind farm data for prediction and discussed some crucial variables such as pressure, temperature, wind speed where without wind speed its difficult to predict wind power All these variables improve the forecasting accuracy of models. Pinto et al. (2014) has worked on historical real time data of three years for forecasting wind speed of 5-minute interval and defined that wind electricity generation highly reliant on wind speed. Kumar and Malik (n.d.)have worked on 26 different cities of India and extracted the dataset from NASA. The input variables which used for wind speed prediction are: Air temperature, Elevation, relative humidity, heating degree days, cooling degree days, longitude, latitude, solar radiation and earth temperature in their research. Sreelakshmi and Kumar (2008) wind speed forecasting plays an important role in the field of missile, target tracking, power shortage and satellite launching. She used moisture content, humidity, atmospheric pressure and rainfall recorded from nearby weather station as input variables for forecast wind speed. Kwon (2010) seasonal plays important role producing wind energy. Therefore, author investigate their study in seasonal effect at the time of producing wind energy. The case study on Kwangyang Bay 3 months of Datasets and applied Monte Carlo models and concluded that wind energy is high in winter season as compared to summer season. Taylor et al. (2009)have worked on wind power density forecasting of UK wind farm. In dataset have some seasonality and presented a plot between wind speed and month of year which show wind speed is high in the month of January, February, march in winter season and low in the month of May, June, July in summer season. The above literature review helps to identify the input variables to perform forecasting wind electricity generation.

2.3 A Review of Existing Models and Techniques of Forecasting Models

Rallapalli and Ghosh (2012) case study of importance of electricity demand of India. In September 2011, total installed capacity 182,245 MW. In future India will be facing high electricity shortage[1] According to Central Electricity Authority (CEA) printed energy Shortage 8.5% and peak demand 9.8% [2]. According to Regions energy and peak demand in India are shown in Table 1

Electricity shortage demand highly effect the economy of the nation and due to deficit of electricity need to complete the demand by region. This project trying to resolve energy crisis with the help of prediction technique can easy to forecast the electricity demand. This study used ARIMA-EGARCH, neural network and exponential shooting

[1]https://www.technologyreview.com/s/542091/indias-energy-crisis/
[2]http://www.cea.nic.in/

Table 1: Energy deficit and Peak Shortage

Region	Energy deficit (millions unit)	Peak Shortage (MW)
Western	11.0	10.9
Southern	10.5	14.5
Eastern	7.7	11.6
Northern	10.9	11.9
North-Eastern	0.3	5.9

techniques. Electricity parameter is most vital for growth of any country. Maharashtra faced lots of electricity crisis and shortage in many areas besides industrial and technology field are rapidly rise. Discussed the electricity uses in Domestic, commercial, Industrial, agricultural, public services and railways. Mostly electricity demand fulfills by thermal 51% and by renewable sources only 13%. Need more power plants in future to meet energy demands as per Kale and Pohekar (2012).The use of Renewable sources is highly neglected in India. Author used solar energy technology for producing electricity. The study discussed about the benefits of Indias weather. North region of India has enough sun shine for produce electricity with the use of sun. moreover, constructed the project with the help of solar technology to set up the solar power plant in Gujarat location as per Kale and Pohekar (2012) Mabel and Fernandez (2008) Tamilnadu state in India started the wind energy generation program because of ample amount of geographical location. This area surrounded by three mountains which produce high amount wind. Author used past three years wind farm data for forecasting electricity with the help of neural network techniques and Matbox tool used for modelling. Meenal et al. (2018)Manalli produced solar energy in 26 states of India with the help of Random forest (RF) algorithm using WEKA tool. Singh et al. (2016)Agriculture plays important role for developing country like India. Electricity demand is increasing and conventional sources are fixed. The wind, tidal, Bio energy and water can complete the electricity demand. Author used the Indian agriculture data for wind speed forecasting using ANN techniques. As perKočenda and Černỳ (2015) investigated a nature of time series analysis. Any data that change periodically over time is called time-series. Time series data have worked on some important properties for analysis. When time depend on one variable called univariate time series and when depends on multiple variable called multivariate time series. The time series data consist of many components such as trend, seasonal pattern, white noise and irregular patterns. Trend defined as data grown over time in a long period without irregular effects. A component of seasonal pattern influenced by seasonal factor (the month, the year etc). white noise component does not show any data that would useful for analysis and with arbitrary variables with zero means and no correlation with other variables. Mahalakshmi et al. (2016) worked on forecasting time data and use time series models such as ARIMA. According to As' ad (2012) ARIMA model appropriate for forecasting electricity demand. As' ad (2012) experimented with ARIMA model to find best model for forecast electricity demand. ARIMA model applied on four different type of time series data three, six, nine and twelve months. Compared the forecast performance using root mean square error (RMSE) and mean absolute percentage error (MAPE) and determined that earlier three-month data appropriate for forecast seven days ahead and past six months data is best for one day ahead forecast. Eldali et al. (2016)shows how ARIMA model help in wind power forecasting. Author used real time hourly data on Electrical Reliability Council of Texas to improve the day ahead forecasting. The benefit of ARIMA model is to be worked under when data is non- stationary. Das and Banerjee (2017)worked on fluctuated wind speed data of Maharashtra city of India. Wind speed data is non-stationary time series data and applied Fractional ARIMA (FARIMA) model which have ability to capture long range correlation and compared the model with ARIMA and ARMA and concluded that FARIMA model is appropriate for long range correlation data. Zhang et al. (2017)build hybrid model for forecasting short term wind speed. Authors used Empirical mode decomposition (EEMD) for extracting data from periodic series and non-linear time series, SARIMA for periodic time series data and Adaptive neural network based fuzzy inference system (ANFIS) for non-linear time series data. the predicted result compared with MAPE and found that Empirical model gave more accurate results. Erdem and Shi (2011) introduced different approaches of ARMA model such as linked ARMA model used for predicting wind direction, traditional ARMA model for predicting wind speed, vector Autoregression (VAR) and restricted version of (VAR). all approaches applied 1 hour ahead forecasts of hourly average wind attributes of two different wind sites in North Dakota, USA. The performance is evaluated by MAE and traditional ARIMA outperformed with others model. Lau and McSharry (2010) proposed multistep density forecast for 64 offshore wind farms in Ireland using ARIMA-GARCH model for multistep forecast and exponential smoothing model forecast density. This approach is able to forecast wind speed 15 minutes to 24 hours ahead. Liu et al. (2012) conducted a comparison between two time series models for predicting wind speed. They applied ARIMA-ANN, ARIMA-Kalman hybrid and Pure ARIMA model to compared their performance on the basis of MAPE, MSE, and MASE. When ARIMA Kalman hybrid model compared with ARIMA model, then ARIMA Kalman gave better percentage than ARIMA and when ARIMA ANN model compare with ARIMA then again ARIMA ANN gave three steps better MAPE error with ARIMA. Moreover, when ARIMA Kalman and ARIMA ANN compare itself then Kalman part gave better performance than ARIMA ANN. And final they concluded that ARIMA Kalman and ARIMA ANN are good for wind speed prediction and also applicable for real time wind power. Zhang et al. (2017)developed the self-adaptive auto regressive integrated moving average with exogenous variable (Adaptive ARIMAX) model which used for wind farm dataset of north region of China and also compared with ARIMAX model for Weather Research Forecasting (WRF) and ARMAX model. Performance compared with RMSE. Adaptive-ARIMAX model performs best with other time series models. Ramasamy et al. (2015) used Artificial Neural Network (ANN) model for prediction of wind speed of 11 different location in Himalaya regions of India. The wind data of Hamirpur location used for training and testing with parameters such as temperature air pressure and solar radiation are used as Input to forecast wind speeds. The output of Mean Absolute Percentage Error (MAPE) and Correlation Coefficient are 4.65 percent and 0.97. Authors also compared the model with Gurgaon city of Haryana with MAPE and Correlation Coefficient are 6.50 percent and 0.99 respectively with high prediction accuracy with settled model. and concluded that wind speed prediction in Himalaya regions are suitable for micro wind turbines. do Nascimento Camelo et al. (2018) produced hybrid model for forecast monthly region of wind speed in Brazil north region using ARIMAX (for linear function) and ANN (for non-linear functions) and shows that combined model performed better as compare to forecast separately. Chen et al. (2018) proposed a novel method called EnsemLSTM using Long Short-Term Memory neural network (LSTMs), External Optimization (EO) and Support Vector Regression Machine (SVRM) for Short Term Wind Speed forecasting in Mongolia and China. Additionally, compared the performance of EnsemLSTM model with other models and demonstrated that proposed EnsemLSTM model give better results. In terms of Hybrid Model, Liu et al. (2015) forecast the wind speed using Wavelet packet decomposition (WPD) after they used density based spatial clustering of application with noise (DBSCAN) and for find the structure of wind series, after applied the Elman neural network (ENN) and combined these and they created (WPD-DBSCAN-ENN) model. Afterwards performance compared with WPD-ENN hybrid model and ENN mode with error rate and justify that WPD-DBSCAN-ENN model performed better with other models. Khosravi et al. (2018)developed a study based on Osorio wind farm in Brazil and applied Multilayer feed forward neural network (MLFFNN), Group method of data handling (GMDH), Fuzzy interference System (FIS), Support vector regression (SVR) and adaptive neuro-fuzzy inference system (ANFIS) are used to predict wind speed data of Osorio farm and all the models inspected 10 min, 15 min and 30 min intervals and verified by GMDH model and also compared their performance where GMDH model outperformed with low RMSE and MAE rate.

