REPUBLIC OF TÜRKİYE
KIRŞEHİR AHİ EVRAN UNIVERSITY
INSTITUTE OF NATURAL AND APPLIED SCIENCES
DEPARTMENT OF MECHANICAL ENGINEERING

WIND ENERGY FORECASTING METHODS: A CASE STUDY OF THE LONG SHORT-TERM MEMORY MODEL (LSTM)

ALI ABDULRAHMAN HUSSEIN SALIHI

MSc THESIS

KIRŞEHİR
2024

REPUBLIC OF TÜRKİYE
KIRŞEHİR AHİ EVRAN UNIVERSITY
INSTITUTE OF NATURAL AND APPLIED SCIENCES
DEPARTMENT OF MECHANICAL ENGINEERING

WIND ENERGY FORECASTING METHODS: A CASE STUDY OF THE LONG SHORT-TERM MEMORY MODEL (LSTM)

ALI ABDULRAHMAN HUSSEIN SALIHI

MSc THESIS

SUPERVISOR
ASST. PROF. DR. Merdin DANIŞMAZ

KIRŞEHİR
2024

KIRŞEHİR AHİ EVRAN UNIVERSITY
INSTITUTE OF NATURAL AND APPLIED SCIENCES
MSc THESIS
ETHICS DECLARATION

In this thesis study, which I have read and understood the Kırşehir Ahi Evran University Scientific Research and Publication Ethics Directive and which I have prepared in accordance with the Kırşehir Ahi Evran University Institute of Natural and Applied Science Thesis Writing Rules;

- I have obtained the data, information and documents I have presented in the thesis within the framework of academic and ethical rules,
- I present all information, documents, evaluations and results in accordance with scientific ethical rules,
- I have cited all the works I have benefited from in the thesis by making appropriate references,
- I have not made any changes in the data used and the results,
- This study, which I have presented as a thesis, is original,

Otherwise, I declare that I accept all legal actions to be taken against me in this regard and all loss of rights that may arise against me. ...../……./20…

Ali Abdulrahman Hussein SALIHI

LIST OF CONTENTS

Page No

LIST OF CONTENTS ..... **I**
ACKNOWLEDGMENTS ..... **III**
GENİŞLETİLMİŞ ÖZET ..... **IV**
ABSTRACT ..... **VI**
LIST OF TABLES ..... **VIII**
LIST OF FIGURES ..... **IX**
LIST OF ICONS AND ABBREVIATIONS ..... **XI**
1. INTRODUCTION **1**
1.1. Wind Power 4
1.1.1. Wind power potential assessment process and its stages 4
1.1.2. Wind distribution 7
1.2. Advantages and Disadvantage of Wind Turbine 8
1.2.1. Usage of Generators with Wind turbine 10
1.3. Numerical Weather Prediction & Wind Forecasting 11
1.3.1. The basics of wind power forecasting 13
1.3.2. Importance of Short-term forecasting 17
1.4. Neural Networks 18
1.4.1. Types of RNNs 18
1.4.2. Long short-term memory 19
1.5. LSTM RNN (Recurrent Neural Networks)-Based Forecasting 20
1.5.1. Sequence-to-sequence LSTM RNN 21
1.5.2. LSTM network layer 22
1.6. Summarize of the Introduction 23
1.7. Problem Statement 24
1.8. Objectives of the Study 24
1.9. Significance of the Study 25
2. LITERATURE REVIEW **27**
3. MATERIAL AND METHOD **49**
3.1. Dataset Description 49
3.2. Introducing the Data to the Paython Program 51
3.2.1. Application of exploratory data analysis using pandas profiling and then some boxplots 51

3.3. Data Overview 52
3.3.1. Variables 52
3.3.2. Missing values 67
3.4. Model Description 68
3.4.1. SARIMAX Model 68
3.4.2. XG Boost 72
3.4.3. Random forest regressor 75
3.4.4. Long-Short-Term-Memory (LSTM) model 77
4. RESULTS AND DISCUSSIONS 79
4.1. Basic Arima Model 79
4.2. Pattern of Power Generation Versus Wind Speed 82
4.2.1. Seasonal ARIMA (SARIMA) model 85
4.2.2. Whole dataset on same plane 85
4.2.3. Graph of predicted versus actual for last 15 days of the dataset 86
4.2.4. Graph using 80% of the dataset for training 86
4.3. Extreme Gradient Boost (XGBOOST) 87
4.3.1. Long Short-Term Memory (LSTM) 88
4.4. Evaluation of The Performance of The Forecasting Models 91
5. CONCLUSION AND RECOMMENDATION 95
5.1. Initial Data Examination and Processing 95
5.2. Data Overview and Preprocessing 95
5.3. Methodology and Modeling Comparison 95
5.4. Recommendation 95
6. REFERENCE 97
CURRICULUM VITAE 105

ACKNOWLEDGMENTS

For their continual encouragement and counsel during my master's programme, Asst. Prof. Dr. Merdin DANIŞMAZ, my supervisors, have my sincere gratitude. Their knowledge and tolerance have been a great help to me and were essential to the achievement of this thesis.

I want to convey my most profound appreciation to the Mechanical Engineering Department in Engineering and Architecture Faculty at Kırşehir Ahi Evran University for giving me the chance to pursue my master's degree. Their support and assistance throughout this research journey have been invaluable.

I also extend my gratitude to my best friend Mohammed Oral ALHURMUZI for all the information and assistance he gave me throughout my research, not to mention his technical aid.

I appreciate my friends and family's affection and assistance throughout this journey. I would not have finished my adventure if it were not for their support and inspiration.

Last, I thank everyone who participated in my study and was willing to share their knowledge. With their help, this work was completed.

January, 2024 Ali Abdulrahman Hussein SALIHI

GENİŞLETİLMİŞ ÖZET

YÜKSEK LİSANS TEZİ

RÜZGÂR ENERJİSİ TAHMİNİ YÖNTEMLERİ: UZUN KISA SÜRELİ BELLEK MODELİ (LSTM) ÖRNEĞİ

Ali Abdulrahman Hussein SALIHI

KIRŞEHİR AHİ EVRAN ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
MAKİNE MÜHENDİSLİĞİ ANABİLİM DALI

Danışman: **Dr. Öğr. Üyesi Merdin DANIŞMAZ**
Yıl: 2024 Sayfa: 105
Jüri: **Dr. Öğr. Üyesi Merdin DANIŞMAZ**
Prof. Dr. Ali Osman KURBAN
Dr. Öğr. Üyesi Omer Adil ZAİNAL

Bu tez, rüzgâr enerjisi tahmininin alanını ilerletmeyi ve sürdürülebilir enerji yönetimini teşvik etmeyi amaçlayarak üç ayrı araştırma bileşeninden elde edilen bulguları sentezlemektedir. İlk çalışma, Long Short-Term Memory (LSTM) ve diğer metodolojileri kullanarak rüzgâr enerjisi tahminini araştırır. Araştırma, rüzgâr hızı verilerine dayalı güç üretimini öngörme odaklıdır ve eksik değerler ve mevsimsel desenlere yönelik zorlukları ele alır. İlk analizlerden elde edilen sonuçlar, ARIMA modelleri ve rüzgâr hızı ile güç üretimi arasındaki korelasyon değerlendirmelerini içerir ve özellikle Temmuz, Ağustos ve Eylül aylarında güç üretiminde belirgin zirvelerin olduğunu, bunun rüzgâr hızı dalgalanmaları ile uyumlu olduğunu ortaya koyar. Çalışma, 2.5 metrenin üzerindeki rüzgâr hızlarının güç üretimini başlattığını, 8 m/s civarında zirve yaptığını ve grafiksel temsillerin rüzgâr hızı ile güç üretimi arasında bir sigmoid ilişkiyi gösterdiğini belirledi. Ardından, SARIMA modelinin başarısızlığının ardından alternatif modelleme yaklaşımları keşfedildi. XG Boost, Random Forest Regressor ve LSTM, görselleştirme ve istatistiksel analiz yoluyla veri setinin özelliklerinin detaylı bir incelemesi ile birlikte değerlendirildi. Eksik hücrelerin yaygınlığı, titiz veri işleme öneminin vurgulanmasına neden oldu. Veri genel bakışı ve ön işleme aşaması, veri ithalatı sürecini, tarih sütununun tanınmasını, yinelenen girişlerin işlenmesini ve Pandas profillemesi ile boxplot keşfini detaylandırdı. "Active Power" ve "Ambient Temperature" gibi temel değişkenler ele alındı, eksik değerlerin zorluğu ve gereksiz değişkenlerin tanımlanmasıyla ilgili olarak vurgulandı. Son aşama, kullanılan yöntemi kapsayan, doğru analiz için eksik veri noktalarına ve anormalliklere odaklanan bir yöntemdir. Titiz temizlik süreci, model seçimi (SARIMA, XG Boost, Random Forest Regressor, LSTM) ve bunların performansı tartışıldı. Ayrıca, veri doğruluğunun önemi, rüzgâr hızının güç üretimine etkisi ve rüzgâr enerjisi dinamiklerini etkili bir şekilde yakalamak için çeşitli modelleme yöntemlerinin gerekliliği vurgulandı. Bu bulgulara dayanarak, rüzgâr enerjisi tahminini ve sürdürülebilir yönetimi ilerletmeye yönelik bir dizi öneri getirildi. Veri seti kalitesini ve güvenilirliğini artırmak için eksik değerlerin, aykırı değerlerin ve gürültünün işlenmesini içeren gelişmiş veri ön işleme yöntemleri önerildi. Daha doğru tahminler için klasik istatistik metodolojilerini ve makine öğrenimi algoritmalarını birleştiren hibrit modelleme teknolojileri önerildi. Rüzgâr enerjisi tahminine etki eden meteorolojik ve coğrafi unsurların özellik mühendisliği metodolojilerine eklenmesi, güç üretimini daha iyi anlamak için önerildi. İlgili değişkenlerle rüzgâr enerjisi üretimi arasındaki ilişkiyi anlamak için daha yorumlanabilir modeller geliştirmek, bilinçli kararlar için vurgulandı. Model ortalaması ve yığma gibi ensemble

öğrenme yöntemleri, model kusurlarını en aza indirerek tahmin doğruluğunu artırmak amacıyla önerildi. Dinamik hava durumu desenlerini ve çevresel koşulları yakalamak için gerçek zamanlı veri akışları ve gelişmiş izleme sistemlerinin kullanımı, uyarlanabilir tahmin modelleri için teşvik edildi. Tahmin modelinin parametre ayarlarıile ilgili hassasiyet çalışması, rüzgâr enerjisi üretimini etkileyen en ilgili değişkenleri belirlemek amacıyla önerildi. Farklı coğrafi konumlar ve çevresel koşullar arasında tahmin modellerinin güvenilirliği ve genelleme yeteneğinin sağlanması, kapsamlı geriye dönük test ve çeşitli veri setlerinde doğrulama dahil olmak üzere, sıkı model doğrulama ve doğrulama ile vurgulandı. Değişen iklim dinamikleri ve küresel enerji talepleri karşısında sürdürülebilir enerji altyapısı planlamak için uzun vadeli rüzgâr enerjisi üretimi tahmin çalışmaları önerildi.

Son olarak, akademik kurumlar, endüstri paydaşları ve devlet kurumlarının iş birliği yaparak dünya genelinde rüzgâr enerjisi tahmin teknolojileri ve sürdürülebilir enerji uygulamaları için bilgi, veri ve yenilikçi çözümleri paylaşmaları teşvik edildi. Bu kapsamlı yaklaşım, rüzgâr enerjisi tahmininin ilerlemesine katkıda bulunmayı ve sürdürülebilir enerji yönetimi uygulamalarını teşvik etmeyi amaçlamaktadır.

Anahtar Kelimeler: Rüzgâr Enerjisi Tahmini, LSTM Modelleme, Güç Üretimi Tahmini, Zaman Serisi Analizi, Yenilenebilir Enerji İçin Makine Öğrenimi

ABSTRACT

MASTER'S THESIS

WIND ENERGY FORECASTING METHODS: A CASE STUDY OF THE LONG SHORT-TERM MEMORY MODEL (LSTM)

Ali Abdulrahman Hussein SALIHI

KIRŞEHİR AHİ EVRAN UNIVERSITY
INSTITUTE OF NATURAL AND APPLIED SCIENCES
DEPARTMENT OF MECHANICAL ENGINEERING

Supervisor:	**Asst. Prof. Dr. Merdin DANIŞMAZ**
	Year: 2024 Pages: 105
Juries:	**Asst. Prof. Dr. Merdin DANIŞMAZ**
	Prof. Dr. Ali Osman KURBAN
	Asst. Prof. Dr. Omer Adil ZAİNAL

This thesis synthesizes findings from three distinct research components, aiming to advance the field of wind energy prediction and promote sustainable energy management. The initial study explores wind energy prediction utilizing Long Short-Term Memory (LSTM) and other methodologies. The investigation focuses on forecasting power output based on wind speed data, addressing challenges related to missing values and seasonal patterns. Results from initial analyses, including ARIMA models and correlation assessments between wind speed and power output, revealed distinct peaks in power output during specific months, notably July, August, and September, corresponding with wind speed fluctuations. The study identified that wind speeds above 2.5 meters per second initiate power generation, peaking around 8 m/s, with graphical representations indicating a sigmoid relationship between wind speed and power output. Subsequently, alternative modeling approaches were explored after the failure of the SARIMA model. XG Boost, Random Forest Regressor, and LSTM were considered, with a detailed examination of the dataset's properties through visualization and statistical analysis. The prevalence of missing cells underscored the importance of meticulous data handling. The data overview and preprocessing phase detailed the process of data importation, recognition of the date column, handling of duplicate entries, and exploration through Pandas profiling and boxplots. Key variables such as "Active Power" and "Ambient Temperature" were discussed, along with the challenge of missing values and the identification of redundant variables. The final phase encapsulated the methodology used, emphasizing the importance of addressing missing data points and anomalies for accurate analysis. The rigorous cleaning process, model selection (SARIMA, XG Boost, Random Forest Regressor, LSTM), and their respective performance were discussed. Furthermore, the significance of data accuracy, the impact of wind speed on power output, and the necessity for varied modeling methods to capture wind energy dynamics effectively were highlighted. Building on these findings, several recommendations for advancing wind energy prediction and sustainable management were proposed. Advanced data pre-processing methods were suggested to enhance dataset quality and dependability, including the handling of missing values, outliers, and noise. Hybrid modeling technologies that combine classical statistical methodologies and machine learning algorithms were recommended for more accurate predictions. Incorporating meteorological and geographical elements into feature engineering methodologies was suggested to better understand power output. Developing more interpretable models to comprehend the relationship between relevant variables and wind energy generation was emphasized for informed decision-making.

Ensemble learning methods, such as model averaging and stacking, were proposed to increase prediction accuracy by minimizing model flaws. The utilization of real-time data streams and advanced monitoring systems for dynamic weather patterns and environmental conditions was encouraged for adaptive forecasting models. A rigorous sensitivity study was suggested to assess forecasting model robustness to parameter adjustments, identifying the most relevant variables affecting wind energy generation. Ensuring the reliability and generalizability of forecasting models across different geographical locations and environmental conditions was emphasized through rigorous model validation and verification. Long-term wind energy generation forecasting studies were proposed to plan sustainable energy infrastructure in the face of changing climate dynamics and global energy demands. Finally, collaboration between academic institutions, industry stakeholders, and government agencies was encouraged to share knowledge, data, and innovative solutions for wind energy forecasting technologies and sustainable energy practices worldwide. This comprehensive approach aims to contribute to the advancement of wind energy prediction and foster sustainable energy management practices.