TBATS model have not used prior for forecasting wind electricity generation as per my acquaintance. There is no literature paper found which based on forecasting wind electricity generation. However, this model has been used in many others forecasting applications. Naim et al. (2018) ARIMA, SARIMA model worked to handle single seasonality but when multiple seasonality exists, these models perform moderate. BATS and TBATS relatively new approach to handle complex seasonal data. BROŻYNA et al. (2018) by applying TBATS model forecast the long period of electric energy demand in Poland. Moreover, this model also used for Gold price forecasting Hassani et al. (2015).As wind data set used in this project which also have long, non-linear features and complex seasonality are presents this is the reason to choose this model.

2.4 Conclusion

Based on the results of the reviewed literature there is a need to develop wind electricity forecasting models for Jodhpur and Bengaluru. This review is truly helped to get idea about which model will be using in this thesis.From the investigation of literature review able to answer research objective Table 1. As per review from Kočenda and Černỳ (2015), As' ad (2012) and Zhang et al. (2017), Eldali et al. (2016) provided the knowledge of ARIMA model better for univariate variable and better result than the other time series models for forecasting wind speed. Also give idea of SES, HOLT's model which also used for this project. Therefore, I have applied this model in my project for better results.This study mainly focused on TBATS model. The author Hassani et al. (2015) and BROŻYNA et al. (2018) gave rigid overview of use of forecasting so considered in this project. Earlier literature review as per Khosravi et al. (2018) (2018), Erdem and Shi (2011) evaluated the performance using RMSE, MASE, MAE and MAPE. In that way consider all important, valuable and positive notions.The next chapter presents the scientific methodology approach used and the architectural design.

3 Scientific Methodology Approach Used and Project Design

3.1 Introduction

This methodology section includes the process flow of the project and explanation of modified methodology. moreover, described the the architectural design for this project.

3.2 Scientific Methodology Approach Used

The process of discovering important insights from the raw data for useful decision making is known as data mining according to author Han et al. (2011). The standard process of data mining methodologies such as Cross Industry Process of Data Mining (CRISP-DM), Knowledge discovery and data mining and Simple, Explore Modify, Model, and Assess (SEMMA). This research follows the CRISP-DM methodology. Crisp-dm is a hierarchical procedure which contain six stages breakdown [3]. Generalized CRISP-DM modified according to the requirement as represents in Figure 2.

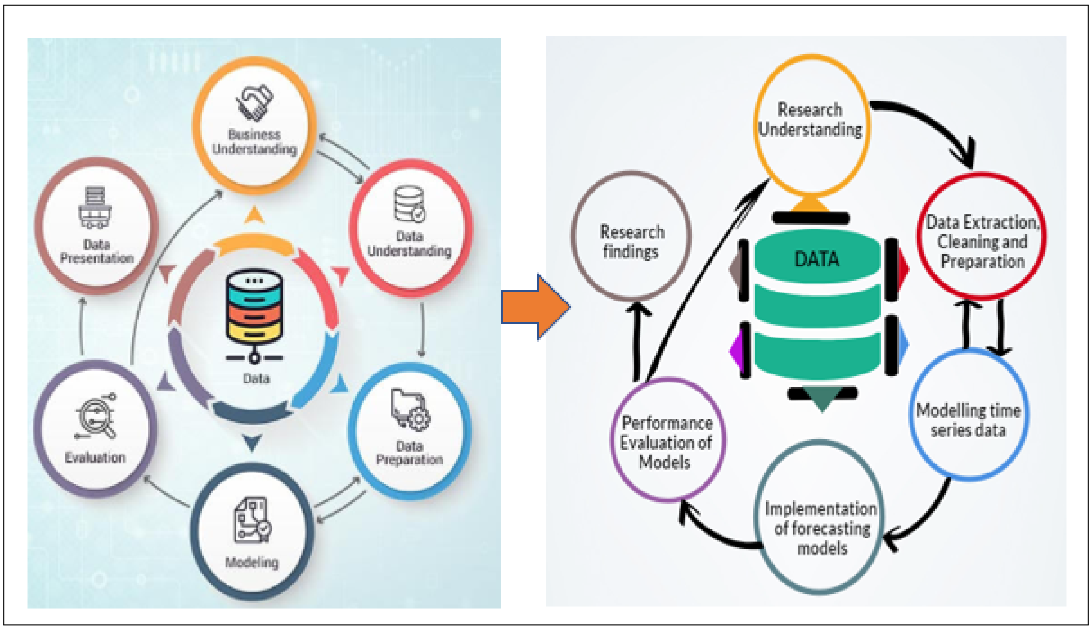

Figure 2: Scientific Methodology Approach Used

First business understanding change as the research understanding of project, second stage modified as data extraction, cleaning and preparation, third stage about the modelling of time series techniques, next stage Implementation of forecasting models, fifth stage discuss the Evaluation and Results, last stage about the final research findings about the project.