Keywords: Wind Energy Prediction, LSTM Modeling, Power Output Forecasting, Time Series Analysis, Machine Learning for Renewable Energy

LIST OF TABLES

Page No

Table 4.1. Model performance comparison for time series forecasting. **92**

LIST OF FIGURES

Page No

Figure 1.1. The first windmill built by Hammurabi. **1**

Figure 1.2. Oldest design drawings of the post-mills by Mariano Jacob (Pilipets et al., 2014). **2**

Figure 1.3. Fixed-speed wind urbine with induction generator (Drago Ban et al., 2021)..... **10**

Figure 1.4. Frame of short-term wind power forecasting..... **13**

Figure 1.5. Frame of wind power forecasting. **17**

Figure 3.1. Dataset statistics..... **52**

Figure 3.2. df_index **53**

Figure 3.3. Active power..... **53**

Figure 3.4. Ambient temperature **54**

Figure 3.5. Bearing shaft temperature **55**

Figure 3.6. Blade1pitch angle..... **56**

Figure 3.7. Blade 2 pitch angle..... **57**

Figure 3.8. Blade 3 pitch angle..... **58**

Figure 3.9. Gear box bearing temperature..... **58**

Figure 3.10. Gear box oil temperature **59**

Figure 3.11. Generator RPM **60**

Figure 3.12. Generator winding 1 temperature **61**

Figure 3.13. Generator winding 2 temperature **62**

Figure 3.14. Hub temperature..... **62**

Figure 3.15. Main box temperature **63**

Figure 3.16. Nacelle position **64**

Figure 3.17. Reactive power..... **64**

Figure 3.18. Rotor RPM **65**

Figure 3.19. Turbine status..... **66**

Figure 3.20. Wind direction **66**

Figure 3.21. Wind speed..... **67**

Figure 3.22. Data if serval year of turkey..... **68**

Figure 3.23. The python codes for SARIMAX Model..... **71**

Figure 3.24. The python codes for XGBOOTS..... **74**

Figure 3.25. The python codes for Random forest regressor **76**

Figure 3.26. The python codes for LSTM..... **78**

Figure 4.1. Power generated **79**
Figure 4.2. Mean daily power generated **79**
Figure 4.3. Wind speed **80**
Figure 4.4. Active power and wind speed *100 ..... **81**
Figure 4.5. Power output and wind velocity *100 ..... **81**
Figure 4.6. Power output versus wind speed. **82**
Figure 4.7. Power generated versus wind velocity. **82**
Figure 4.8. Power output versus wind speed **83**
Figure 4.9. Monthly boxplots of power generated **84**
Figure 4.10. Monthly boxplots of wind Speed **84**
Figure 4.11. Hourly boxplots of wind speed. **84**
Figure 4.12. Active power **85**
Figure 4.13. Graph of predicted versus actual for last 15 days of the dataset. **86**
Figure 4.14. Graph using 80% of the dataset for training. **87**
Figure 4.15. Extreme gradient boost. **88**
Figure 4.16. Long Short-Term Memory (LSTM). **89**
Figure 4.17. LSTM modelling **89**
Figure 4.18. Performance of a predictive model **90**

LIST OF ICONS AND ABBREVIATIONS

Icons		**Described**
%	:	Percentage

Abbreviations		**Described**
ANN	:	Artificial neural network
BE	:	Back engendering
CNN	:	Convolutional neural network
DDM	:	Discrete dark model
DFIG	:	Doubly fed ınduction generator
DNS	:	Direct numerical simulation
LSHVM	:	Least square help vector machine
LSTM	:	Long short-term memory
NWP	:	Numerical weather prediction
PLM	:	Profound learning models
PMSG	:	Permanent magnet synchronous generator
RBF	:	Radial basis function
RNN	:	Recurrent neural network
RSC	:	Rotor side converter
SQIG	:	Squirrel cage ınduction generator
WBO	:	Wavelet brain organization
WECS	:	Wind energy conversion systems
WRIG	:	Wound rotor induction generator
WT	:	Wind turbines
WTG	:	Wind turbine generator

1. INTRODUCTION

Wind energy plays a pivotal role in addressing the global energy crisis by offering a sustainable and renewable source of power. Harnessing the power of the wind helps reduce reliance on finite fossil fuels, mitigating environmental impact and combating climate change. Additionally, wind energy promotes energy independence, fostering resilience in the face of geopolitical uncertainties. As a clean and abundant resource, it contributes to a more sustainable future, ensuring a greener and healthier planet for generations to come.

Even before 7000 years ago, wind provided the power for Egyptian sails. In the seventeenth century BC, the Babylonian ruler Hammurabi purportedly intended to deploy windmills for irrigation. In the 400 BC book Arthashastra, the Indian philosopher Kautilya mentioned windmills. According to some Indian experts, Buddhist monks brought the art of windmills to China, and there is evidence that windmills were effectively used for water pumping throughout the pre-Christian era. Windmills also migrated east as trade along the Silk Road from China to the Middle East grew. In the third century BC, Hero of Alexandria wrote about a windmill with a horizontal axis that could be used to power an instrument. As early as 200 BC, vertical axis windmills were being used in Persia and the Middle East to grind grains (Abdoos, 2016).

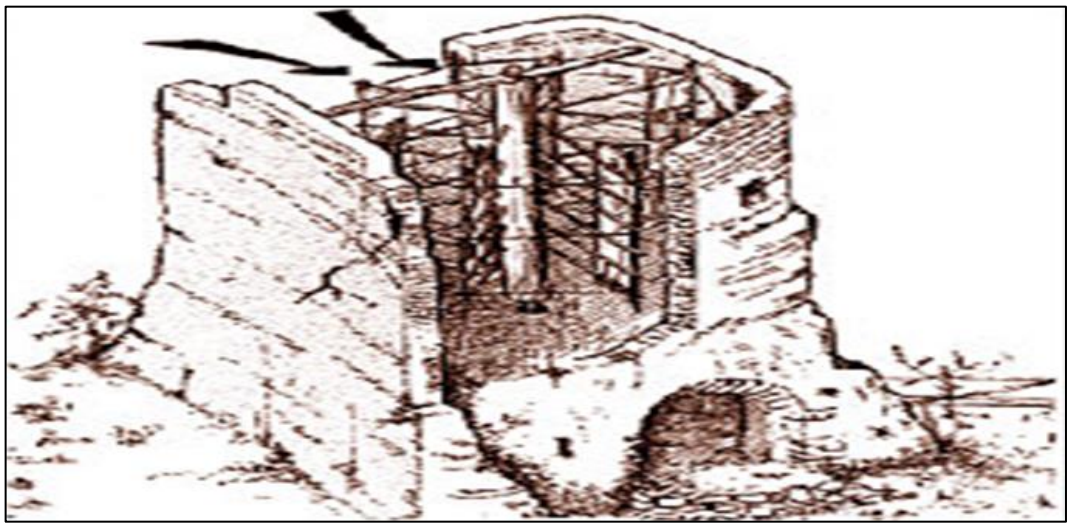

Figure 1.1. The first windmill built by Hammurabi.

By the seventh century AD, building windmills was a well-established trade in the Middle East. Merchants and Crusaders who had returned to Europe in the eleventh century carried the windmill with them. The Persian version was later enhanced by the

Dutch and then by the English. More advanced horizontal-axis windmills were discovered in France and England by the 12th century. There were 10,000 windmills in the Netherlands alone in the 18th century, used for sawing, pumping, and grinding. By the end of the 19th century, windmills were widespread throughout the Great Plains of the United States. Golding has offered an outstanding analysis on the growth of wind machines (Amjady et al., 2011).

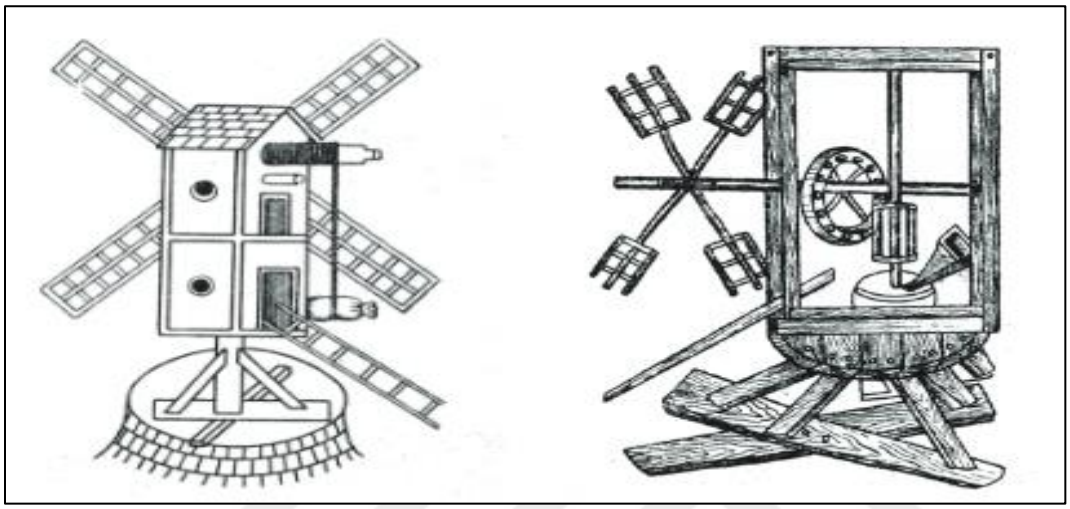

Figure 1.2. Oldest design drawings of the post-mills by Mariano Jacob (Pilipets et al., 2014).

The focal point of energy plans in different nations is progressively moving towards energy preservation and discharge decrease as essential energy utilization and contemporary provokes keep on emerging. Since wind energy is presently generally coordinated into the network, issues like its unpredictability and intermittency cause wind energy prediction slip-ups to straightforwardly affect functional decisions involving it in the power framework. To settle matrix association issues and keep up with network dependability, wind power should be definitively anticipated subsequent to being associated with the lattice since the haphazardness and unpredictability of wind speed bring about power quality that can't satisfy the requirement for framework association (Kusiak & Song, 2010).

Shallow learning models and Profound Learning Models (PLM) are the two general classifications into which AI based prediction calculations can be isolated. Profound learning is a new subfield of AI that tends to the blemishes in shallow learning models and is being used increasingly more in the field of wind power prediction. A genuinely restricted Long Short-Term Memory (LSTM) calculation was put out by Luo (Luo, 2017). The Material science Obliged LSTM (PC-LSTM) model decisively

increments prediction exactness when contrasted with regular AI and measurable procedures. Chen et al. picked highlights with great connection utilizing the Pearson relationship coefficient. The factors used as contribution for the LSTM model are the factors for temperature, moistness, and sunlight based radiation power. The recommended model shows further developed prediction precision when contrasted with a solitary LSTM model, Back Engendering (BE), Radial Basis Function (RBF), and Time Series (TS) calculations (Asis & Dhiren, 2012).

Goa (Gao, 2020) made a prediction model in view of LSTM and Discrete Dark Model (DDM) for non-ideal weather conditions by joining mathematical climate prediction and LSTM to estimate power age under ideal climate conditions. Its prediction results are more exact than those of the Wavelet Brain Organization (WBO), BP, and Least Square Help Vector Machine (LSHVM) in terms of precision. The LSTM organization, one of the previously mentioned profound learning models, succeeds at dealing with time series forecasting because of its unmistakable design and is a rendition to determine the issue of detonating and evaporating slopes in Recurrent Neural Network (RNN). Subsequently, a short-term wind turbine power estimate model in light of LSTM is made in this article (Ayadi et al., 2020).

Forecasting wind energy is a difficult errand. It is trying to match the genuine prerequisite in terms of prediction precision because of the impact of various variables. Along these lines, analysts locally and abroad have proposed various strategies for anticipating wind power. To further develop causality, Shivam (Shivam, 2020) utilized a convolutional Neural Network (CNN) to figure results by taking care of residuals from a one-layered preparing dataset into the organization. Notwithstanding, the CNN model's design is excessively perplexing and it has significantly more hyper-boundaries to deal with this issue. Hence, as of the present moment, a recursive brain network is the best methodology for planning an Artificial Neural Network (ANN) to handle the succession displaying issue. Recursive brain organizations (RNN) take into consideration the development of step conditions, which is reliable with the articles' portrayals of the fleeting coherence of wind speed prediction. Scientists use "gating" techniques in RNN models, which basically contain calculations like Gated Repetitive Units (GRU) and LSTM, to settle the issues of detonating and evaporating angles in RNN models. The exhibition of GRU and LSTM are tantamount in an examination of the group of writing. Yet, while differentiating the two models, apparently the LSTM-based model is more precise. Nonetheless, because of its broad approval, the LSTM-based model seems, by

all accounts, to be more exact than the other two models. Every approval uncovered precise predictions utilizing different wind datasets from around the world (Banna et al., 2014).

1.1. Wind Power

The basic tenet of heat transport on Earth is what drives atmospheric winds, a natural phenomenon. The uneven heating of the Earth's surface by the sun's radiant energy is the primary cause of these winds. Sunlight strikes various parts of the Earth, where it is absorbed and re-radiated at different rates. Variations in temperature and, consequently, atmospheric pressure are caused by this differential heating. Lower air pressure is experienced in warmer regions whereas higher pressure is felt in colder regions. This pressure differential creates the conditions for air masses to move, which is what we usually refer to as wind (Carpinone et al., 2015).

In its purest form, wind is just air moving. It is a dynamic and ever-present force that is crucial in determining the climate and weather patterns on our planet. It has proven extremely helpful to harness the energy of the wind for a variety of uses, especially in the context of the production of renewable energy.

Innovative machinery known as wind turbines is used to collect and transform the kinetic energy of the wind into mechanical energy. According to O'Boyle (O'Boyle, 2017) the name "windmill" originally referred to devices that were principally used to mill grain using wind energy. The phrase "wind turbine" has gained in popularity in modern times, reflecting the wider range of uses for this technology. When used to produce electricity, wind turbines are also known as Wind Energy Conversion Systems and occasionally as wind generators or aero generators (Rakeshchandra et al., 2013).

Wind energy usage has significantly increased recently, solidifying itself as one of the fastest-growing technologies globally. The urgent need to move away from fossil fuels and cut greenhouse gas emissions to combat climate change is the main factor fueling this spike in popularity. In order to battle climate change and fulfill the rising energy needs of a growing population, wind power is an essential component of worldwide efforts. It provides a sustainable and ecologically friendly alternative to existing energy sources (Diaconu, Onea, & Rusu, 2012).

1.1.1. Wind power potential assessment process and its stages

The group of technologies and analytical techniques known as wind power potential assessment is used to determine how much wind resource will be available for

a wind power plant during its useful lifetime. An evaluation of the plant's wind potential gives a broad picture of how much electricity it will produce. The ability to produce energy is a key factor in project success for investors and developers. An extremely significant event is the successful conclusion of the wind power potential assessment process. The ability to accurately estimate energy production in a sizable wind farm depends on much more than just being able to monitor wind speed at a specific moment (Vinhoza, 2021).

The process of assessing the wind power potential can be divided into three fundamental phases, including preliminary area identification, wind resource appraisal, and micro siting. A much bigger region is screened in the first stage's preliminary area identification phase. Based on pertinent information, such as meteorological weather data, wind resource maps, terrain data, topography, and other indications, the proper wind resource areas are selected. A wind resource evaluation is the second stage, where wind measurement programs are used to evaluate the wind resource in a specified location where wind power development is being taken into consideration. This stage involves determining whether the region has sufficient wind resources, evaluating the chosen wind turbines' economic viability, and keeping an eye out for prospective locations for installing wind turbines. A micro siting is the third stage. It refers to the area where one or more wind turbines can be placed in close proximity to one another to optimize the amount of wind energy produced in that particular area (Dolara et al., 2017).

One of nature's most prevalent renewable energy sources is wind. It is an unpredictable, erratic, and uncontrolled variable. Numerous factors, including temperature, stress, topography of the land, the landscape, the region, etc., have an impact on the wind profile. But among the most sustainable energy sources, wind stands out since it has positive environmental impacts and is much simpler than the others. Wind energy is plentiful, renewable, accessible, and clean, similar to using fossil fuels. Additionally, wind uses the least amount of water, emits the fewest greenhouse gases, and requires a limited amount of land. In contrast to photovoltaic, wind generators have lower installation costs and use high-efficiency power converters, making them the most dependent source in the recent past, despite the fact that solar cells have seen a similar level of attention. The Wind Energy Conversion System is a potential source of alternative energy in the future (Duan et al., 2021). Due to its qualities, it has attracted a lot of interest and is an endless supply of energy. Since wind energy has a great potential in the majority of locations worldwide, it stands out among RES and is the most

encouraging. Wind energy is primarily generated by turning wind turbine blades through airflow. Due to variations in wind speed, which convert mechanical power, the production strength of the power changes? Therefore, there is a broad range of energy production employing wind turbines with vertical and horizontal axes in WECS. Less energy is produced by the vertical type design. The use of wind power generating is widespread among electric utilities worldwide. Wind farms offshore are more stable and powerful than those on land (a short distance from the shore). In places like offshore islands, where fuel is frequently expensive and wind initiatives are particularly advantageous, using wind energy might be an intriguing option. However, the expenditures for building and upkeep are significantly higher. The wind farms contain a large number of individual Wind Turbines (WT) (Gao et al., 2020).