[3]https://www.sv-europe.com/crisp-dm-methodology/

3.3 Architectural Design

To develop better, robust and secure project for forecasting wind electricity, architecture diagram is created for better result, performance and accuracy. This diagram presents the techniques, technology and tools used for this project, how they used and in which order to followed, for achieve the aim to generate wind electricity forecasting with better accuracy. This architecture has also connectivity with modified CRISP-DM methodology. Figure 3 illustrates the Architecture diagram of this project.

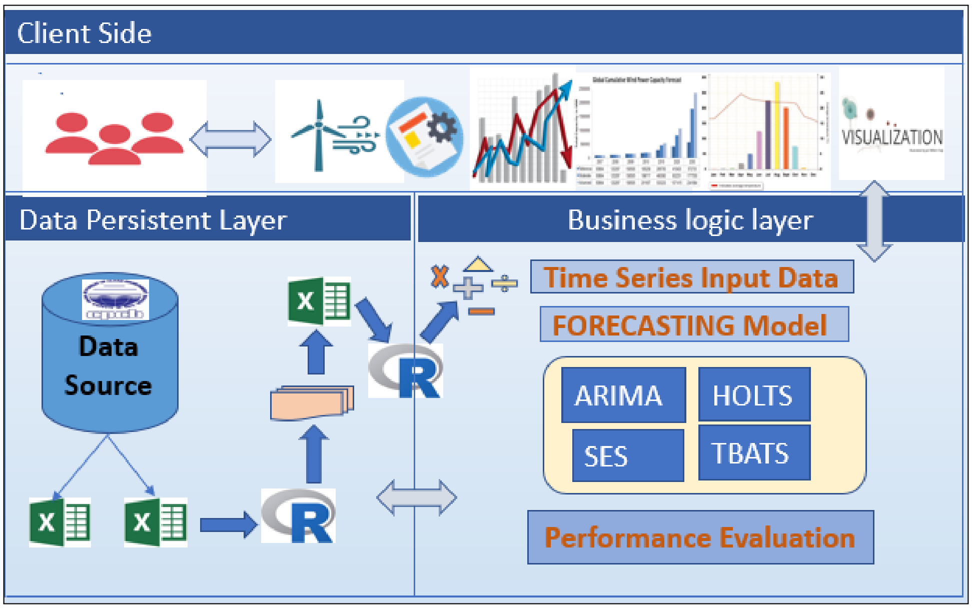

Figure 3: Architecture Diagram

The Architecture diagram follow the three-layer known as Data persistence layer, Problem/Business requirement layer and client persistence layer. Data persistence layer about the data sources, data extraction then cleaning, pre-processed and transformation of the data. this layer is linked with problem/ business requirement layer which provides the solution of how to forecasting techniques help to generate and predict wind electricity. And preceding with Client side, where data analyst team and client interact to give resolution of problem in terms of report and results.

3.4 Conclusion

Hence, for this project a modified CRISP DM methodology approach is used as per its fits best for the project research. Also, process flow diagram helps to solve how data analyst can worked on this problem. Furthermore, a detail discussion is enlightened in the dataset pre-processing and the implementation of the models.

4 Data Preparation and Implementation of Wind Electricity Forecasting Models in Jodhpur and Bengaluru

4.1 Introduction

This chapter discussed about how to extract data, data cleaning and pre-processed the data for the project and implementation of each of the time series model and performance evaluation of the model.

4.2 Central Pollution Control Board Dataset Prepossessing

For the start of any project need proper data set since project is about the forecasting wind electricity. For achieving the aim data extracted from the source Central Pollution Control Board (CPCB) organization under the act of Ministry of Environment and forest, Govt.of India[4]. CPCB website contains Air quality, water quality and Noise Monitoring data. this project considers the Air Quality of two different locations. The first location is Jodhpur city from Rajasthan state with (Latitude 26.2389 N, Longitude 73.0243 E) and another location is Bengaluru city from Karnataka state with (latitude 12.9716N, Longitude 77.5946E). Time period of dataset is from (January 1, 2014) to (July 25, 2018). Jodhpur dataset taken from Jodhpur station and Bengaluru dataset taken from BTM station. The parameters which has taken as per the requirement of forecasting wind electricity problem such as Wind Speed, Wind Direction, Relative Humidity, Temperature, and Bar Pressure. In this way two separate data extracted from the source CPCB website. Briefly explore the parameters. Wind speed is one of the major factors for forecasting wind electricity as per Pinto et al. (2014) and the without wind speed cannot predict the wind electricity so this was the reason to choose this parameter. Temperature and relative humidity are the part of atmosphere which change according to season and help in forecasting environment problem. In Abdullah et al. (2014) research used wind speed, temperature in forecasting for better accuracy. As per the literature review parameters are selected for this project. Figure4 presents the Data pre-processing flow.

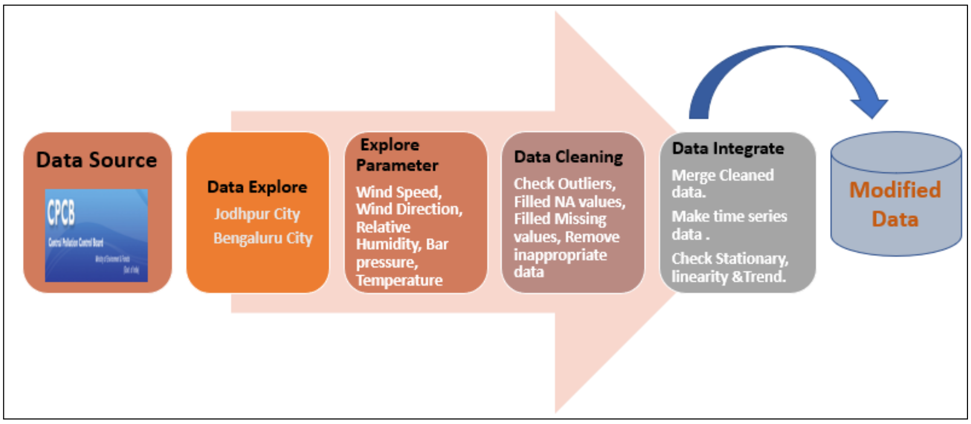

Figure 4: Data preparation process flow diagram

Dataset cleaning is a significant part of the project for better and accurate results

[4]http://www.cpcb.gov.in/CAAQM/frmUserAvgReportCriteria.aspx

because noisy data can lead to inappropriate results. The parameters of the dataset in different csv file so integrate all parameters such as wind speed, wind direction, bar pressure, relative humidity and temperature all are merged in one CSV. This dataset has NA values, missing columns, blank values. After the extraction dataset imported in R using read.csv() function. First dataset explores using summary() functions and check the outliers to check that values out of range. Missing values checked with is.na() and after that missing values filled with na.locf() function means last observation carried forward. Many blank values filled with na.mean(), na.random() functions. Many others values filled using excel.Modified data change in to time series data using R packages: tseries and zoo are the time series package as per Shumway and Stoffer (2011). The resulted time series data univariate and multivariate used for forecasting wind electricity. Data of both the cities examine with the plot to check the noise seasonality and trend. Time series data analyzed for stationary, non-stationary and linearity using qqnorm plot, Box test, autocorrelation function (ACF) and QQline plot. In this way able to answer Sub_Objective(i) from Research Objectives figure 1. Additionally, following discussion enlightened the implementation of forecasting time series model for wind electricity generation.

4.3 Implementation

To more important evidence of the wind time series data needs to see the insights. forecast, zoo, ggplot2, tseries and fpp2 packages required for this project. Download all the packages using install.packages() function. First, read the data by importing using read.csv() function. For converting modified data into time series format used zoo() function. In ts() function set the argument wind speed with frequency 365.25 (mean daily observation data), to check the time series data used is.ts() function. To check data significant or not applied Box.test() with type Ljung. In time series data, it is imperative to identify data have white noise or not. To check the white noise applied autocorrelation function such as acf() and pacf() functions. Resulted wind speed data is not white noise even some relationship with the next variables. After that using plot() function wind speed data presented in a graphical format. For the multiplicative components such as trend, seasonality and noise have presented using decompose() function and then plotted. for the accurate more results wind speed data transformed and stabilized using Box.CoxLamda() function.

4.3.1 Implementation of ARIMA Model

ARIMA is a standard time series model and highly used for forecasting applications. The model has ability to change non-stationary data in to stationary data and worked with univariate model as perBox et al. (2015). The representation of this model is below:

ARIMA (p, d, q)

p is number of auto regressive term;**d** is non-seasonal difference;**q** is moving average;
p difined the number of lags and d represents differencing means subtract current and past values d times and used stabilize the non- stationary data. q shows lags of prediction error. When Autoregressive (d) and moving average (q) are zero then model becomes pure stationary ARMA model, when autoregressive (p) and non-seasonal difference (d) are zero, the model becomes pure stationary Moving average model¡[5].

[5]https://otexts.org/fpp2/arima.html

Implementation Process: ARIMA model applied using auto.arima() function in wind time series data and pass the differencing value and boxcox transformation for stabilized the variance. All the backend coding process in forecast package auto.arima(). Forecasted value of wind speed for next 240 days forecasted using forecast() function and plot the forecasted value using autoplot() function in graphical format. After, that forecasting values extracted and saved in directory using write.csv() function. additionally, check the residuals of forecasted value.

4.3.2 Implementation of TBATS Model

According to De Livera et al. (2011); TBATS model have ability to handle complex seasonality, high frequency data, non-integer stationary and multi seasonality time series data[6].T stands for Trigonometric term used for seasonality; B stands for Box-cox transformations for Heterogeneity; A stands for ARMA error, T stands for Trend and S for Seasonality components. Mathematical representation shown in equation 1

$$\textbf{TBATS}(\mathbf{w},\phi,\mathbf{p},\mathbf{q},\{\mathbf{x}_1, y_1\}, \{x_1, y_1\}, ..., \{x_t, y_t\}) \tag{1}$$

w is Box cox Transformations for heterogeneity; ϕ is Damping parameters (levels and trend); p,q is Arma Parameters; x is Seasonality period; y is Seasonal Components or Fourier pairs;

Implementation Process For TBATS model is applied using tbts() function in wind time series data and available in forecast package. Wind speed forecasted value of next 240 days forecasted by tbts() function and forecasted value plotted by autoplot() function. again, forecasted values file extracted saved in csv in directory using write.csv function.

4.3.3 Implementation of Simple Exponential Smoothing Model

Simple Exponential Smoothing (SES) method help for prediction when no trend or seasonality in the data.SES model forecast value are calculating using weighted averages, and weights are decreasing exponentially[7].The mathematical representation is shown in equation 2 below:

$$x_{t+1} = \alpha x_t + \alpha(1-\alpha)x_{t-1} + ... \tag{2}$$

Implementation Process For Simple Exponential Smoothing model is applied using ses() function in wind time series data and this function presented in forecast package. Wind speed forecasted value of next 240 days (for 8 months) forecasted by ses() function. with the help of autoplot() function presented forecasted value then forecasting values extracted saved as csv in directory using write.csv() function.

4.3.4 Implementation of HOLT'S model

HOLT's method is used when linear trend and seasonal components variation in dataset.The model needs to be separate smoothings constants for slope and intercepts highly usesd in forecasting applications as per Kalekar (2004).The mathematical representation is shown in equation 3 below:

[6]https://robjhyndman.com/hyndsight/tbats-with-regressors/
[7]https://otexts.org/fpp2/expsmooth.html

$$H_{t+c} = (S_t + cd_t)K_{t-U+c} \quad (3)$$

where H is the forecast at c period, S is Smoothing exponential observation, d is trend and t is the time period.
Implementation Process For HOLTS model is applied using holt() function and the function backend forecast process presented in this function. wind speed forecasted value of next 240 days (8 months) forecasted by holt() function. forecasted values presented in graphical format using autoplot() function. again, forecasted values file extracted and saved in csv in directory using write.csv function.

4.3.5 Wind Electricity Generation Computations

Burton et al. (2011)wind electricity generation highly based upon wind speed and wind speed depends upon temperature, pressure, humidity and others weathers factors. When wind strike the blades of turbines and rotate the shaft which produce electricity. According Burton et al. (2011) universal formula to calculate wind electricity generation is described below:

$$\mathbf{E = 1/2}C_p\rho\mathbf{a}v^3$$

E= Electricity generated by wind turbines;
C_p = power coefficient;
ρ = Density of Air(1.25kg/m^3);
a = swept area of rotor;
v = Wind speed;
Forecasted wind speed of Jodhpur and Bengaluru cities are calculated from the above formula. Forecasted values of both the cities are compared to see which cities produced more wind electricity to set up the power plant. Sarkar and Behera (2012) discussed the calculation of wind power for India. [8]parameters of formula selected as Power coefficient value is 0.35, density of air value 1.25kg/m^3, swept area of rotar 8495 m^3, and used average wind speed of forecasted value month August to March 2019 for 100 wind plant. In this way calculated the wind electricity for both cities.
Model of the performance checked using error rates with MAPE, RMSE, MASE and MAE. The 80 % data used for training set and 20% data used for testing set. Mainly focused on the TBATS model how worked and performed in wind speed time series data.

4.4 Conclusion

Now,after the data pre-processed,implemented all models using R.The next section presents the forecasting evaluation of models.

5 Evaluation

5.1 Introduction

The evaluations were conducted using the pre-processed time series dataset was evaluated using i.e. seasonality, white noise, and, trend. This was important before using the

[8]https://www.raeng.org.uk/publications/other/23-wind-turbine

dataset in the implementation of the models.This was significant in order to get better accuracy results.https://v2.overleaf.com/project/5b65b6c01decdc6499ea57f4

5.2 Evaluation of pre-processd Dataset

A comparison of wind time series data of Jodhpur and Bengaluru cities is presented in figure 5. We can see the seasonality and some patterns from both the plots. In Jodhpur wind speed data there is some upward fluctuation occur in the month of July and august. In Bengaluru wind speed data some rising fluctuation in the month of September and October.