Compared to coal or gas-fired power plants, wind is a viable supply of electricity. Onshore small wind farms are able to contribute a little amount of energy to the grid or supply off-grid electricity to some areas far from land. The larger turbines are used for home power supply and are dispersed across a greater geographic area. Any excess power is then sold back to the utility provider via the electrical grid. A small turbine can be used to power boats, caravans, and battery chargers as well as traffic warning signs. Small wind turbines for boats and RVs can be as small as a 50-watt generator. In rural areas, traffic signs are powered by hybrid solar and wind systems. In any case, it is necessary to design compact standalone systems. Small units have direct current output, direct drive, bearings, and lifelong aero elastic blades. The National Renewable Energy Laboratory defined small wind turbines as less than or equal to 100 kilowatts. Energy ranges up to 700 kW now surpass previous ones thanks to later developments. Different types of huge turbines are developing as a substantial source of sporadic RES. little wind turbines are used for on-grid or off-grid homes, remote monitoring, offshore platforms, telecom towers, rural schools and clinics, among other uses (Harrouz et al., 2019).

The original wind turbines were typically constant velocity turbines with straight induction generators (Hau, 2013) and gearboxes connected to the grid. This structure, which is still widespread in Denmark, is the least adaptable and has the most negative impact, necessitating the compensating of installed devices at times. In the wind energy industry, fixed-speed WECS with either a functioning or detached slow down have long ruled. Fixed speed WECS enjoys the benefits of being basic, dependable, and proficient with straightforward and cheap electrical parts and very much demonstrated activity. Then again, the fixed-speed process requires the steady mechanical pressure. Since there

is little flexibility in changing the set generator velocity, their main flaw is rigidity. Furthermore, because the rotor speed is fixed at the grid frequency and is nearly continuous, fixed-speed WECS has very little controllability. The electronic interface power converter, which enables full or partial decoupling from the grid, makes the variable velocity approach possible. A wind turbine can be projected for a procedure with constant or variable velocity. Compared to its steady-speed equivalents, variable wind turbines can produce up to 15% more energy, but they need digital power converters to supply their loads with a fixed frequency and fixed voltage. This plan provides variable speed operation with a power converter for electronics over a sizable but constrained area, and its controlling controllers run the generator. Wind powers varied output, however, has significant consequences across shorter time frames that are remarkably consistent from year to year (Hau, 2013).

Electricity is used in conjunction with other sources to provide a consistent supply. A reduced capacity to replace conventional output may result from the requirement to restructure the grid and the proportional increase in wind power. To overcome these challenges frequently When wind production is low, power-management techniques can be used to have a variety of capacities, spread turbines geographically, dispatch able backup sources, appropriate hydroelectric power, to neighboring territories, and import the electricity. Furthermore, weather forecasting allows the electric-power network to be ready because the variations that occur are predictable. Connection to the grid is required for current conversion systems of wind energy with an efficient power converter due to the variable wind speed qualities for a stable task gigantic measure of energy storage or other source of energy when the turbine is filled for an isolated region in as a voltage source (Frandsen, S, 1992).

1.1.2. Wind distribution

Wind turbines are strategically positioned within the Earth's atmospheric boundary layer, which rises anywhere from a few hundred meters to several kilometers above the surface, and use the kinetic energy of flowing air to generate power. This area, where the Earth's surface and atmosphere physically interact, is characterized by a variety of dynamic characteristics and behaviors that have a substantial impact on wind energy production. The height above ground at which wind turbines are sited is an important factor to take into account since it affects the type and quantity of wind resource that can be used to generate electricity.

Practical factors and the peculiarities of the local wind regime are frequently taken into account when determining the height at which wind turbines operate. This height is typically defined as the altitude above the Earth's surface at which turbulence practically vanishes. Due to interactions with the topography and surface features, turbulence is common in the lowest part of the atmospheric boundary layer, close to the Earth's surface. Wind speed profiles have a tendency to stabilize and follow more consistent patterns as one moves above this layer. For wind energy projects, this change in wind behavior is essential because it enables the construction and positioning of turbines that can effectively capture the energy from the stronger, less turbulent winds (Wang, 2011).

The atmospheric boundary layer is characterized by an increase in wind speed with height. Frictional forces and drag from the Earth's surface cause this phenomenon, called wind shear. Buildings, trees, and hills that interrupt the airflow close to the ground result in slower wind speeds. However, wind speeds often rise as one ascends above the surface. The design and positioning of wind turbines must take this vertical gradient in wind speed into account. Wind turbines are frequently mounted on tall towers to access the stronger and more reliable winds that can be found at higher altitudes in order to enhance energy output. With the help of this method, wind energy projects can more effectively produce electricity while utilizing a bigger amount of the available wind resource.

1.2. Advantages and Disadvantage of Wind Turbine

The goal of wind farms and wind turbines is to capture wind energy and transform it into useable energy, such as electrical and mechanical energy. Given the fundamental principle of energy conservation, it follows that kinetic energy collected from the wind will result in a decrease in downstream kinetic energy relative to kinetic energy upstream of the wind turbine. Subsequently, the wind downstream of a wind turbine is tempestuous and has a lower speed; this wind is the turbine's wake. Thusly, bunching turbines in ranches makes two critical issues: diminished power yield because of wake speed lacks and more noteworthy unique burdens on the edges because of higher choppiness levels. The power loss of a downstream turbine in full-wake conditions can without much of a stretch methodology 30-40% comparative with the upwind turbines, and weakness burdens can depend on 80% more noteworthy than the upstream turbines, contingent upon the setup and wind states of a wind ranch. This wake will begin to spread and gradually reestablish to free stream conditions as the wind stream moves further downstream (Louka et al., 2008).

Since the beginning of the expanded interest in the utilization of wind energy in the last part of the 1970s, wind turbine wakes have been an examination region. The streamlined features of wind turbines might show up moderately direct from an external perspective. However, the truth that the admission is consistently helpless to stochastic wind fields and that slowdown is an inborn part of the functional climate for machines without pitch guideline confuses the definition. In spite of the fact that the wind turbine is among the earliest methods of harnessing wind energy (together with the sailing boat), some of the most fundamental aerodynamic principles guiding the flow are still poorly understood.

Most research on wakes has distinguished between near-wake and far-wake regions; the relationship between the two regions is still not well understood. The area up to three diameters downstream of the rotor is considered to be the near wake. The rotor's impact is most noticeable in this situation. The near-wake zone is characterized by strong turbulence produced by the blades, shear, and tip vortices degrading that transport a variety of length scales. The area beyond the near wake is called the far wake (Luo et al., 2017).

Since the speeds upstream and downstream of a wind turbine conveyed in the climate are regularly in the scope of 5-25 m/s, it is ok to expect that the stream field in the wakes of wind turbines is incompressible. The best relative Mach number in light of the edge tip speed is commonly under 0.2 in estimations, in any event, when the rotor is displayed straightforwardly, and the incompressible the streamlined features of wind turbine wakes can be demonstrated utilizing the Navier-Stirs up conditions. Notwithstanding offering a complete model for the portrayal of violent streams, this arrangement of conditions is trying to settle. The presence of the non-straight convective component, which creates an extensive variety of time and length scales, makes fierce streams testing. For example, the biggest tempestuous scales in the air limit layer are on the request for 1 km, while the littlest scales are on the request for 1 mm. The scales are significantly more moment inside the sharp edge limit layers. The Reynolds number (Re), a dimensionless metric that addresses the proportion of convective powers to thick powers in a stream, determines the scope of scales. Huge upsides of the Reynolds number, which are met in the sharp edge and wake calculations, bring about a wide assortment of scales, which drives up the expense of programmatic experiences. It is not possible to resolve every scale in the flow using so-called Direct Numerical Simulation. On the basis of the

behavior of the big scales, turbulence models must be built, representing the impact of the unresolved small scales (Memarzadeh & Keynia, 2020).

1.2.1. Usage of Generators with Wind turbine

Generators can be connected to wind turbine systems either synchronously or asynchronously. Low speed or high speed drive trains are coupled to the generator depending on the needs for generator speed. No concurrent wind turbine generators incorporate the squirrel cage induction generator (SCIG) and wound rotor induction generator (WRIG). Simultaneous wind turbines are the ongoing business standard. The two most famous sorts of wind turbine generators are the Permanent Magnet Synchronous Generator (PMSG) and Doubly Fed Induction Generator (DFIG). When the squirrel cage induction generator is operating in generator mode, unfavorable slip occurs. The turbine speed is adjusted by the gearbox to the appropriate rated generator speed. The main drawback is that it offers little assistance with velocity control. The concept of variable velocity is employed by WRIG. Controlling energy production and generator slip is done by adjusting the rotor's power. Reactive power and inrush current are reduced by using soft starters (Nielsen et al., 2006).

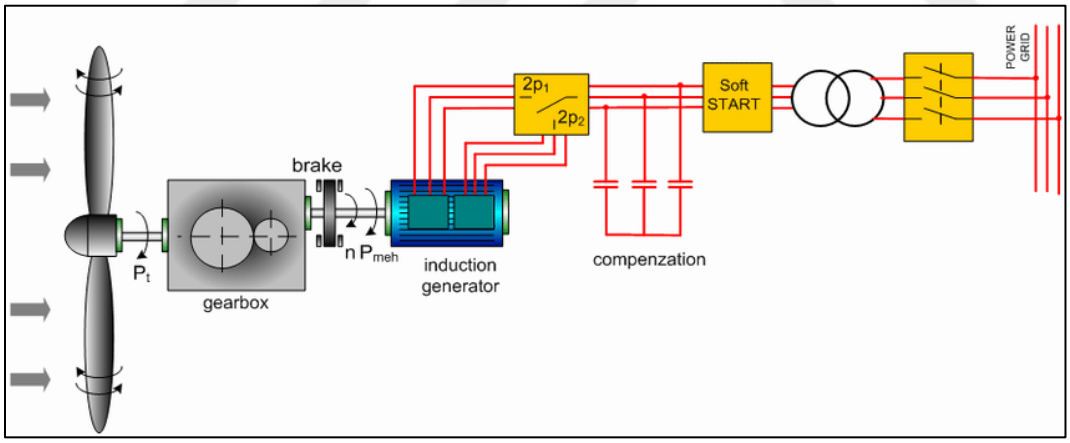

Figure 1.3. Fixed-speed wind urbine with induction generator (Drago Ban et al., 2021).

DFIG is a WRIG with rotor windings connected to the AC-AC source and stator windings that are directly connected to three phases, the constant frequency grid, and the grid. Through the stator and rotor of the generator, it transmits wind energy in two directions. It can be supported by either a rotor side or a grid side converter. The stator receives the voltage from the grid, and the power converter then induces the voltage to the rotor. Depending on the application stage, or slip location, energy is either supplied or consumed. For instance, the system absorbs energy if the slide is supplied with adverse

side energy. The active and reactive power flow between the grid terminals and the generator is controlled by the power converter in the DFIG linking capacitor. A crowbar is installed between the generator and the converter in the wind energy system to prevent short circuits. Depending on the rotor velocity of the generator, the Rotor Side Converter (RSC) controls the DFIG wind turbines' flux when operating at the slip frequency. The entire effective and reactive power control capacity of the converter is used to determine the power rating of the RSC (O'Boyle et al., 2017). A magnet is used as the excitation current's source rather than a coil in a permanent magnet synchronous generator. The generator is connected using the full-scale converter according to the needs of the grid. The converter aids in controlling the generator's effective and reactive energy output to the grid. This kind of generator is also used in several wind turbine variants. The parallel component of the stator field and the perpendicular component, which effect electromagnetic torque, both have an impact on the generator voltage. The load generator controls the voltage. The angle between the rotor and stator regions will be more than 90 degrees and correspond to the generator voltage if the load is inductive. This is seen as a generator that is overexcited. Extremely durable magnets of unrivaled quality have underlying and warm issues. They require a fitting cooling framework since attractive materials are temperature delicate and can lose their attractive properties whenever presented to high temperatures (Osório et al., 2014).

1.3. Numerical Weather Prediction & Wind Forecasting

The decision of the particular NWP model is a vital stage in the improvement of a NWP-based wind power figure model. Geological locale, goal (both spatial and fleeting), figure skyline, required accuracy, computation time, and number of runs are critical choice factors. The powerful focus, which portrays the adiabatic non-gooey stream, the actual conditions making sense of fluctuation of the meteorological cycles (like disturbance and radiation), and the data gathering programming code are the three essential pieces of NWP models. Thus, instead of just foreseeing the wind, the result of a NWP model is an intensive figure of the condition of the climate at a particular time. NWP projections are used by many different companies, sectors, and governmental organizations; they are not just prepared for the power sector. NWP is sensitive to beginning conditions; hence ensemble forecasting is utilized to get around this. The Kalman filter can eliminate systematic forecast errors in NWP wind speed estimates, as demonstrated by Louka (Murali et al., 2014).

Most NWP do exclude sea models since climatology is utilized to portray ocean surface water temperatures. Hurricane Group Model by the Japan Meteorological Organization is one illustration of a particular NWP model that has been made to distinguish storms in the Pacific and Atlantic. To suit the necessities of their clients, most of meteorological administrations solely offer on-shore and close shore climate predictions. Accordingly, the objective of current worldwide NWP models has been to create more exact climate predictions for the land. To address their conditions, worldwide NWP models essentially require land surface factors, especially temperature. For time frames longer than 4 hours, NWP holds the best. Most of these devices, which are known as deterministic, spot, or point predictions, just produce a solitary expected incentive for each figure timetable notwithstanding the way that most models are multi-step and proposition look-ahead times for a few skylines. Thus, their application to stochastic advancement and chance investigation is confined.

One more group of NWP models was made at the provincial and musicale levels to focus on neighborhood climate occasions specifically. The hydrostatic estimated time of arrival model, the HIRLAM model, and the ALADIN model are a couple of models. Extra models incorporate the more present day Weather Conditions Exploration and Figure (WRF) territorial model as well as the openly available MM5 local model made at Pennsylvania State College and used by the Public Focus of Climatic Exploration in the US of America (USA). To figure wind power in a country or a district of a country, some NWP models are applied at the territorial level. It could require a great deal of investment to foresee the wind power yield from each and every wind ranch, consequently a technique known as "up scaling" is used all things considered. The wind power creation from an example number of wind ranches fills in as the establishment for reference information for scaling. Since the conjecture mistake is found the middle value of over the whole area, up scaling may seem to decrease it. When downscaling, physical as well as measurable models are utilized to make additional exact geological data from coarse NWP yields. Like NWP, physical downscaling models work at improved goal across a more modest locale. The power or potentially wind speed at a genuine wind ranch and NWP are used in factual downscaling models to make an exchange capability that can be utilized to estimate wind power from more wind ranches in a district (Saidi et al., 2019).

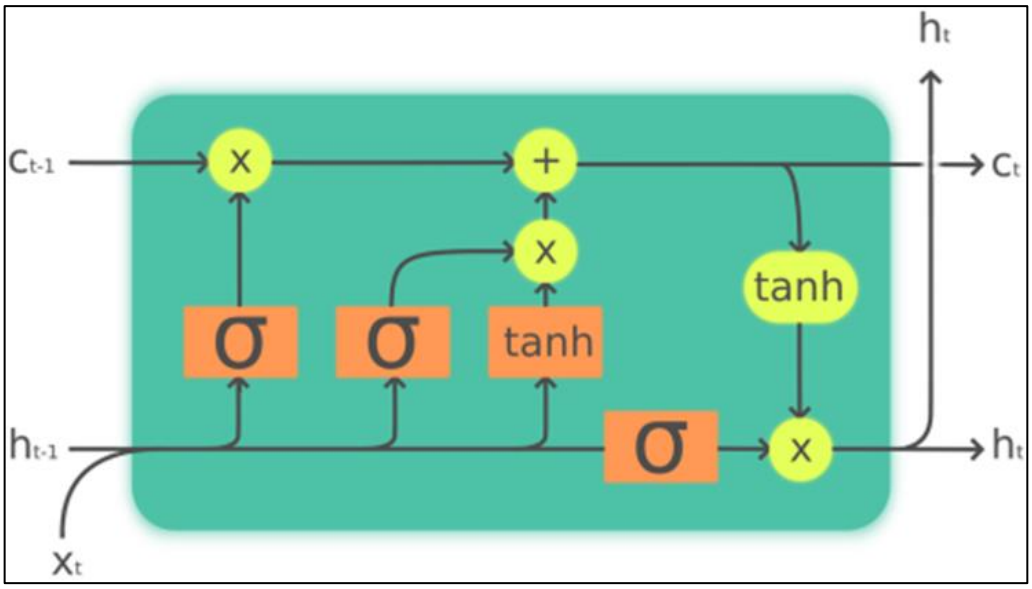

Figure 1.4. Frame of short-term wind power forecasting.