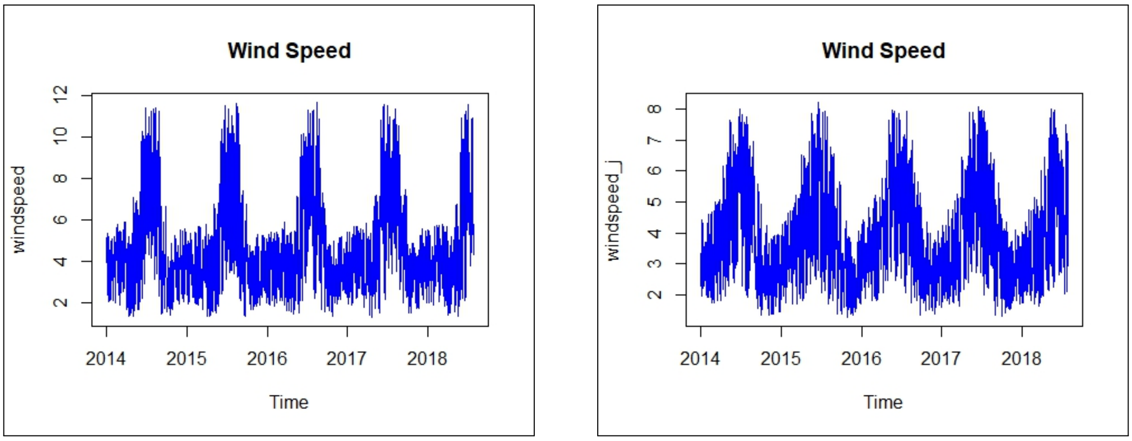

Figure 5: Bengaluru and Jodhpur Wind Speed Plot

Wind Speed data is verified by Ljung Boxtest to check data is white noise or not. The p-value is very small means data is not white and can be further used for forecasting. Additionally, for the better analysis about the time series data, applied Auto correlation function (ACF) to check the relationship between the points and and partial auto correlation function (PACF) to check for correlation with many lags such as relation with presents data with the past data. from both the figures 6. we can see that Jodhpur and Bengaluru, many spikes are significantly greater than zero, so data is not white noise and have some pattern, seasonality and can be further used for forecasting models.

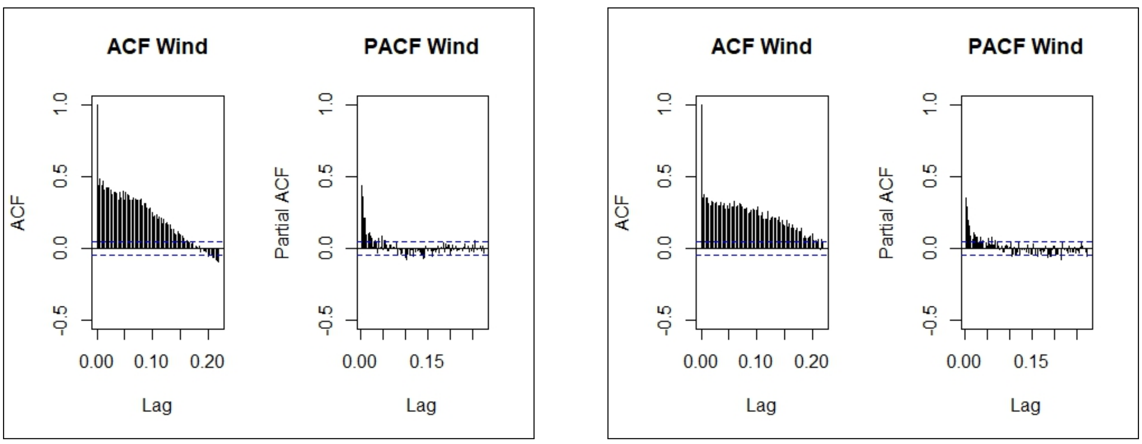

Figure 6: Jodhpur and Bengaluru ACF PACF Plot

Time series decomposition of Jodhpur and Bengaluru component shown in figure 7. Decomposition help to adjust the seasonality, to detect the trend, detrend and verify the remaining white/random noise from the data. from these graphs clearly, understand the components of wind speed data of both the cities. overall observation of both the cities approximately same although Bengaluru is more fluctuates as compared to Jodhpur. In terms of trend Bengaluru component shows upward trends and in Jodhpur have some variations in trends. In terms of seasonal components, both the cities show yearly seasonality. After carefully considering the seasonality, observations randomness and trend found that data is stationary and suitable for forecasting.

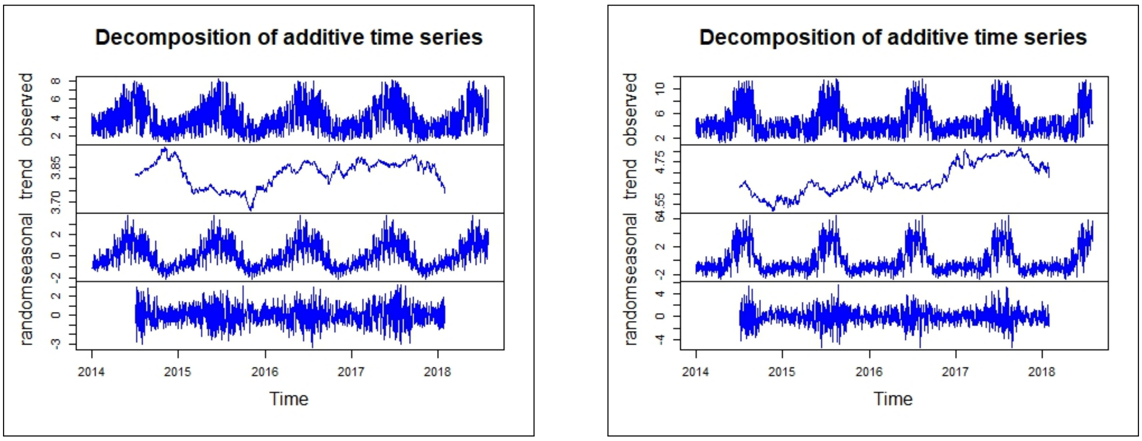

Figure 7: Jodhpur and Bengaluru Decomposition Plot

5.3 Evaluation of ARIMA model

ARIMA model forecasting evaluation of next 8 months of Jodhpur and Bengaluru are shown in figure 8. The model shows the maximum likelihood estimation. Blue lines shaded area shows the exact forecasted value of next 240 days. The model evaluated ARIMA(0,0,0)(0,1,0)[365] for both the cities, where (0,1,0) means one seasonal differencing with 365 days on daily observations. After that check the residuals of resulted forecasting values.