1.3.1. The basics of wind power forecasting

The three principal classes of wind speed or power gauge methods are physical, measurable, and keen methodologies. Analysts have given the savvy approach different names, including information-based approaches, computational knowledge strategies, man-made brainpower techniques, and numerous others. It is critically important to change from deterministic forecasting, otherwise called point forecasting, to probabilistic forecasting since wind energy age is arbitrary. The forecasting of wind power and wind speed additionally displays this irregularity. Probabilistic forecasting could offer the quantitative vulnerability data that is important to control the silliness of the activity of the power framework. Moreover, numerous specialists attest that the mixture approaches address the fourth way for wind prediction. While half and half models are commonly portrayed as a mix of physical, factual, or clever methodology approaches with information pre-handling or potentially post-handling strategies, Wang (Wang J. H., 2016) characterized crossover models as the blend of physical and factual techniques. Furthermore, spatial connection procedures address an unmistakable prediction methodology. The arrangement and presentation of different wind prediction frameworks, as well as their advantages and viability, are the fundamental subjects of this part.

While making these models, actual methodologies consider actual information like geology, air temperature, tension, and environment factors. These procedures need a ton of computation time since they are reliant upon a bunch of numerical conditions for

these multivariate actual boundaries. This at last outcomes in the actual strategies being unacceptable for forecasting short-term wind information. Regardless of having a bigger registering intricacy, these techniques give off an impression of being exceptionally precise for enormous scope and long-term prediction. Among actual models, the Mathematical Climate Prediction (NWP) approach is incredibly famous. Albeit well known Markov models and more adjusted and high level NWP approaches have been utilized in wind speed applications, these techniques are not generally utilized because of the expanded processing intricacy and absence of availability to all market members of the fundamental actual data. A physical method may not be as effective as the spatial correlation method. This approach bases the forecast on the sites and the sites that are close by. Correlated wind speed measurements at multiple locations at once are challenging, though (Shahid et al., 2020).

The factual techniques, conversely, ordinarily construct the models in light of the amounts of authentic information. Measurable models are dependable for making short-term gauges and are nearly easy to send. For long-term forecasting, be that as it may, factual techniques truly do less well. In measurable strategies, ARMA, ARIMA, the Pattern Succession Based Forecasting (PSF) technique, Kalman channels, model-based approaches, Molecule Multitude Streamlining, and a lot more techniques are utilized for prediction. For the nonlinear properties of wind information, the prediction utilizing such factual strategies was not adequate. In any case, these methods are all the more broadly utilized on the grounds that they are less pricy, prominent, and have more reasonable techniques.

Like the factual methodology, the Counterfeit Prediction system is likewise reasonable for short-term prediction, but these strategies portray the relationships in a profoundly nonlinear manner as opposed to using deterministic approximations. Well known strategies utilized in the man-made reasoning methodology incorporate ANN, fluffy rationale, SVM, and Radial Base Function (RBF). These refined procedures precisely expect short-term wind. These procedures, notwithstanding, have the drawback of being "black box" strategies since appreciating their laws is exceptionally difficult. Yet again nonetheless, fluffy rationale can be utilized to gauge such guidelines, but since these methods manage such countless factors, perception is a test. Canny techniques commonly catch nonlinear collaborations inside wind information and produce preferable predictions over factual and actual methodologies. Besides, as expressed in, the

coordination of at least two fake methodologies has exhibited its adequacy in wind predictions (Shivam et al., 2020).

The prediction with models of factual or canny strategies was not as viable as trusted since wind information are incredibly sporadic and intermittent. Thus, there has been a propensity to embrace half breed approaches for wind forecasting lately. By consolidating numerous models, these half breed methods were put out. Each prediction model has frequently recognized a couple of advantages and downsides. The thought behind hybridization was to use numerous prediction models while limiting the shortcomings of each model. The mixture procedures perform obviously better than anyone prediction technique. These cross breed strategies are much of the time alluded to as a gathering forecasting method. Cross breed strategies are commonly separated into helpful and serious classes. In cutthroat strategies, different prediction models were utilized to make predictions on similar information simultaneously, and the normal of the last predictions made by each model was taken as the last prediction. Conversely, in agreeable strategies, prediction errands were isolated into more modest ones, and each sub-task was relegated to the taking part prediction models in the crossover technique in light of these more modest assignments' attributes. By including the result discoveries from every individual technique, a definitive estimate is gotten (Simon & Bruce, 2012).

The hybrid models that increased wind prediction accuracy by fusing two prediction techniques. However, the prediction performance has been raised to a higher degree by the combination of pre-processing or/and post-processing approaches with one or more prediction methods. The most often used pre-processing techniques for wind applications are MLP, EWT, EMD, EEMD, and WD. These techniques alter wind data by extrapolating or breaking it down into components with various frequency. On the other hand, post-processing techniques classify and modify the anticipated wind data in accordance with other existing predictions to improve prediction accuracy.

In general, there are three types of wind power forecasting models: model-driven, data-driven, and hybrid. Model-driven forecasting models are used to predict wind power. The model-driven approaches call for a wealth of meteorological expertise as well as knowledge of the different physical components influencing wind power. While in data-driven techniques, forecasting is done using data-driven statistical models. With the development of artificial intelligence and data sciences, this approach can produce predictions that are more accurate. The only requirements for such models are historical data. Numerous studies have examined the effectiveness of various data-driven models,

including simple persistence models and more complex models like SVM, NN, ARIMA, and many others. However, due to the highly unpredictable and erratic nature of wind power time series, accurate forecasting becomes challenging. The hybridization strategy, which involves integrating two or more models to anticipate the data for wind power, is frequently employed to address this issue. The various hybrid models (WMD-LSSVM-AR, WRF-SSA, EMD-LSSVM, grey relational analysis, and wind speed distribution based hybrid models) and numerous studies have demonstrated how hybrid models are superior to individual or single techniques (Srivastava et al., 2020).

Understanding the features in order to develop a prediction approach becomes challenging due to the chaotic and extremely complex nature of wind speed and power data. The analysis of the data features becomes crucial in order to offer a stable prediction model. In order to more thoroughly study the time series characteristics, it may be preferable to decompose such time series.

A particularly well-known and successful strategy is the hybridization of decomposition techniques. The wind power time series is divided into different subseries using a decomposition approach, and the cumulative forecasts of each subseries are considered as forecast outcomes. The most popular decomposition techniques for predicting wind power time series are the WT and EMD (Sun & Zhao, 2020).

While the EMD technique uses a preset methodology regardless of the type of data, the decomposition with wavelet transform requires prior knowledge of the data. Different hybrid EMD models that combine different prediction techniques have demonstrated improvements in prediction accuracy. It is suggested to use the EMD-ANN model, in which each subseries of wind speed is anticipated using the ANN technique. EMD-ARIMA, EMD-SVM, and numerous other techniques were presented based on a similar premise (Syu et al., 2020).

However, the EMD method's mode mixing issue has negatively impacted the accuracy of the results. The EEMD approach was put forth by Wu and Huang to lessen the consequences of the mode mixing issue. Some hybrid EEMD models are EEMD-GA-BP, EEMD-SVM, and EEMD-SSA-ENN. When compared to EMD models, these models demonstrated noticeably greater prediction accuracy for wind speed and power data. The upsides of such models over clear factual, keen, and cross breed models with other pre- and post-handling strategies for short-term wind predictions are examined in an itemized and careful survey that makes sense of the meaning of EMD/EEMD based half and half models, the various methods of hybridization, and the prevalence of such models.

1.3.2. Importance of Short-term forecasting

For diverse goals, several forecasting horizons have been utilized. The four main time scales for wind power forecasting are actually extremely short-term (seconds to 30 minutes), short-term (30 minutes to 6 hours), medium-term (6 hours to 1 day), and long-term (1 day to 1 week). Supplying operators with means of ensuring that the turbine may not be impacted by strong winds. Short-term projections are the main concern of the economic load dispatcher. Choose when to turn on or off extra power generation.

While medium-term projections are utilized to empower energy exchanging, long-term conjectures are normally used to plan fixes and upkeep. The most broadly involved models for both short-term and super short-term wind power conjectures are those in view of brain organizations. It is a major reward since they can work with nonlinear information.

Short-term forecasting is utilized to oversee planning, load following, and clog and has a time frame of 30 minutes to 6 hours. The strategy attempts to give the most ideal power booking and dispatch for the next day in light of the data given by the generators. Forecasting by and large empowers functional organizers to design the lattice and the creation of power. It would be trying to increase and down steam-based creating rapidly without the perceivability of RE power (Toubeau et al., 2021).

The capacity of the power matrix framework to deal with critical expansions in wind power yield is seemingly the most concerning issue while coordinating a tremendous volume of wind power information. Different geographic and worldly scales affect wind incline occasions, and a blend of up inclines and down slants with different levels of force might happen (Syu et al., 2020).

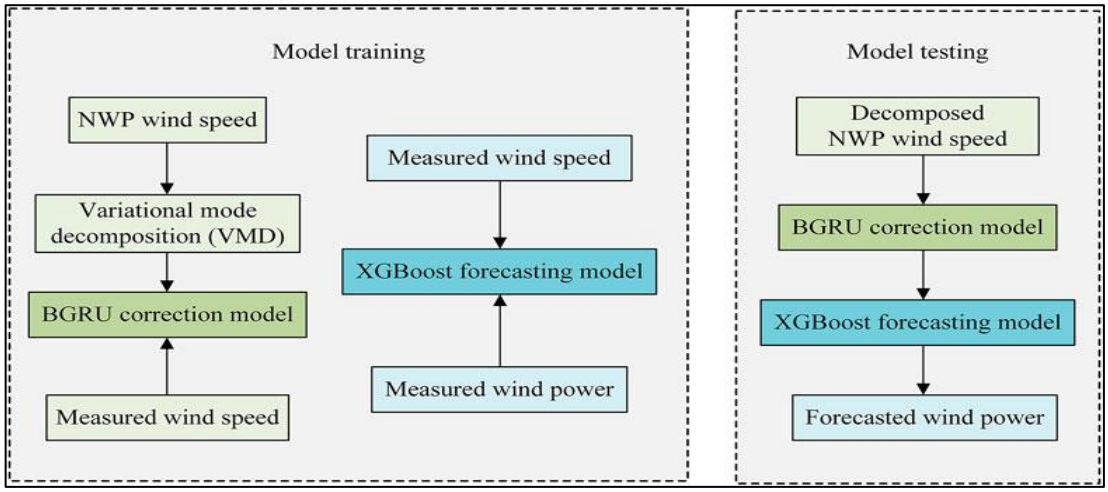

Figure 1.5. Frame of wind power forecasting.

1.4. Neural Networks

Repetitive brain organizations, rather than ordinary feedforward brain organizations, include a memory part. These brain organizations might deal with inputs that are given as a series by utilizing the condition of their inside memory. These are normally favored when the outcome relies upon various sources of info. Regular language handling, text prediction, and discourse acknowledgment are a couple of such purposes.

The name "Recurrent" in this neural network denotes that the same set of operations are carried out again on each element in the input sequence in turn, and the results of each operation contribute to the RNN's final prediction. As the only type of artificial neural networks with this capability at the moment, RNNs are anticipated to provide improved prediction accuracy for sequential data or in situations where context is important, such as text prediction.

Although the aforementioned advantages of RNNs are valid in theory, plain vanilla RNNs can only look back a small number of steps. This is explained by the way plain RNNs are taught.

The process of backpropagation is used to train neural networks. This algorithm can be changed to better suit the network based on the type of neural network being used. Backpropagation through Time (BPTT) is the name of the training algorithm used in RNNs. This approach uses sequential input and goes back in time steps to modify the weights and biases. The way BPTT operates conceptually is by unrolling each input time step. At each time step, errors are calculated and accumulated. The weights are then updated before the network is rolled back up. For instance, a single weight update for an input sequence with 1,000 time steps needs the calculation of 1,000 mistakes, one for each time step. The gradient, also known as the partial derivative of the error function with respect to the current weight, determines how much the weights are updated. With plain RNNs, this gradient tends to decrease after several training rounds until it reaches a point where it is so minute that it scarcely modifies the weights and the neural network stops learning altogether. The term "vanishing gradient problem" applies to this(Vermeer et al., 2003).

1.4.1. Types of RNNs

Recurrent neural networks are a fundamental class of neural networks utilized in numerous sequential data applications, including time series forecasting, speech recognition, and natural language processing. Long Short-Term Memory Networks

(LSTMs) and Gated Recurrent Unit Networks (GRUs) are two RNN varieties that have grown significantly in prominence because of their capacity to address some of the problems that standard RNNs have inherently (Venayagamoorthy et al., 2012).

Long Short-Term Memory networks, also known as LSTMs, are a particular type of RNN created to get around the issue of vanishing gradients, which frequently prevents conventional RNNs from capturing long-term relationships in sequential data. This is accomplished by LSTMs by including a memory cell and a group of gates that control the information flow within the cell. The main strength of LSTMs is their ability to retain information over long periods of time, which makes them ideal for applications requiring the modeling of intricate temporal relationships. The input gate, forget gate, and output gate enable LSTMs to efficiently collect and use long-range dependencies by allowing them to selectively update and access data from prior time steps.

The vanishing gradient problem is also addressed by Gated Recurrent Units (GRUs), a different RNN variant that uses less parameters than LSTMs. GRUs are an improved iteration of LSTMs that streamline the architecture by utilizing fewer gates. The reset gate (rt) and the update gate (zt) are the only gates present in a typical GRU cell. The update gate controls how much new information should be incorporated, whereas the reset gate controls how much of the old cell state should be forgotten. In some situations, this simplified architecture might increase the computational efficiency of GRUs and make training them simpler.

Depending on the precise requirements of the task at hand, LSTMs or GRUs are selected. The profound grasp of long-range dependencies required by LSTM tasks makes them ideal for jobs like speech recognition and machine translation. GRUs, on the other hand, are preferable when the vanishing gradient problem must be solved while also prioritizing computational simplicity and efficiency. It has been discovered that GRUs perform well in tasks like sentiment analysis and text production.

1.4.2. Long short-term memory

The Long Short-Term Memory (LSTM) network is a significant headway in the field of repetitive brain organizations (RNNs), and it was made fundamentally to take care of the disappearing slope issue, which has tormented standard RNNs for quite a while. This problem occurs when gradients, which are used to update the weights of the network during training, shrink dramatically as they go back in time. Traditional RNNs struggle with the vanishing gradient problem, making it difficult to identify long-range dependencies in sequential data.

By adding specialized architectural elements that make it easier to preserve and manage information across long time horizons, LSTM networks get around this restriction. The network's goal is to offer a sort of short-term memory that can store data for thousands of time steps, enabling the capture of "long" dependencies within the data. The name "Long Short-Term Memory" itself reflects this goal.

The relative insensitivity of LSTM networks to the size of the data gap over other RNNs, hidden Markov models, and many other sequence learning techniques is one of its main advantages. Traditional RNNs frequently have trouble processing sequences with gaps or missing data, whereas LSTMs are far better at doing so. As a result, they are incredibly adaptable for a variety of applications where the data may contain erratic time intervals or gaps.

The composition of LSTM units, which typically consists of a cell, an input gate, an output gate, and a forget gate, is a crucial aspect of their architectural design. Together, these parts control how information moves throughout the network. The cell functions as the LSTM's long-term memory's central memory unit and may hold values for any length of time (Vermeer et al., 2003).

The forget gate is crucial in determining whether or not to keep certain pieces of information from the previous state. The forget gate decides which components of the prior state are relevant for the current context by assigning values between 0 and 1, with 1 suggesting information to be maintained and 0 indicating information to be discarded.

Similar to the forget gate, the input gate is essential in determining which fresh pieces of information should be kept in the present state. To assess the significance and relevance of incoming data, it employs a similar technique to the forget gate.

The output gate, which is responsible for deciding which information from the current state should be output, also takes into account the prior state. The LSTM network can keep and utilise valuable long-term dependencies when making predictions, both for the present time step and for future time steps, thanks to this selective output of pertinent information.