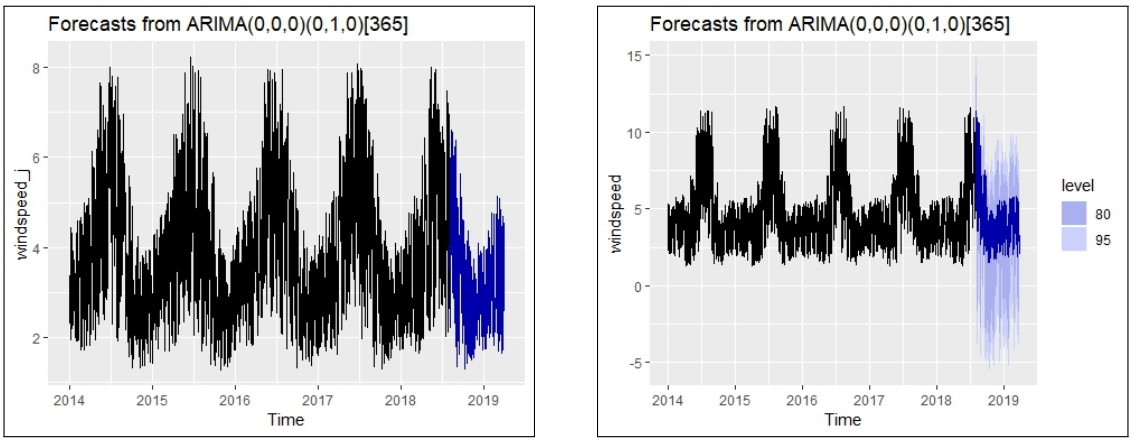

Figure 8: ARIMA Model forecasting:Jodhpur and Bengaluru

5.4 Evaluation of SES Model

Simple Exponential Smoothing (SES) model forecasting evaluation of next 240 days of both the cities are shown in figure 9. SES model based on most recent observation and least square estimation. The straight-line presented a forecast of estimated mean values. The 80% probability (in dark blue shaded area) associated with the interval and 90% probability (in light blue shaded area) associated with in this interval. From the forecast value of this model will not provide the value of every day so as per the research objective forecasting wind electricity generation, this model is not good.

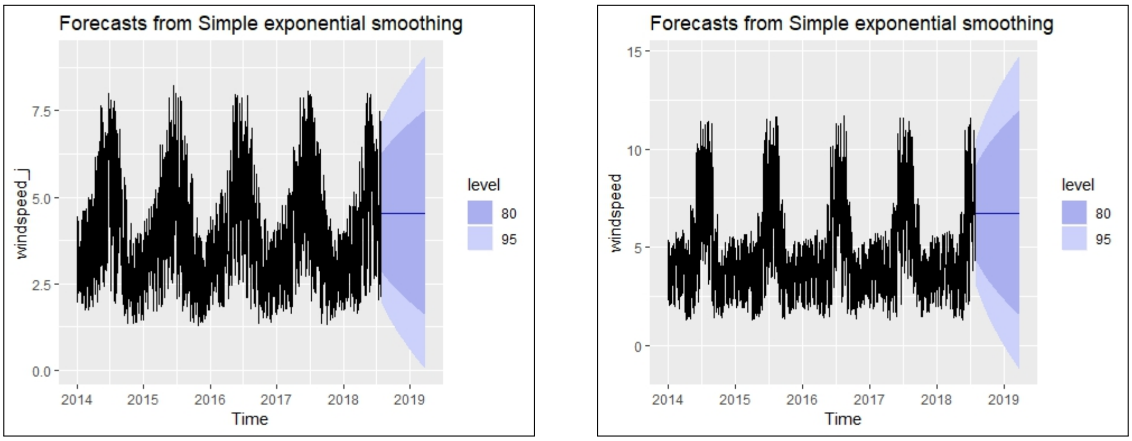

Figure 9: SES Model forecasting:Jodhpur and Bengaluru

5.5 Evaluation of HOLT's Model

HOLTs model forecasting evaluation of next 240 days (8 Months) of both the cities are shown in figure10. HOLTs model is the expansion of SES model and worked for linear trend. The model produce forecast with trend continue with the same slope indefinite in the future. The 80% confidence shows forecast values will be in that interval and 95% confidence shows forecast values of wind speed will be in that confidence intervals.

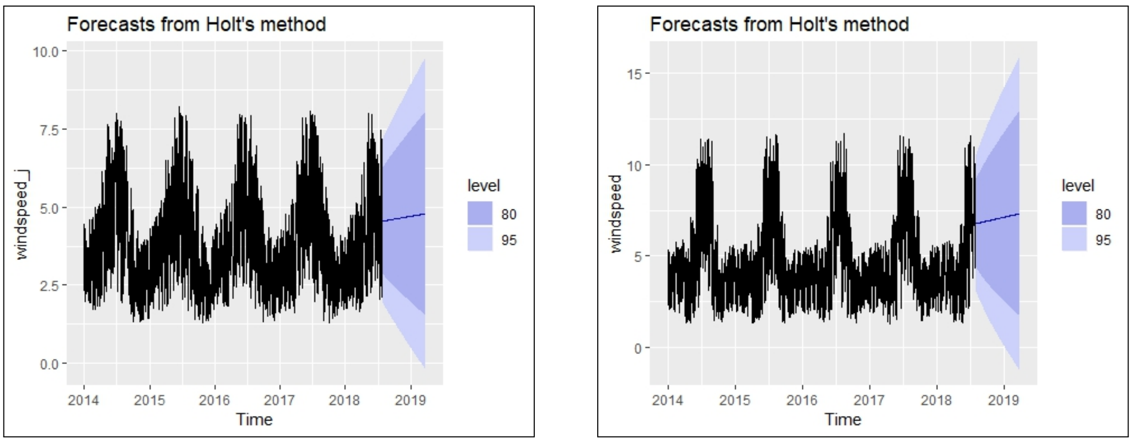

Figure 10: HOLT's Model forecasting:Jodhpur and Bengaluru

5.6 Evaluation of TBATS model

TBATS model forecasting evaluation of next 240 days of Jodhpur and Bengaluru are shown in figure 11. the model covered seasonality, heterogeneity, short term dynamics, multiple seasonality and automated. The Blue line in the graph shows the forecasted value of the model. The 80% probability of forecasting value (in dark blue shaded area) associated with this interval and 95% probability of forecasting value (in light blue shaded area) associated with in this interval For Jodhpur city, TBATS (0.036, 0,0, -, ¡365.25,3¿), first part is boxcox transformation is 0.36 close to square root, next part is arma where p, q is 0 means white noise error used, next is damping parameter so,- means no damping trends, last part is Fourier terms seasonal period 365.25 yearly and 3 Fourier terms are selected. For Bengaluru City, TBATS(0.132,0,0,0.8,-). boxcox transformation is 0.132, next part is arma where p, q is 0 means white noise error used, next is damping parameter 0.8.the graphical representation is shown below.

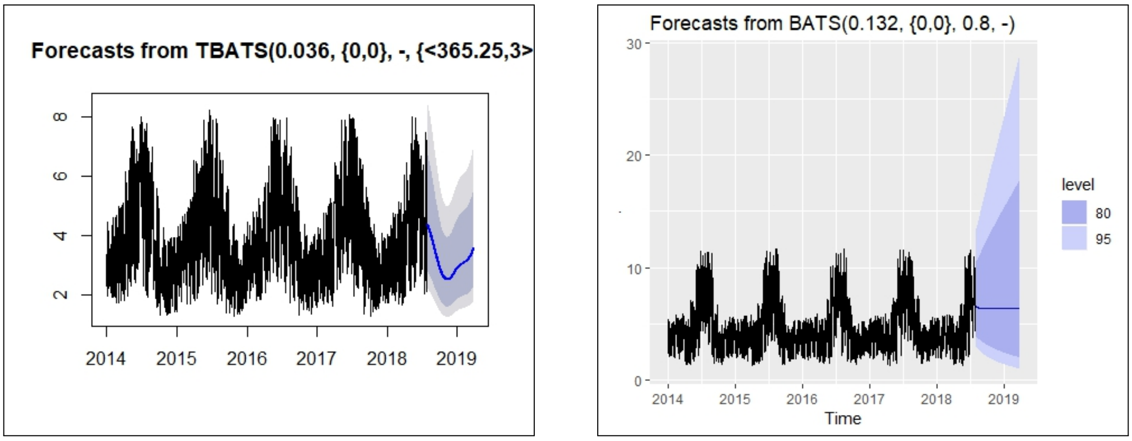

Figure 11: TBATS Model forecasting:Jodhpur and Bengaluru

5.7 Conclusion

Based on developed and evaluated ARIMA, TBATS, SES and HOLT's models presented in chapters 4,5 using that able to answer the Sub_objectives(i) to (v) in figure 1.following section about the accuracy and final wind electricity generation part.

6 Results

6.1 Introduction

This Section is about the choose the model on performance and further discussed the wind electricity generation Comparison.

6.2 Performance Evaluation of Forecasted Models

The purpose of the wind speed forecasting for wind electricity generation evaluation and forecasting have been accomplished. Now all applied models have been evaluated using accuracy measurement: Root mean square error (RMSE), Mean absolute error (MAE), Mean absolute percentage error (MAPE), Mean absolute scaled error (MASE).

Model	RMSE	MAE	MAPE	MASE
ARIMA	2.100819	1.688017	0.4838731	1.076602
TBATS	1.413076	1.193134	0.3423096	0.7609699
SES	2.029466	1.559754	0.3361735	0.9947968
HOLT's	2.105108	1.613743	0.3430939	1.02923

Figure 12: Models Performance

From the model performance table13 clearly understand that TBATS model outperformed with the other models. ARIMA model perform better than Simple Exponential Smoothing model (SES) model. SES performed better than HOLTs model. The HOLT's model error rate maximum and TBATS model error rate in terms of RMSE, MAE, MAPE and MAE values are least with other models. Hence, TBATS model outperformed and best Technique for modelling this type of data.

6.3 Wind Electricity Generation Comparison

The aim of the project is forecasting wind electricity generation and compared the result of two cities. First all the forecasted value of wind speed of next eight months forecasted by time series models for Jodhpur and Bengaluru. The wind speed parameter is selected to calculate the power forecasting of wind with the help of wind power forecasting formula. The table 2 shows the total electricity forecasted in the months for both cities. From the graphical presentation clearly comprehend that Jodhpur city more produced electricity in every month as compared to Bengaluru city. Hence, we can say that investing money to set up the wind power on same amount in both the cities, the electricity generates more in Jodhpur. Thus, in terms of set up wind power plant, Jodhpur is the better place for forecasted high amount electricity generation. So, in this way able to answer the research objective 5.

Table 2: Comparison:Wind Electricity Generation Forecast

Month	Jodhpur (KW)	Bengaluru (KW)
August	1675846	1202671.507
September	1653430	1051511.7
October	1653402	900198.7463
November	1362909	702762.4051
December	994270	600605.184
January	902020	600605.184
February	838553.3	375747.0225
March	1475846	1002671.507

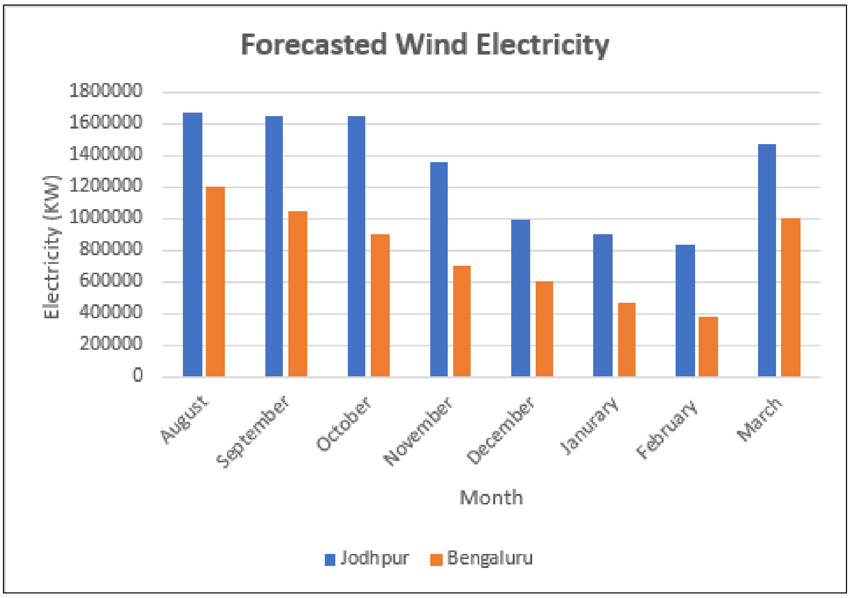

Figure 13: Forecast Wind Electricity generation:Jodhpur and Bengaluru

6.4 Conclusion

Compared the developed models in terms of RMSE, MAE, MASE, MAPE and resolved the objective 3. After that Choose the best model on the basis of error rates and solved and using that model able to compute the forecast electricity production and resolved objective 4. After that compared the electricity results for both cities. Jodhpur location better than Bengaluru and we can give the suggestion Indian government for Jodhpur city to set up the wind mill refer in this way solved the objective 5, refer figure1

7 Conclusion and Future Work

This project as resulted in wind electricity generation forecasting models which will be significant for both Jodhpur and Bengaluru city in India. so this will resolve in helping to solve energy crisis which is currently effected to community in this city. In addition project is resulted in contribution to the knowledge and also in to data analysis Industry. The project aim is to the forecasting wind electricity generation compared with two cities which have been successfully achieved by the time series models. Accurate forecast can significantly help in sinking the cost and enlightening the constancy of wind electricity. From the research able to forecast the next 8 months values of wind speed from August 2018 to March 2019. In terms of performance TBATS model performance outperformed with all others models. So, this is the first research to explore and applied TBATS model in wind energy forecasting. Although result of ARIMA model also performed good. SES and HOLTs model is not good option for forecasting wind speed.

While worked on this thesis the some question come in my mind for future work in this field. future study mainly focuses on the forecast the electricity with other renewable sources such as solar, tidal and water. This study will use deep learning and ARIMAX-ANN, fuzzy models on same dataset and then will check the performance. Data of some wind park and wind farm and data of past 10 years need for accurate wind speed results.

To decrease the white noise effect properly will used some other models. There are tons of villages in India where still lack of electricity so will worked on these cities to fulfill the electricity need.

Acknowledgement

I would like to show gratitude to my supervisor Dr. Catherine Mulwa who encourage and help me in this thesis. she has guide me and clear the doubts on time.
Furthermore, I also owe special thanks to my parents Mr. and Mrs. Gupta for their inspiration and emotional Support.Thank you very much!

References

Abdullah, A., Saleh, A., Moustafa, M. and Abo-al Ez, K. (2014). A proposed framework for a forecasting system of wind energy power generation, *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)* **3**(4).

As' ad, M. (2012). Finding the best arima model to forecast daily peak electricity demand.