1.5. LSTM RNN (Recurrent Neural Networks)-Based Forecasting

Due to their exceptional capacity to identify long-term dependencies within sequential data, recurrent neural networks (RNNs) have grown significantly in favor in time series forecasting. They differ from other neural network architectures because of this innate quality, which makes them an effective option for applications involving time

series prediction. RNNs do face some difficulties, the most noteworthy of which is the vanishing/exploding gradient problem.

The backpropagation of gradients across the network during training is what causes the vanishing/exploding gradient problem. The hidden layers and accompanying time steps are connected in deep neural networks, such as those employed in RNNs for time series forecasting, by multiplicative operations. Gradients can therefore either disappear completely (vanishing gradient) or explode out of control (exploding gradient) when they are propagated backward through these actions. Deep RNN training can be substantially hampered by this problem, which also affects how well they can detect long-range dependencies in time series data.

LSTM-RNNs offer a wide range of uses, depending on the particular specifications of the work at hand. They are frequently grouped depending on their use in the context of time series forecasting. For instance, in regression assignments where the objective is to estimate the value(s) at one or more future time points, LSTM networks can be used to forecast single-step or multi-step time series data. Because of their adaptability in modeling various forecasting horizons, LSTM-RNNs are effective tools for a variety of time series prediction applications.

1.5.1. Sequence-to-sequence LSTM RNN

When it comes to time series forecasting, the Long Short-Term Memory Recurrent Neural Network (LSTM-RNN) is a potent and popular architecture. The capacity of LSTM-RNNs to learn and represent sequential relationships within the data is one of its key properties. This is accomplished by teaching the network to forecast the value of the following time step based on the data it has observed thus far in the sequence.

When used for time series forecasting, the LSTM-RNN is trained so that it can predict the value of the following time step for each iteration or time step in the input sequence. This method is also known as "autoregressive prediction" or "one-step-ahead prediction." It indicates that as the network moves through the sequence, it is constantly adjusting its internal state depending on the input data and its own forecasts in order to produce increasingly accurate forecasts. Because it captures the intrinsic temporal dependencies present in time series data, this autoregressive nature is particularly well-suited for such data.

A sequence-to-sequence LSTM-RNN approach produces results or predictions that are simply the training sequences with values pushed forward by one-time step. In other words, the LSTM-RNN generates a prediction for the value at the following time

step for each time step in the input sequence. Up until the required forecasting horizon is achieved, this prediction serves as the basis for the subsequent iteration. As they learn to take into account both short-term and long-term dependencies in the data, LSTM-RNNs learn to include complicated temporal patterns through this iterative prediction process.

1.5.2. LSTM network layer

In the field of profound learning, Long Short-Term Memory (LSTM) networks have turned into a powerful and significant device, particularly for undertakings including consecutive information, similar to time series examination, regular language handling, and discourse acknowledgment. LSTM networks are a specific sort of repetitive brain organization (RNN) that conquers the evaporating slope issue, which is one of the primary disadvantages of traditional RNNs. Because of the vanishing gradient problem, the network has a hard time understanding long-range dependencies in sequential input when gradients are incredibly small during training(Vinhoza & Schaeffer, 2021).

The capacity of LSTM networks to maintain information over longer time periods is one of the primary characteristics that set them apart from conventional RNNs. This is accomplished by using a memory cell, which acts as a storage container inside of each LSTM block. The network can preserve and capture long-term dependencies in the data thanks to the memory cell's ability to store and access data from earlier time steps. This capability is especially important for jobs like time series forecasting and natural language understanding, where knowledge of the past and present is necessary for making correct predictions.

Each LSTM block consists of a memory cell, an input gate, a forget gate, and an output gate. These gates are in charge of managing the information flow inside the LSTM block. The input gate controls how much of the new data should be added to the memory cell through the use of weights and biases. The degree to which the LSTM cell updates its internal state in response to the current input and the knowledge from the previous time step is effectively controlled by this gate.

Another vital part of the LSTM is the forget gate. It regulates how much of the data from the memory cell should be remembered or forgotten. The loads and inclinations of this door empower the LSTM network keep a specific memory of earlier perceptions by determining whether data is as of now not relevant and ought to be erased. To wrap things up, the result entryway determines the amount of the substance of the ongoing memory cell ought to be utilized to register the result of the LSTM block, which is

similarly administered by its loads and predispositions. Then, this result is shipped off extra layers or put to use in prediction.

1.6. Summarize of the Introduction

The selection of a Numerical Weather Prediction (NWP) model plays a crucial role in enhancing wind power forecasts. Factors like location, forecast duration, accuracy needs, and computational constraints determine the model choice. NWP models analyse meteorological conditions, providing a comprehensive outlook on weather instead of just wind prediction. Their sensitivity to initial conditions demands ensemble forecasting to mitigate errors. However, most NWP models focus on land-based weather, lacking sea models. Specific models like the Hurricane Group Model target storm detection in the Pacific and Atlantic. Regional models, such as HIRLAM and WRF, zoom in on local weather events. Upscaling and downscaling methods adjust the wind power forecast from regional to larger scales or vice versa.

Wind power forecasting involves physical, statistical, and intelligent methods. Physical approaches consider climate elements like temperature, demanding more computation time but excelling in long-term predictions. Statistical methods, like ARMA and Kalman filters, suit short-term forecasts due to their reliance on historical data. Intelligent techniques like ANN and SVM prove effective for short-term wind predictions, yet their black box nature poses challenges in understanding the process.

Hybrid models combining multiple approaches improve wind prediction accuracy. These models often integrate preprocessing and post-processing techniques like MLP and EEMD, enhancing the forecasts. Short-term forecasting, crucial for grid planning and load management, relies on models predicting from seconds to six hours ahead. Meanwhile, LSTM-based Recurrent Neural Networks stand out in capturing long-term dependencies in sequential data like time series forecasts. They address issues like the vanishing gradient problem in traditional RNNs, allowing for better understanding of complex temporal patterns.

LSTM networks, equipped with memory cells and gates, maintain information across time steps, crucial for tasks like time series forecasting and natural language processing. They overcome the vanishing gradient problem, enabling the capture of long-range dependencies in sequential data. LSTM variants like LSTMs and GRUs further refine these capabilities, each excelling in specific scenarios. For time series predictions, LSTM-RNNs, especially in sequence-to-sequence approaches, offer iterative forecasting,

predicting subsequent time steps based on past data, capturing intricate temporal patterns in the process.

1.7. Problem Statement

The problem of Wind Energy Prediction using Long Short-Term Memory (LSTM) stems from the critical need for more precise forecasting in the renewable energy sector, particularly in wind power generation. Conventional prediction models often fall short in capturing the intricate and non-linear nature of wind patterns, leading to inaccuracies that hinder efficient utilization of wind resources for electricity generation.

The challenge at hand involves developing an advanced predictive model based on LSTM architecture that can adeptly navigate through the complexities of wind data. This necessitates overcoming inherent obstacles like the vanishing gradient problem, which limits the ability of traditional models to grasp long-term dependencies in sequential data.

The primary goal is to engineer an LSTM-based forecasting system capable of providing robust predictions of future wind energy outputs across varying time horizons. Such a system would not only empower energy grid operators and wind farm managers with more accurate insights but also enable better planning, scheduling, and integration of renewable energy sources into the existing power infrastructure.

Achieving this goal involves leveraging the unique capabilities of LSTM networks to capture temporal relationships, comprehend intricate wind patterns, and generate forecasts that aid in optimizing energy production and grid stability. Ultimately, the aim is to enhance the efficiency and reliability of wind energy generation by harnessing the potential of LSTM-based predictive models (Wang, Guo, & Huang, 2011).

1.8. Objectives of the Study

1. Improve dataset quality through advanced data preprocessing to handle missing values, outliers, and noise, ensuring reliability for accurate wind energy modeling.

2. Investigate hybrid modeling methods merging statistical approaches with machine learning algorithms for more precise and robust wind energy predictions.

3. Incorporate meteorological and geographical elements affecting wind energy generation into feature engineering to better comprehend power output dynamics.

4. Develop interpretable models that elucidate the relationship between key variables and wind energy generation, aiding informed renewable energy policy decisions.

1.9. Significance of the Study

The significance of a study on "Wind Energy Prediction Using Long Short-Term Memory (LSTM)" lies in its pivotal contributions to the renewable energy landscape. By exploring advanced predictive models in the context of wind power generation, this study addresses critical gaps in the field. The utilization of LSTM models, alongside other forecasting methods, not only enhances the understanding of wind energy dynamics but also offers valuable insights into the efficacy of diverse modelling approaches.

This research bears significance in multiple dimensions. It contributes to advancing predictive accuracy in renewable energy forecasts, particularly in wind power generation, which is instrumental in energy planning and resource allocation. By dissecting the dataset's intricacies, such as missing values and anomalies, and applying meticulous data preprocessing techniques, this study sets a precedent for robust and reliable modelling in renewable energy studies. The comparison and evaluation of various models, including SARIMA, XG Boost, Random Forest Regressor, and LSTM, provide a comprehensive understanding of their strengths and limitations. This comparative analysis not only emphasizes the importance of selecting suitable modelling methodologies but also sheds light on the challenges in predicting wind energy outputs over extended periods. The study's exploration of the relationship between wind speed and power output, uncovering the sigmoidal function that governs their correlation, is a significant finding. This nuanced understanding of the relationship between these variables, derived through curve fitting and high R-squared values, contributes to the accuracy of short-term wind power generation projections. Such insights are crucial for effective energy management and resource allocation in the renewable energy sector (Wing et al., 2012).

This study's significance lies in its role as a pioneering effort in leveraging advanced predictive modelling techniques to enhance renewable energy forecasting, specifically in wind power generation. Its findings not only pave the way for more accurate predictions but also offer actionable insights for policymakers, energy planners, and researchers striving towards a more sustainable and efficient energy future.

2. LITERATURE REVIEW

The literature review within the domain of wind energy prediction using advanced modeling techniques represents a comprehensive exploration of prior research, methodologies, and findings aimed at forecasting power output from wind turbines. This critical analysis encapsulates a diverse array of studies, encompassing traditional statistical approaches and cutting-edge machine learning algorithms, which have been pivotal in advancing the understanding of wind power generation dynamics. Within this burgeoning field, scholars and researchers have extensively delved into the application of various predictive models to comprehend and forecast wind energy outputs. Traditional methods such as Seasonal Autoregressive Integrated Moving Average (SARIMA) models have been foundational in capturing seasonal trends within wind energy datasets. However, limitations in these approaches have spurred the exploration of more sophisticated techniques to tackle the complexity inherent in wind power prediction.

Recent advancements have witnessed a paradigm shift towards employing machine learning algorithms like Extreme Gradient Boost (XG Boost), Random Forest Regressor, and Long Short-Term Memory (LSTM) networks. These methodologies offer enhanced predictive capabilities, leveraging the intricacies of wind speed, temperature, and turbine-specific variables to improve forecast accuracy. This review amalgamates and critically evaluates these diverse methodologies, scrutinizing their efficacy, strengths, and limitations in predicting wind energy outputs. It highlights the successes of advanced machine learning techniques in capturing complex data dynamics, while also acknowledging challenges, such as handling missing values, anomalies, and the temporal aspect of wind power generation.

In Portugal, (Wang et al., 2016) describe a novel method for forecasting short-term wind energy. Particle swarm optimization (PSO), wavelet transform, and the adaptive network-driven fuzzy inference system (ANFIS) are the three distinct systems that make up this novel methodology. Their main goal is to improve wind power prediction accuracy, which is essential for the smooth functioning of wind energy systems and the grid integration of renewable energy sources. The higher performance of the suggested model in comparison to other comparable forecasting systems is one of the study's primary findings. Two commonly utilized error metrics, Normalized Mean Absolute Error (NMAE) and Mean Absolute Percentage Error (MAPE), are used to quantitatively illustrate this superiority. The outcomes show that the suggested model

regularly outperforms its alternatives in terms of NMAE and MAPE values. This suggests that the successful prediction error reduction provided by the integration of PSO, wavelet transform, and ANFIS in the forecasting process makes it a promising strategy for wind power forecasting. The study also considers how effectively their hybrid model can be computed. It is interesting that the average computational time needed for the proposed model remains tolerable despite the difficulty of integrating three different systems. This issue has practical implications since accurate and timely wind power projections are crucial for maintaining the grid's stability and managing the available energy supplies. The proposed approach's practical applicability and ability to retain a reasonable processing time while obtaining improved forecasting accuracy are highlighted.

The complete method presented by (Wang et al., 2011) aims to improve the precision of short-term wind speed forecasts. The creators proposed a crossover forecasting approach that consolidates the utilization of Long Short Term Memory (LSTM) networks for profound learning time series prediction with four unique modules: Wavelet changes (WT), Crow search calculation (CSA), common data (MI), and entropy-based highlight determination (FS). They utilized information from two geologically unmistakable regions, Sotavento in Galicia, Spain, and Kerman in the Center East, which is situated in the southeast of Iran, to direct their examination determined to assess the adequacy of this imaginative wind speed forecasting approach. Wavelet transforms, the first module, is a potent signal processing method used to split up time series data into various scales and frequencies. With the use of this decomposition, the wind speed data may be analyzed in greater detail, capturing both short- and long-term trends. The Crow search algorithm, the second module, is an optimization method that draws inspiration from crows' foraging habits.

It is used to enhance the forecasting model's overall performance and optimize its parameters. Mutual information, the third module, is employed for feature selection, assisting in the determination of the most pertinent input variables for the wind speed forecasting model. This stage is essential for lowering the data's dimensionality and increasing the forecasting process' effectiveness. To make sure that only the most insightful variables are employed in the prediction model, the fourth module, entropy-based feature selection, further refines the selection of input features. Memarzadeh et al. used actual wind speed data gathered from two different geographic places to assess the proposed method. These two locations-Kerman, Iran, and Sotavento, Spain-provide a thorough evaluation of the model's applicability to various climatic and geographic

contexts. Their study's numerical results showed that the hybrid forecasting method performed better than other wind speed forecasting methods, proving its superiority in predicting short-term wind speeds.

By utilizing a hybrid model known as Wave Net Long Short-Term Memory (WN-LSTM), (Wing et al., 2012) presented an innovative method for forecasting short-term wind power. This model uses many activation kernels, including Morelet, Gaussian, Shannon, and Ricker, and combines components of Wave Net and LSTM, two well-known neural network designs. The main objective of this research was to reduce the necessity for wavelet and gradient transformations in the non-linear mapping process while improving the accuracy of wind power estimates by utilizing this hybrid method. The authors used seven different wind farms throughout Europe to test the performance of their suggested WN-LSTM model. Using accepted criteria such the Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE), they assessed its efficacy. While the MAPE gives a percentage-wise evaluation of prediction accuracy, the MAE measures the average magnitude of errors between expected and actual values. According to the findings of their studies, the MAE showed a substantial percentage gain of up to 30% when compared to conventional forecasting methods. This shows that the WN-LSTM model fared better than the conventional approaches, highlighting its potential to enhance short-term wind power projections. The researchers independently ran their model numerous times to ensure its dependability and robustness, minimizing the possibility that random fluctuations would have an impact on the findings. The study also incorporated interval forecasting in addition to conventional point forecasts, assessing prediction uncertainty. Fisher's and Tukey's tests based on ANOVA (Analysis of Variance) were used to achieve this. With the help of the intervals, which added a degree of ambiguity to the forecasts, it was possible to gain a deeper knowledge of the potential discrepancies between the predicted and observed power outputs. Importantly, the study showed that the WN-LSTM model could provide reasonably accurate interval forecasts, with a comparatively low variance of about 0.02 at a 95 percent confidence level.

The development of a software-based computing model for precise forecasting of future demand in the context of renewable energy, particularly wind energy, was the main goal of (Xu et al., 2015) study. The authors noted that a combination of linear and non-linear methods was used in modern state-of-the-art forecasting systems. In order to maximize its use, the authors emphasized the crucial role that wind energy plays within the renewable energy industry and the necessity of precise prediction models. They

suggested using three different neural network-based models to handle this problem: the recurrent neural network (RNN), the gradient boosting machine (GBM), and the long short-term memory (LSTM). These neural network models were developed with the goal of predicting, using data on wind velocity, the power output produced by wind turbines. Their study's main goal was to evaluate the output parameter values of these three neural network models to compare how well they performed. This analysis sought to ascertain whether model RNN, GBM, or LSTM was better at predicting wind turbine power output. This comparative investigation offers important insights into the usefulness of neural network-based algorithms in renewable energy prediction, in addition to advancing wind energy forecasting.

A fascinating wind speed forecasting model using recurrent neural networks (RNNs), more precisely the Gated Recurrent Unit (GRU), is presented by (Yoon and Kun, 2013). In order to maximize resource allocation and system stability, wind speed forecasting is an essential component of the production of renewable energy and grid management. The goal of this study was to create a model for forecasting short-term wind speeds that is incredibly accurate. The researchers installed a specially made anemometer, most likely near a site of interest, and gathered wind speed data continually for the first six months in order to build their model. The GRU model was trained using these data as the basis. Being an RNN subtype, GRUs are well renowned for their effectiveness at capturing sequential patterns, making them appropriate for time-series data like wind speed. This study stands out for its emphasis on forecasting wind speed in brief 15-minute periods, which enables more accurate and prompt predictions. The following three 15-minute time interval forecasts were to be produced using the GRU model. For applications like wind energy generation and grid management, where prompt modifications are frequently required, this real-time forecasting is vital. The researchers used Mean Absolute Error (MAE) and Root Mean Square Error (RMSE), two widely used forecasting metrics, to assess the effectiveness of their GRU-based model. These indicators are common ways to evaluate how accurate a prediction model is. The researchers contrasted the results of their GRU model with those obtained from plain basic RNN and Long Short-Term Memory (LSTM) models in order to make relevant findings.

By fusing multiple cutting-edge methodologies, Zeng (Zeng and Qiao, 2011) present a novel strategy for enhancing short-term wind power forecasting. Their suggested model uses Sample Entropy (SE), an improved vibrational mode

decomposition (IVMD) technique, and an LSTM neural network with Correntropy enhancement as its foundation. Their main objective is to improve the precision and efficacy of wind power forecasts, especially in the near future. The authors first decompose the initial wind power data using improved variation mode decomposition (IVMD). IVMD is a method that breaks down complicated time series data into more manageable, comprehensible parts. In this instance, the best parameter K for IVMD is chosen using the Maximal Correntropy Criterion (MCC), which aids in the extraction of valuable subseries from the wind power data. Following the decomposition phase, the fragmented subseries are rebuilt using Sample Entropy (SE). The complexity or irregularity of time series data is quantified by SE. By capturing the underlying patterns and dependencies in the wind power data, the scientists hope to improve the forecast accuracy of their model by applying SE. The integration of MCC with the conventional Mean Squared Error (MSE) in the LSTM network is one of the study's major advances. By using the MCC in addition to the loss function, the LSTM network is made more resilient and adaptable to the unique properties of wind power data. This fusion of MCC and MSE aids in the creation of an original and reliable hybrid forecasting model for wind power. The authors used actual data from two wind farms in China for four evaluations to verify the efficacy of their suggested strategy. The results of these analyses, which were conducted at varied sample intervals, consistently showed that the suggested strategy beat the majority of established techniques for wind power forecasting. As a result, it appears that the IVMD, SE, and MCC-enhanced LSTM model combination has the potential to greatly enhance short-term wind power projections, which are essential for effective energy management in the context of renewable energy sources.

Using a hybrid forecasting model, (Righter, 1996) sought to enhance short-term wind energy projection. To improve the accuracy of wind power forecasts, this hybrid model integrated a number of techniques, including Convolutional Long Short-Term Memory networks (ConvLSTM), variational mode decomposition (VMD), and error analysis. The incorporation of these techniques into a thorough framework was one of their research's major accomplishments. Their strategy relied heavily on the VMD method, which divided the input wind power into various frequency components. The researchers were able to learn more about the underlying spatiotemporal patterns of the wind data thanks to this breakdown. The basic forecasting engine was built on top of this information. In essence, VMD acted as a stage in the pre-processing process that helped to extract significant features from the wind power time series data. The researchers

combined an LSTM network with a Convolutional layer to improve forecasting accuracy even more. Recurrent neural networks of the LSTM or Long Short-Term Memory variety are frequently employed in time series prediction challenges. The model improved its ability to capture both the geographical and temporal characteristics of the data by merging Convolutional and LSTM layers, making it well-suited for wind power forecasting. The individual subseries predicted by VMD provided the foundation for the initial forecasting results derived from this hybrid model. To get an overall forecast, this anticipated subseries was then combined. The researchers didn't stop there, either. They realized that wind power series have erratic features and needed to be improved. The study used LSTM to predict the variations in the first forecasting results in order to address this. The model was able to better capture and adjusts the abnormalities and fluctuations in the actual wind power series as a result of this step. LSTM was essentially utilized to model the variances from the initial projections in order to improve the predictions.

Enhancing Long Short-Term Memory (LSTM) network designs and their practical validation were the main topics of (Andrew and Zhe, 2010). Time series forecasting is one of the disciplines where LSTM networks have found widespread use. By incorporating sophisticated architectural adjustments, the scientists hoped to increase the precision with which these networks predicted upcoming occurrences. In their study, the scientists not only suggested new LSTM network topologies but also methodically assessed how well they performed in real-world scenarios. In order to evaluate the models' performance, real-world data has to be used. They hoped to close the gap between theoretical developments in neural networks and their actual application in predicting by doing this. Their study's investigation of methods for recalibrating the model is one of its standout features. When fresh data becomes available, traditional machine learning techniques frequently involve starting over with a completely new dataset to train the model. However, suggested a different strategy in which the parameters of the neural network are changed in response to fresh information learned over time. This method takes less time and enables the model to continuously adjust to changing circumstances, which eventually improves forecast accuracy. A case study done in Belgium was used to show the application of their research. The authors calculated the potential financial savings associated with increased forecast accuracy brought about by model recalibration. They discovered that the expenses related to system assessment might be decreased by recalibrating the model and improving the consistency of its forecasts. This shows that

investing in model recalibration can result in real savings and increased forecasting accuracy, both of which can be extremely beneficial in financial and economic contexts.

In order to overcome the difficulties of wind power forecasting, (Amjady, 2011) saw it as a non-linear multivariate function with hyper plane singularities and geographical in homogeneities. In the field of renewable energy, wind power forecasting is crucial because it permits effective integration of wind energy into power systems. The researchers decided to employ ridge lets as an efficient basic set to build the wind power function in order to address this challenging problem. Ridge lets are mathematical tools that can effectively capture spatial abnormalities and singularities in data, which makes them a good option for modeling wind power production. Recurrent neural networks (RNNs) were used by the researchers in their study to forecast wind energy. Their method was distinct since it used ridge functions as the initiation functions for the nodes in the RNN's hidden layers. The goal of this ridge function integration was to improve the neural network's capacity to represent the complicated and non-linear nature of wind power generation. The study article provided a unique stochastic search method called NDE (Natural Dual Estimation), which was used to train the suggested wind power forecast engine and establish the parameters of the model. The computational effectiveness of NDE and its need for a small number of samples during the training and validation phases were highlighted. In actual situations where processing resources and data may be constrained, this efficiency is essential. The adaptability of the suggested NDE technique is one of its significant advantages. The possibilities of discovering a worldwide optimum value for addressing the difficulties of wind power forecasting are increased since it enables a thorough search and analysis of prospective solutions from different research paths. This adaptability can help boost the precision of wind power estimates, which will be advantageous to both power systems and specific wind farms. The research effort carefully examined wind speed forecasting in addition to wind power forecasting, which is a crucial component of comprehending and properly utilizing wind energy. This study advances renewable energy systems and helps ensure the reliable integration of wind energy into the larger energy grid by addressing both wind power and wind speed predictions.

Ewea (2009) Developed a unique method for forecasting wind energy by using a discrete time-based Markov chain model. This study's main goal was to provide a more practical and effective method of calculating the distribution of wind energy, especially for short-term forecasting, without applying any constricting assumptions. The

importance of this research rests in its potential to increase wind power prediction precision, which is essential for maximizing wind energy grid integration and guaranteeing grid stability. The authors created a special model that used time series analytic methods to make it easier to estimate wind power distributions in order to accomplish their objectives. Their strategy aims to increase the application and reliability of wind power forecasting techniques by doing away with the requirement for constrictive assumptions. This is crucial when discussing renewable energy sources like wind power, which can be extremely volatile and difficult to anticipate with precision. The research report went into greater depth about the specifics of the suggested approach, offering a thorough explanation of both first- and second-order Markov chain models. A greater comprehension of the fundamental ideas and workings of the forecasting technique is made possible by this analytical investigation of the models. It is also a useful tool for forecasters of renewable energy who are both scholars and practitioners. Carpinone et al. used their suggested method to real-time wind power data as part of their research's conclusion to show how it may be used in practice. This empirical validation demonstrated both the efficiency of their method and its potential to promote the integration of renewable energy sources and grid management.

Support Vector Machine (SVM) regression was used in a novel way for wind power forecasting by (Frandsen et al., 2006). This study aims to evaluate SVM's performance in forecasting wind power output and compare it to other forecasting techniques. The authors came at numerous important results through extensive simulations. Zeng and colleagues discovered during their experiment that the SVM-driven regression model had impressive accuracy in predicting wind power output. The tight agreement between the predicted and expected values proved the model's accuracy and dependability. The results also demonstrated the durability of the SVM model in capturing intricate patterns in wind data by demonstrating how well it captured the anticipated changes in wind power. This study's comparison of the SVM model with the RBF-neural network-driven model and a persistence model was one of its most important contributions. With a predictive horizon of roughly 16 hours, the SVM model surpassed both of these models in terms of short-term wind power forecasting, showing an amazing improvement of more than 26%. This result highlighted SVM's potential as a useful tool for improving short-term wind power estimates. The study also demonstrated the SVM model's advantage over other diligence methods in terms of forecasting wind power accuracy. It was shown that the SVM-driven approach could produce more precise

forecasts, which is essential for maximizing the grid integration of wind energy and enhancing the overall effectiveness of wind power generation. The study did highlight one drawback of the suggested SVM model, though. The historical data lost correlation with the current wind power conditions as the forecast horizon grew. This constraint indicates that additional meteorological factors, like as pressure, temperature, and others, may need to be included in the model in order to produce reliable 24-hour wind power projections. The accuracy of long-term wind power forecasting could be improved by combining these factors with Numerical Weather Prediction (NWP) data, thereby overcoming the difficulties brought on by the expanding prediction horizon.

For the investigation of wind power projections in the Portuguese system, a unique methodology known as the Hybrid Evolutionary Adaptive (HEA) methodology was presented by (Jensen, 1983). This methodology's main goal was to give three-hour wind power estimates with 15-minute intervals, which was its core focus. This method was created to improve the precision and robustness of wind power forecasting while addressing the inherent difficulties presented by non-stationary data sets. To accomplish its goals, the HEA methodology incorporated a number of models and techniques. It combined features from the Wavelet Transform (WT) model to take advantage of filtering effects, the Evolutionary Particle Swarm Optimization (EPSO) model for evolutionary optimization, the Adaptive Neuro-Fuzzy Inference System (ANFIS) model to implement an adaptive architecture, and the Model Identification (MI) model to choose input data and improve overall robustness. This multi-model method made it possible to analyze wind power predictions in detail while taking into account various aspects of data processing and optimization. The research work selected test cases that were comparable to those used by other approaches in order to produce a transparent and precise comparative analysis, allowing for a meaningful comparison. In order to concentrate only on the fundamental elements of wind power forecasting, exogenic variables were purposefully ignored. Results from the HEA methodology were said to be extremely precise and effective in lowering predicting uncertainty for wind energy. Notably, this method's Mean Absolute Percentage Error (MAPE) score of 3.75% demonstrated a high degree of prediction accuracy. Further demonstrating the efficacy of the methodology, the average error variance and Normalized Root Mean Square Error (NRMSE) were discovered to be 0.0013 and 2.66%, respectively. The HEA methodology's capacity to lessen computing complexity was one of its key advantages. Without sacrificing the precision of the outcomes, it was able to produce real-time wind power projections in less

than 40 seconds per iteration. For practical applications where precise and timely forecasts are important, this reduction in computing time is crucial.

A thorough description of the statistical techniques used by the ANEMOS project for forecasting short-term wind power was presented by (Katic et al., 1986). The difficulties in estimating wind power, particularly in the setting of wind farms, were addressed in large part by this research. The forecasting procedure included a number of discrete processes, each of which added to the precision and dependability of wind power estimates as a whole. The meteorological (MET) forecasts were scaled down as one of the first steps in this process to fit them with the unique circumstances of the potential wind farm. This downscaling procedure was essential for modifying the general MET forecasts to the specific characteristics of the wind farm, improving the precision of power forecasts. The researchers then made use of wind power curves created from previously collected data from the wind farm. The relationship between wind speed and the accompanying power output was shown by these curves. Since these curves took into consideration the particular performance traits of the wind turbines in the farm, they allowed for more accurate estimates of power production. Another essential element of the forecasting strategy used by the ANEMOS project was dynamic models. Using these models, which took into consideration the dynamic nature of wind conditions, it was possible to predict changes in wind speed and power production over time. The researchers sought to produce short-term forecasts that were accurate and responsive to shifting environmental conditions by include these dynamic components. The study also stressed the significance of uncertainty estimation. It was crucial to put a number on how unreliable the predictions were given the inherent diversity of wind patterns. This made it possible to offer probabilistic forecasts, which could help with grid management and energy market decision-making. Last but not least, the ANEMOS project tackled the issue of scaling up the forecasts to determine the whole regional wind power generation. Data from a small number of wind farms that were used as reference locations had to be combined for this stage. The researchers attempted to give a precise evaluation of the overall wind energy generation in the area by using proper scaling methodologies.

In especially for short-term horizons, (Ishihara et al., 2004) make a significant addition to the field of wind power prediction. Accurate forecasting of wind performance and accompanying electric energy generation has emerged as a critical component of assuring grid reliability and stability as a result of the rapid growth in wind power integration within power systems. The authors addressed the urgent requirement for

increased forecast accuracy by introducing a brand-new and incredibly powerful hybrid Wind Power Forecasting (WNF) technique. The capacity of the suggested strategy to attain an average Mean Absolute Percentage Error (MAPE) of roughly 5.99% is one of its significant accomplishments. For grid operators, energy market participants, and policymakers, this denotes a remarkable degree of precision in projecting wind power generation. The low MAPE shows that the model's forecasts closely match real wind power output, lowering operational uncertainty in the power system. Furthermore, it is important to emphasize how effective the suggested strategy is in terms of computing. According to the authors, the hybrid WNF approach's average computation time is under a minute. This quick processing time is a big plus, especially in operational environments where decisions must be made in real-time or almost real-time. It illustrates that the model provides efficient forecasts in addition to being accurate, making it an effective option for predicting wind generation. The hybrid WNF method outperforms a number of current approaches, including ARIMA (Autoregressive Integrated Moving Average), NNWT (Neural Network with Wavelet Transform), persistence models, and conventional neural network (NN) methods, according to the study's comparative analysis. This indicates the proposed method's superiority in terms of precision and computing effectiveness. According to the study's findings, the hybrid WNF strategy put forth by (Venayagamoorthy et al., 2012). Is a promising method to resolving the mounting problems caused by the integration of wind power into power systems.

By incorporating intelligent and hybrid pattern recognition technologies, (Werle, 2008) provide a novel approach for wind power forecasting. Variational Mode Decomposition (VMD), a well-known signal processing method, is one of the primary methods used in this study. To perform time-series decomposition on the wind power data, a vital step in comprehending the underlying patterns and trends, is the main goal of using VMD. The researchers have implemented a unique feature selection strategy powered by gravitational search optimization (GSO) to increase the effectiveness and interpretability of their forecasting model. This feature selection method tries to remove useless data, hence lowering memory needs and improving the forecasting device's overall efficiency. The study uses Extreme Learning Machine (ELM) in addition to VMD and the GSO-driven feature selection model to create the connection between the desired projected output and exemplar patterns. ELM is a machine learning method that is well-known for being quick and straightforward, making it a good fit for this application. The structure of the researchers' forecasting model is also adjusted using the cross-validation

method. In addition to ensuring that the model is accuracy-optimized, this phase also provide a way to assess how well it is working. In order to determine the most efficient arrangement, the study thoroughly analyzes numerous selected attributes and decomposition mechanisms. According to the simulation results, the forecasting model that chooses 20 features and ten different decomposition types performs better and makes less forecasting errors. Surprisingly, this improved performance is seen for forecasting periods that are both short-term (1 hour) and very short-term (10 minutes). The researchers used historical wind power data collected from twelve different wind farms to apply the suggested model to validate it. The results of this real-world application demonstrate how well the model predicts wind power generation, highlighting its potential use in the renewable energy market.