Bakshi, R. (2002). Wind energy in india, *IEEE Power Engineering Review* **22**(9): 16–18.

Box, G. E., Jenkins, G. M., Reinsel, G. C. and Ljung, G. M. (2015). *Time series analysis: forecasting and control*, John Wiley & Sons.

BROŻYNA, J., Mentel, G., Szetela, B. and Strielkowski, W. (2018). Multi-seasonality in the tbats model using demand for electric energy as a case study., *Economic Computation & Economic Cybernetics Studies & Research* **52**(1).

Burton, T., Jenkins, N., Sharpe, D. and Bossanyi, E. (2011). *Wind energy handbook*, John Wiley & Sons.

Chen, J., Zeng, G.-Q., Zhou, W., Du, W. and Lu, K.-D. (2018). Wind speed forecasting using nonlinear-learning ensemble of deep learning time series prediction and extremal optimization, *Energy Conversion and Management* **165**: 681–695.

Das, J. and Banerjee, R. (2017). Estimation of long range correlations and farima modelling of wind speed in maharashtra, *Asia-Pacific Power and Energy Engineering Conference (APPEEC), 2017 IEEE PES*, IEEE, pp. 1–6.

Daut, I., Razliana, A., Irwan, Y. and Farhana, Z. (2012). A study on the wind as renewable energy in perlis, northern malaysia, *Energy Procedia* **18**: 1428–1433.

De Livera, A. M., Hyndman, R. J. and Snyder, R. D. (2011). Forecasting time series with complex seasonal patterns using exponential smoothing, *Journal of the American Statistical Association* **106**(496): 1513–1527.

do Nascimento Camelo, H., Sérgio Lucio, P., Verçosa Leal Junior, J. B., von Glehn dos Santos, D. and Cesar Marques de Carvalho, P. (2018). Innovative hybrid modeling of wind speed prediction involving time-series models and artificial neural networks, *Atmosphere* **9**(2): 77.

Eldali, F. A., Hansen, T. M., Suryanarayanan, S. and Chong, E. K. (2016). Employing arima models to improve wind power forecasts: A case study in ercot, *North American Power Symposium (NAPS), 2016*, IEEE, pp. 1–6.

Erdem, E. and Shi, J. (2011). Arma based approaches for forecasting the tuple of wind speed and direction, *Applied Energy* **88**(4): 1405–1414.

Han, J., Pei, J. and Kamber, M. (2011). *Data mining: concepts and techniques*, Elsevier.

Hassani, H., Silva, E. S., Gupta, R. and Segnon, M. K. (2015). Forecasting the price of gold, *Applied Economics* **47**(39): 4141–4152.

Hyndman, R. J. and Koehler, A. B. (2006). Another look at measures of forecast accuracy, *International journal of forecasting* **22**(4): 679–688.

Kale, R. V. and Pohekar, S. D. (2012). Electricity demand supply analysis: current status and future prospects for maharashtra, india, *Renewable and Sustainable Energy Reviews* **16**(6): 3960–3966.

Kalekar, P. S. (2004). Time series forecasting using holt-winters exponential smoothing, *Kanwal Rekhi School of Information Technology* **4329008**: 1–13.

Khosravi, A., Machado, L. and Nunes, R. (2018). Time-series prediction of wind speed using machine learning algorithms: A case study osorio wind farm, brazil, *Applied Energy* **224**: 550–566.

Kočenda, E. and Černỳ, A. (2015). *Elements of time series econometrics: An applied approach*, Charles University in Prague, Karolinum Press.

Kumar, G. and Malik, H. (n.d.). Generalized regression neural network based wind speed prediction model for western region of india, *Procedia Computer Science* **93**: 26–32.

Kwon, S.-D. (2010). Uncertainty analysis of wind energy potential assessment, *Applied Energy* **87**(3): 856–865.

Lau, A. and McSharry, P. (2010). Approaches for multi-step density forecasts with application to aggregated wind power, *The Annals of Applied Statistics* pp. 1311–1341.

Liu, H., Tian, H.-q. and Li, Y.-f. (2012). Comparison of two new arima-ann and arima-kalman hybrid methods for wind speed prediction, *Applied Energy* **98**: 415–424.

Liu, H., Tian, H.-q. and Li, Y.-f. (2015). Four wind speed multi-step forecasting models using extreme learning machines and signal decomposing algorithms, *Energy conversion and management* **100**: 16–22.

Mabel, M. C. and Fernandez, E. (2008). Analysis of wind power generation and prediction using ann: A case study, *Renewable energy* **33**(5): 986–992.

Mahalakshmi, G., Sridevi, S. and Rajaram, S. (2016). A survey on forecasting of time series data, *Computing Technologies and Intelligent Data Engineering (ICCTIDE), International Conference on*, IEEE, pp. 1–8.

Meenal, R., Selvakumar, A. I., Brighta, K., Joice, S. C. J. and Richerd, C. (2018). Solar radiation resource assessment using weka, *2018 2nd International Conference on Inventive Systems and Control (ICISC)*, IEEE.

Mentis, D., Siyal, S. H., Korkovelos, A. and Howells, M. (2016). A geospatial assessment of the techno-economic wind power potential in india using geographical restrictions, *Renewable energy* **97**: 77–88.

Naim, I., Mahara, T. and Idrisi, A. R. (2018). Effective short-term forecasting for daily time series with complex seasonal patterns, *Procedia Computer Science* **132**: 1832–1841.

Pinto, T., Ramos, S., Sousa, T. M. and Vale, Z. (2014). Short-term wind speed forecasting using support vector machines, *Computational Intelligence in Dynamic and Uncertain Environments (CIDUE), 2014 IEEE Symposium on*, IEEE, pp. 40–46.

Rallapalli, S. R. and Ghosh, S. (2012). Forecasting monthly peak demand of electricity in indiaa critique, *Energy policy* **45**: 516–520.

Ramasamy, P., Chandel, S. and Yadav, A. K. (2015). Wind speed prediction in the mountainous region of india using an artificial neural network model, *Renewable Energy* **80**: 338–347.

Sarkar, A. and Behera, D. K. (2012). Wind turbine blade efficiency and power calculation with electrical analogy, *International Journal of Scientific and Research Publications* **2**(2): 1.

Shumway, R. H. and Stoffer, D. S. (2011). Time series regression and exploratory data analysis, *Time series analysis and its applications*, Springer, pp. 47–82.

Singh, A., Gurtej, K., Jain, G., Nayyar, F. and Tripathi, M. (2016). Short term wind speed and power forecasting in indian and uk wind power farms, *Power India International Conference (PIICON), 2016 IEEE 7th*, IEEE, pp. 1–5.

Sreelakshmi, K. and Kumar, P. R. (2008). Performance evaluation of short term wind speed prediction techniques, *IJCSNS International Journal of Computer Science and Network Security*, Citeseer.

Taylor, J. W., McSharry, P. E., Buizza, R. et al. (2009). Wind power density forecasting using ensemble predictions and time series models, *IEEE Transactions on Energy Conversion* **24**(3): 775.

Zhang, J., Wei, Y., Tan, Z.-f., Ke, W. and Tian, W. (2017). A hybrid method for short-term wind speed forecasting, *Sustainability* **9**(4): 596.