In order to improve the accuracy of wind power estimates for short-term horizons, (Crasto and Gravdahl, 2008) presented an innovative technique. They employed NWP-data correction models as part of their methodology in an effort to correct inaccuracies in the Numerical Weather Prediction (NWP) data. Using several data mining approaches, the authors used their suggested model to find and classify mistakes in the NWP data. The raw and anomalous NWP data was then adjusted and standardized before being sent into the Wind Power Forecasting (WPF) engine. One of the noteworthy accomplishments emphasized in this study is the applicability of their model, which significantly reduced Wind Forecasting (WFO) mistakes. This result highlights how their method can be used in practice to increase the accuracy of wind power estimates. Nevertheless, despite the enthusiasm for this development, a number of problems need to be taken into account. First of all, the study falls short of providing a thorough analysis of the root reasons of the various inaccuracy patterns found in the NWP data. To successfully address these mistake causes and further improve the performance of the model, it is essential to understand them. It is difficult to execute targeted adjustments without a detailed explanation of the causes of these mistakes. Second, the study heavily relies on a data-driven strategy without giving a detailed explanation of the methodology or model architecture used. This lack of transparency may prevent the suggested algorithm from being widely accepted and applied by the scientific and practical communities. A thorough framework and model description that enables greater comprehension and reproducibility is crucial for the model to be broadly accepted and incorporated into practical applications. Last but not least, the neural network module is the only one used to choose the threshold for AWPF (Wind Power Forecasting), which may not be the most

reliable or adaptive method. Establishing more reliable and adaptive criteria for choosing this threshold is essential to guaranteeing the suggested method's dependability in real-world circumstances. This will increase the model's ability to adjust to changing circumstances and boost its effectiveness in real-world wind power forecasting applications.

Crespo et al. (1999) has provided a thorough review of the benefits and drawbacks of incorporating renewable energy sources into our energy infrastructure. The growing acknowledgement of renewable energy as an essential component of future energy generation has given this subject a great deal of recent attention. The switch to renewable energy sources has become urgently necessary due to the persistent global energy crisis and the environmental issues connected to traditional fossil fuels. The environmental sustainability of renewable resources is one of the main benefits noted in the literature. Renewable energy sources, such as solar, wind, and hydroelectric power, produce little to no greenhouse gas emissions in contrast to fossil fuels, making them crucial for reducing global warming. Additionally, the almost limitless supply of these resources lessens our reliance on limited fossil fuel reserves. Because it diversifies the energy mix and lessens reliance on geopolitically unpredictable regions for energy supply, the switch to renewable energy also helps to ensure energy security. Renewable resource integration does present certain difficulties, though. Numerous drawbacks are mentioned by Ayadi et al., including intermittency and variability. The production of renewable energy is reliant on fluctuating natural elements like sunshine and wind. In order to maintain a steady supply of electricity, this intermittency can compromise the stability of the energy grid, necessitating the development of advanced energy storage devices and grid management techniques. The focus of current research is changing to the incorporation of renewables into smart grids in order to address these issues and realize the full potential of renewable energy. To improve energy management, smart grids include cutting-edge control systems, real-time monitoring, and communication technology. With this strategy, renewable energy can be distributed effectively, grid instability problems are reduced, and extra energy can be stored or sent to areas where it is most required. The incorporation of renewable resources into smart grids is emerging as a crucial study field with the potential to change our energy systems and provide a more sustainable energy future as the globe struggles with the need for sustainable energy solutions and persistent energy shortages.

The significance of binomial energy management in successfully managing power flows within a PV-Wind hybrid power system is examined by (Frandsen, 1992). In their study, they proposed a novel strategy for creating an energy management system designed for this particular objective using fuzzy logic. They used the SIMULINK-MATLAB environment, which allowed them to build and test a reliable simulator for the suggested hybrid power system, to validate their methodology. They conducted a number of tests utilizing actual meteorological data gathered from the Adrar region as part of their research, which was a crucial component. This information was used to assess how well their system performed. Additionally, they made a comparison between the simulation results and a determined realistic load demand pattern. Their conclusions were based in applicability and relevance to situations seen in the real world thanks to the rigorous methodology they used. These experiments had very good results. The simulation findings of the study showed that their "coupled approach" to energy management surpassed conventional power management techniques significantly. This shows that their fuzzy logic-based energy management system may be more able than traditional approaches to optimize power flows in a PV-Wind hybrid system.

Barthelmie et al., (2006) Focused on a short-term forecasting horizon of just 24 hours to address the urgent demand for precise and fast predictions of wind farm reducibility. Due to the fact that it enables energy management to efficiently plan and optimizes their resources, this timeframe is crucial for the effective operation and integration of wind energy into the grid. The researchers used feed-forward artificial neural networks, a machine learning method renowned for its capacity to predict intricate, nonlinear relationships in data, to accomplish this goal. The creation of a variety of prediction models, each adapted to a particular facet of wind farm performance, was a crucial component of their research. The objective of these models was to offer trustworthy predictions of wind power output within a 24-hour window. The meticulous process used to identify the ideal design for each neural network model is what distinguishes this study from others. The researchers used a simulation model that includes systematically changing the crucial artificial neural network parameters rather than relying exclusively on theoretical considerations. They were able to optimize the neural network architectures using this empirical method to get the most accurate wind power estimates. The researchers compared the results with forecasts produced by numerical weather prediction (NWP) models in order to evaluate the effectiveness of their neural network-based prediction models. The effectiveness of alternative prediction

techniques is measured against the performance of NWP models, which are frequently used for forecasting weather and wind energy. Dolara and his colleagues were able to validate the dependability and accuracy of their artificial neural network models for forecasting wind farm producibility through the use of this comparison.

Mechali et al. (2006) proposed a fresh and creative method for making wind energy forecasts. Aguilar and his team established a thorough probability-based forecasting method, which differs from the majority of prior research in the field, which mostly concentrated on point forecasts for wind speed and subsequently produced energy level estimates using wind farm power curves. Their strategy included wind power projections for each forecasted wind speed at different lead times. This change in approach was made to address the inherent uncertainty in wind speed forecasts, which is a crucial component of wind energy generation forecasting. In the past, wind energy projections mainly depended on point estimates of wind speed that were deterministic, ignoring the probabilistic nature of wind speed changes. The Double Seasonal Holt Winters model and conditional density kernel estimation were combined with time series approaches by Aguilar et al., in contrast, to produce complete probability-based forecasts. They were able to produce a complete probability distribution for wind power at various wind speeds and lead periods using this method rather than just a single point estimate. This method's adaptability to real-world situations was one of its key benefits. The authors used data from a real wind farm in Brazil to validate their methods. Positive outcomes from this empirical examination proved the practical viability and efficacy of their suggested strategy. Aguilar and his team provided a more robust and trustworthy method for wind energy forecasting, which is essential for optimizing the operation and integration of wind energy into the grid, by taking into consideration the uncertainty associated with wind speed estimates.

This technique was essential in minimizing restrictions like turbine radius and inter-turbine distances while also maximizing the wind farm's overall energy output. Kusiak and his team attempted to reconcile the requirements of safety and turbine spacing with the spatial arrangement of the turbines, which impacts wake losses and energy capture. They did this by employing a multi-objective optimization framework. The authors also offered possible directions for future research, highlighting the significance of taking various terrain heights and wind turbine specifications into account. These suggested additions demonstrate how flexible and scalable their concept is, since it can be adjusted to fit a variety of geographic areas and wind farm layouts. Researchers and

wind farm developers can make well-informed choices to improve energy generation while taking into consideration site-specific limits and characteristics by accounting for various terrain heights and turbine specs.

The complex trade-off between energy generation and noise reduction in the design of wind farm layouts is explored in depth by (Wing Yin Kwong, 2012). Their main goal was to give engineers useful information for the creation of efficient design processes that balance these two crucial criteria. The researchers used genetic algorithms to assess populations of potential solutions in order to accomplish this. The authors looked at single and multi-objective optimization situations for wind farm layout optimization. They started by thinking about how to maximize energy production. It is crucial to create layouts that can wring the most energy from the prevailing wind patterns because wind farms are primarily intended to harvest wind energy. In addition to serving economic interests, increasing energy production also advances the development of renewable energy, addressing sustainability issues. The study also focused on the equally crucial goal of reducing noise levels near the wind farm's edge. Due to the fact that wind farms are frequently situated in rural regions, noise pollution is a rising concern in the area. Keeping noise levels to a minimum is crucial for the social and environmental acceptability of wind farms. The authors sought to achieve a compromise between effectively utilizing wind energy and limiting negative effects on nearby residents and ecosystems by reducing noise at the boundary. The study examined a wide range of potential wind farm layouts using Genetic Algorithms, a potent optimization tool inspired by natural selection. The layouts that represent trade-offs between energy generation and noise reduction were able to be identified thanks to this method. The study gave engineers and other stakeholders a thorough grasp of the design space by looking at a variety of possibilities, enabling them to make deft choices.

The paper covers a variety of forecasting techniques, including neural networks, the Adaptive Neuro-Fuzzy Inference System (ANFIS), Computational Fluid Dynamics (CFD), and numerical weather prediction (NWP). The performance and applicability of these methodologies in the analysis of various wind farms are examined in the paper. ARMA modeling is one of the main methods included in the literature review. When analyzing time series data to predict wind speeds, ARMA models are frequently utilized. In order to capture temporal relationships and fluctuations in wind data, they offer a statistical methodology. The review highlights the applicability of ARMA models in predicting wind power generation and examines their advantages and disadvantages in

wind forecasting. Another important part of the review is computational fluid dynamics (CFD). At wind farm locations, CFD simulations are used to examine the turbulence and wind flow patterns. This method makes it possible to comprehend the intricate aerodynamic interactions between wind turbines in great detail. The literature study examines how CFD modeling helps to enhance the design of wind farms and optimize the location of turbines for greater energy output. A crucial part of wind forecasting is numerical weather prediction (NWP). The usage of NWP models, which are complex mathematical representations of atmospheric conditions, is extensively discussed in the paper. NWP models provide useful insights into wind patterns, enabling more precise forecasts of wind direction and speed. In order to improve the predictive capacities of wind forecasting systems, the research investigates how NWP data might be integrated into them. The review also emphasizes the function of artificial intelligence methods in wind prediction, including neural networks and the Adaptive Neuro-Fuzzy Inference System (ANFIS). In order to estimate the future of the wind, neural networks may understand intricate relationships within the data. ANFIS develops a hybrid model that can handle nonlinear and unpredictable data by fusing fuzzy logic and neural networks. The review covers the benefits of these AI-based methods and how they can be used to forecast wind speed and output.

To simulate several scenarios and assess the potential repercussions of establishing such a renewable energy plant in the coastal region, their research used the SWAN (Simulating Waves Nearshore) ghostly prototype model. The study has several facets and concentrated on three main goals. The research's first goal was to look into the topographical and geographical implications of what the hybrid farm would do to the coastal environment. This involves examining potential changes to the shoreline and underwater topography that could result from the construction and operation of the hybrid wave-wind farm. It was essential to comprehend these changes in order to evaluate the project's possible ecological and environmental effects. Coastal regions are extremely sensitive to changes in wave patterns, currents, and sediment transport; therefore, the second important component of the study was to investigate the effect of the hybrid farm on the dynamic forces that influence the shoreline. Therefore, the researchers sought to quantify how the wave-wind farm's existence may alter these dynamic forces. This knowledge was essential for forecasting the project's potential effects on coastal stability, sedimentation, and erosion. Examining the variations in wave characteristics and wave energy distribution close to the proposed wave-wind farm was the study's third goal. The

researchers specifically looked into how the hybrid farm's presence changed wave heights, directions, and frequencies. It was essential to comprehend these changes in order to assess the viability and effectiveness of the wave energy conversion technology used in the hybrid system. They covered topics such turbine blade design, wind power characteristics, and estimates for output power. One of the major outputs of their research was the creation of a 1 kW, 1 m diameter wind turbine that was painstakingly developed with the aid of specialist software tools. The study went deep into assessing the performance of this turbine design from several perspectives in an effort to efficiently harness wind energy. The study of the turbine blade design was one crucial area covered in the research. To evaluate the power production and efficiency of the turbine blades at various tip-speed ratios, Sarkar et al. carried out comprehensive testing. These tests offered insightful information about how various turbine blade arrangements affected the wind turbine's overall performance. The researchers also performed calculations using software tools to support their experimental results, providing a thorough understanding of the turbine's potential. The study also considered a number of environmental aspects that might affect the functioning of the turbine. These included the wind direction, which has a direct impact on how well the turbine can catch wind energy. In addition, issues like corrosion, water vapor infiltration, thermal expansion, mechanical load considerations, and the effects of seasonal climate changes between summer and winter were covered in the research. In order to create a more realistic and practical design that could endure actual climatic conditions, it was essential to comprehend how these environmental factors interacted with the turbine's component parts. Discussions about the aging of wind turbine parts were a key component of the research. Long-term operation of wind turbines can cause component wear and degradation, which can reduce overall performance. These aging mechanisms were examined by Sarkar and his team, who also suggested methods for reducing their negative consequences. This component of the study was crucial in assuring that the wind turbine's design would be durable and dependable over its operating lifespan in addition to being effective at the time of construction.

The authors also emphasize their plans for future work to enhance forecasting models for photovoltaic (PV) energy sources. This implies expanding the scope of the integration of renewable energy with the aim of improving prediction algorithms to incorporate other renewable sources like solar energy. A step closer to a greener future would be made possible by the eventual integration of such enhanced forecasting models

into actual energy operations. This would increase the effectiveness and sustainability of the use of renewable energy sources.

The field of monitoring the state of wind turbines holds tremendous promise for copula analysis, as discussed in their research. Copulas are statistical tools that enable the modeling of the structure of variable dependence. Copula analysis can be used in the context of wind turbines to analyze and quantify the correlations between different parameters, such as wind speed and active power output. Given the nonlinear and non-stationary nature of data from wind turbines, this is especially crucial. The presenting of an example featuring the use of copulas to assess the intricate correlations between wind speed and active power output was one of their study's major achievements. They used copulas to capture and comprehend the complex relationships between these factors, which might change dramatically over time as a result of shifting environmental conditions. This showed how copula analysis could be a useful method for monitoring the status of wind turbines. Additionally, the results showed that copula analysis might be applied to more intricate situations. To examine the interdependencies and correlations among various wind turbines operating close to one another, the evaluation can, for example, include copulas for numerous turbines. Copulas could also make it easier to translate data from one turbine while taking alternate outcomes into account, allowing for a thorough analysis of data discrepancy at various manufacturing altitudes.

The study on Wind Energy Prediction Using Long Short-Term Memory (LSTM) marks a significant leap in the realm of renewable energy forecasting, distinguishing itself from past studies through various noteworthy aspects. Comparisons with prior research reveal key advancements and methodological differences, contributing to the evolution of wind power prediction methodologies. Past studies predominantly relied on conventional statistical models like Seasonal Autoregressive Integrated Moving Average (SARIMA) to forecast wind energy outputs. However, the current study diverges by incorporating sophisticated machine learning algorithms such as LSTM networks alongside ensemble methods like Extreme Gradient Boost (XG Boost) and Random Forest Regressor. This departure signifies a shift towards leveraging the robustness of deep learning architectures to capture intricate temporal relationships in wind data. Prior studies focused on addressing seasonal trends and basic correlations between wind speed and power output, the present research extends its scope. It delves deeper into the nuances of wind turbine behaviour, scrutinizing the impact of variables like ambient temperature, blade pitch angles, gearbox conditions, and generator RPM on power generation. This

comprehensive analysis expands the understanding of multifaceted factors influencing wind energy, thereby enhancing prediction accuracy.

Past endeavours often encountered challenges related to missing data, outliers, and the temporal aspect of wind power generation. In contrast, the current study demonstrates a meticulous approach to data preprocessing, emphasizing the careful management of missing values, detection, and exclusion of anomalous data points. Additionally, it incorporates visualization and statistical analysis to comprehend dataset properties, ensuring a robust foundation for modelling. Moreover, previous studies primarily focused on short-term wind power forecasting. In contrast, the present study extends its scope to incorporate long-range dependencies in sequential wind data through LSTM networks. This facilitates a deeper exploration of wind speed-power generation correlations, even though the LSTM model exhibited challenges in effectively capturing these complex dynamics. The current study stands out by embracing a holistic approach, amalgamating advanced machine learning techniques, meticulous data treatment, and an in-depth analysis of various parameters affecting wind energy generation. This departure from traditional methodologies demonstrates a significant stride towards more accurate and comprehensive wind power predictions, offering valuable insights for sustainable energy management and decision-making.

The literature review encompasses diverse studies on wind energy, emphasizing novel approaches to wind forecasting, wind farm optimization, grid integration, and environmental impact assessment. Aguilar (Aguilar et al., 2014) proposed a novel probability-based forecasting method, departing from traditional deterministic models, to account for uncertainty in wind speed estimates, validating their approach using real wind farm data. Harrouz (Harrouz et al., 2019) Had underscored the importance of accurate predictive models for maximizing wind energy utilization while addressing the unpredictability of wind patterns. Banna (Banna, 2014) investigated grid placement's influence on wind farm stability, concluding that increased wind energy penetration enhances grid stability. Kusiak (Kusiak et al., 2010) has introduced a multi-objective optimization model for wind farm layout design, optimizing turbine placement for maximum energy output while considering wake losses. Wing (Wing Yin Kwong, 2012) addressed the energy-generation versus noise-reduction trade-off in wind farm layouts, using genetic algorithms to identify layouts balancing these factors. Fugon (Fugon, 2008) focused on short-term wind power forecasting, highlighting Random Forest's superiority in forecasting accuracy for wind power integration. Murali (Murali, 2014) explored

feasible offshore wind farm locations along the Indian coast, considering technological, environmental, and financial aspects. Rakeshchandra (Rakeshchandra, 2013) reviewed forecasting methods, including ARMA modelling, CFD simulations, NWP, and AI-based approaches, to aid wind farm operators in estimating wind speeds. Diacon Sorin (Diacon Sorin, 2013) evaluated the coastal effects of a proposed hybrid wave-wind farm, assessing topographical changes, dynamic forces on the shoreline, and variations in wave characteristics. These studies collectively offer insights into diverse facets of wind energy, ranging from forecasting techniques to environmental impact assessments, aiming to optimize wind energy utilization and grid integration while considering environmental implications by (Ali Abdulrahman Salihi and Merdin Danişmaz, 2023).

3. MATERIAL AND METHOD

3.1. Dataset Description

The research methodology is inherently rooted in Python as the primary tool for conducting the entire analysis. Python's versatility and extensive libraries make it an ideal choice for handling the intricate aspects of data manipulation, exploratory data analysis (EDA), and modeling.

Initially, Python facilitates the crucial phase of data processing. Through various Python libraries, the dataset undergoes meticulous cleaning processes. Missing values are addressed, outliers are handled, and any redundant or irrelevant data points are removed. This preparatory step ensures the dataset's integrity and quality, setting the stage for subsequent analyses. The subsequent stage delves into exploratory data analysis, leveraging Python's capabilities to unveil key insights within the dataset. Python's libraries, such as Pandas, Matplotlib, and Seaborn, are instrumental in visualizing distributions, correlations, and identifying anomalies. These visual representations aid in comprehending the dataset's underlying characteristics, relationships, and potential patterns. Following EDA, the methodology progresses toward predictive modeling using a neural network. Python's Tensor Flow or Keras libraries are often employed to construct and train the neural network model. The network's architecture, training parameters, and optimization techniques are carefully configured within Python, enabling the model to learn and extract complex patterns inherent in the dataset. The pinnacle of the methodology is encapsulated within an interactive Python file (.ipynb). This file acts as a comprehensive repository, housing not only the executable code but also detailed documentation and interpretations of the neural network model's performance and insights derived from the analysis. This interactive file serves as a transparent and accessible platform, allowing readers to explore the research process, comprehend the model's behavior, and understand the obtained results thoroughly. Throughout this methodological journey, Python serves as the cornerstone, offering a rich ecosystem of tools and functionalities. Its seamless integration across data processing, exploratory analysis, modeling, and documentation ensures a robust and transparent framework for conducting simulations, analyses, and presenting findings.

The CSV file utilized in the thesis served as the primary source of data, housing a structured dataset in comma-separated values format. This file contained multifaceted information crucial for the research, capturing various parameters related to the operation

of a system or machinery. Among the vital variables extracted from this file, "active power" stood as a pivotal metric. It represented the actual power consumed or produced within the system, reflecting its active energy usage or generation. Additionally, the "df index" referred to the indexing or referencing mechanism within the data frame, enabling swift access to specific data points or subsets for analysis. Another significant variable, "angle blades," delineated the angular positioning of individual blades, possibly indicating the rotational behavior or alignment of specific components within the system. The CSV file, alongside these variables and others extracted from it, formed the fundamental dataset driving the analysis, enabling insights into operational patterns, energy usage, and machinery behavior within the studied system.

The data utilized in this section has been sourced from the International Energy Agency (IEA) through the Wind Power Association. The Wind Power Association serves as a primary disseminator of this data, making it accessible to researchers and corporations globally.

This study is carried out in Turkey, and it is drawing its ideas from the wind turbines concept that has been so successfully adopted in Türkiye. The primary purpose is to extract data from the CSV file in order to make a prediction about the power output that will occur over the next 15 days. To achieve this goal, a wide variety of forecasting techniques are utilized, beginning with the application of SARIMA (Seasonal Autoregressive Integrated Moving Average). This is done mostly as a result of the obvious seasonal trends that can be seen within the dataset. However, due to the poor performance of SARIMA, the research turns its focus to more effective machine learning algorithms, such as XG Boost and Random Forest Regressor, which produce better results. This was done in order to improve the accuracy of the study. In addition, the study investigates the use of LSTM, which stands for long short-term memory. This is a sort of recurrent neural network, but unfortunately, it does not generate the results that are wanted. Throughout the entirety of the process of analysis, the study makes heavy use of common data exploration tools, such as visualization and statistical analysis, in order to get a thorough comprehension of the dataset's innate properties. This multi-faceted strategy has been carefully constructed to determine the most accurate way for estimating power output within the particular context of Iraq. It does this by taking ideas from the successful wind turbines model in Türkiye in order to boost the forecasting skills for the setting of Turkey.

3.2. Introducing the Data to the Paython Program

To use this data correctly, we need to make sure the information in the file is seen as dates. We do this by telling the system that the "Unnamed:0" column is about dates. This helps us work with the data in terms of time, which is really important for tasks like forecasting or understanding how things change over time.

First, I checked if there were any entries that appeared more than once in the dataset. Turns out, there were 23,039 records that were exact duplicates, meaning they showed up more than once. To make things neat and clear, I got rid of these extra copies from the data. After doing this, we're left with 95,185 unique entries. These unique entries are all different and not repeated, which is super important for us to get accurate insights from the data. Removing these duplicates is a common way to clean up data, making it better for analysis and giving us more trustworthy information to work with.

3.2.1. Application of exploratory data analysis using pandas profiling and then some boxplots

The analysis of the dataset begins by examining the distributions of each variable. This exploration is facilitated by utilizing the pandas profiling package, which not only reveals the data distributions but also highlights missing values within the dataset. During this examination, it becomes evident that the variables "Control Box Temperature" and "WTG" do not provide meaningful information and may not be useful for the analysis, prompting their potential exclusion.

Additionally, a correlation graph is initially presented, but further interpretation is reserved for a later stage of the analysis. Another important step in data cleaning involves checking for and handling duplicate entries. If duplicate rows are found in the dataset, they are removed to ensure the data is free from redundancy.

As a noteworthy observation is made about the "Active Power" variable. It's discovered that the box plot for this variable includes values below zero, suggesting the presence of negative power values, which can be unusual. To maintain data integrity and relevance, these negative power values are proposed for removal from the dataset. The initial stages of data analysis involve examining variable distributions, identifying less useful variables, addressing missing values, exploring correlations, and checking for and removing duplicate entries. Additionally, a decision is made to exclude negative values in the "Active Power" variable, which might require further investigation.

3.3. Data Overview

The dataset comprises 22 variables and encompasses a total of 95,185 observations. Among these observations, approximately 27.2% of the cells contain missing data, totalling 570,424 missing entries. There are no duplicate rows within the dataset. In terms of memory usage, the entire dataset occupies 16.0 MiB, with an average record size of 176.0 bytes.

The variables in the dataset fall into three main types. There is one variable categorized as Date Time, capturing temporal information. Additionally, there are 19 numeric variables that likely represent various measurements or numerical data points. Lastly, two variables are categorized as categorical, implying they contain distinct categories or labels rather than continuous numerical values.

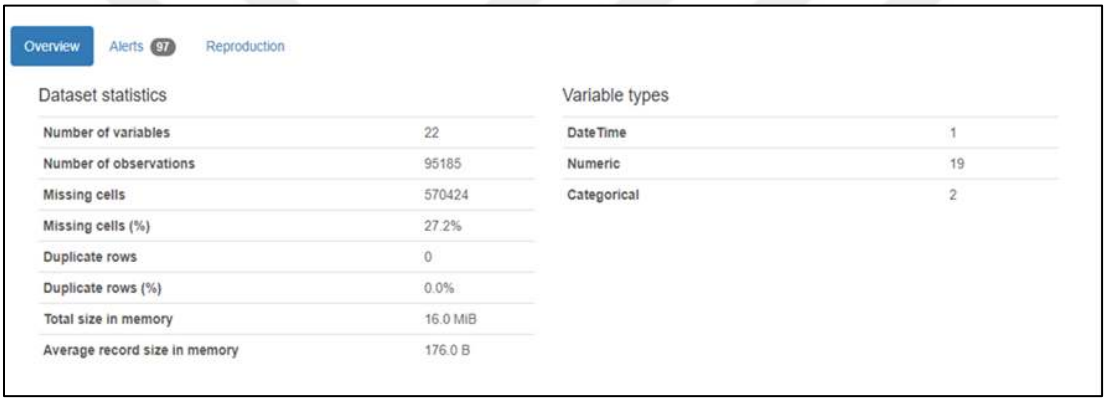

Figure 3.1. Dataset statistics

3.3.1. Variables

The "Date" variable, also referred to as "df_index," represents a chronological timeline in the dataset (Fig. 3.2). Here's an explanation of this specific variable:

The "Date" variable contains 95,185 distinct values, accounting for 100.0% of the observations in the dataset. This means that each observation has a unique date associated with it, and there are no duplicate dates. There are no missing values in this variable, indicating that all 95,185 observations have a valid date timestamp, resulting in a missing percentage of 0.0%. The memory size occupied by this "Date" variable is approximately 743.8 KiB, which is the amount of computer memory needed to store and process this data efficiently. The minimum date in this variable is "2017-12-31 00:00:00+00:00," which signifies the earliest date and time in the dataset.

The maximum date in this variable is "2020-03-30 23:50:00+00:00," indicating the latest date and time recorded in the dataset. In summary, the "Date" variable in this

dataset represents a time series spanning from December 31, 2017, to March 30, 2020, with no missing values and a memory size of approximately 743.8 KiB. It provides a chronological reference for the dataset's observations over this time period.

Figure 3.2. df_index

The "Active Power" represents the real power output of a wind turbine, which is the portion of generated power that can perform work and contribute to the overall electrical energy supply. Active power variable is a real number (belonging to the set of real numbers, $\mathbb{R}$) within this dataset. It exhibits a high correlation with other variables in the dataset, indicating that it may be closely related to or influenced by other factors under study. There are 94,084 distinct values for the "Active Power" variable, which accounts for approximately 99.4% of the observations in the dataset. This suggests that the majority of the data points for this variable are unique, reflecting a wide range of active power values.

In terms of missing data, there are 561 missing values for "Active Power," constituting a relatively small percentage of 0.6%. This indicates that most of the observations have valid entries for this variable, although a small proportion is missing.

There are no infinite values recorded for "Active Power," meaning that all values fall within a finite range of real numbers.

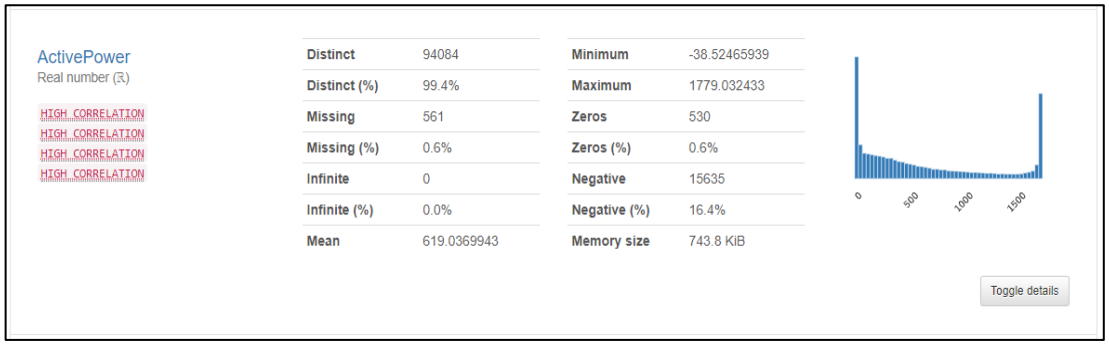

Figure 3.3. Active power

The "Ambient Temperature" is a critical environmental parameter influencing the efficiency and performance of wind turbines, affecting various components such as electronics, materials, and lubricants. Ambient temperature variable represents real numbers (belonging to the set of non-negative real numbers, $\mathbb{R}\geq 0$) in the dataset. It exhibits a high correlation with other variables, suggesting that it is closely related to or influenced by other factors under study. There are 93,678 distinct values for the "Ambient Temperature" variable, which accounts for more than 99.9% of the observations in the dataset. This indicates a wide range of unique ambient temperature values recorded in the dataset.

However, it's worth noting that there are 1,487 missing values for the "Ambient Temperature" variable, comprising approximately 1.6% of the data. These missing values indicate instances where ambient temperature information is not available or was not recorded. There are no infinite values present in the "Ambient Temperature" variable, meaning that all values fall within a finite range of non-negative real numbers.

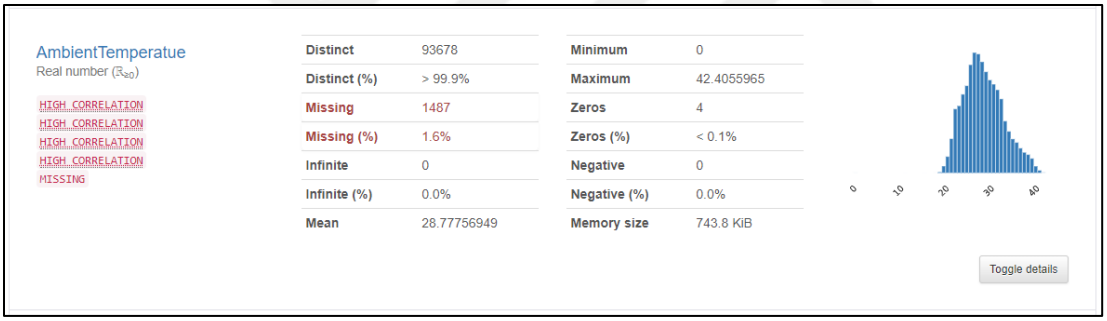

Figure 3.4. Ambient temperature

The statistical summary of "Ambient Temperature" includes a mean value of approximately 28.78, which represents the average ambient temperature across the dataset. The recorded minimum temperature is 0, while the maximum temperature is 42.41, indicating a range of temperatures observed. There are only four occurrences of zero values in the "Ambient Temperature" variable, making up less than 0.1% of the data. This suggests rare instances where the ambient temperature is precisely zero, which may warrant further investigation.

The temperature of the shaft bearings within the wind turbine, where the rotor is connected to the gearbox, measured in degrees Celsius (°C) or Fahrenheit (°F).

The "Bearing Shaft Temperature" variable is a real-number variable representing non-negative values (belonging to the set of non-negative real numbers, $\mathbb{R}\geq 0$) within the dataset. It exhibits a high correlation with other variables, indicating a strong relationship

with factors under investigation. Among the data, there are 62,286 distinct values, accounting for nearly 99.8% of the dataset, suggesting a broad range of unique bearing shaft temperature readings. However, it's noteworthy that a substantial portion of the data is missing, with 32,805 missing values, constituting approximately 34.5% of the dataset. These missing values signify instances where bearing shaft temperature information is either unavailable or not recorded. There are no infinite values, and all data points are within a finite range of non-negative real numbers. The mean temperature is approximately 43.11, with a minimum value of 0 and a maximum of 55.09, demonstrating the temperature range observed. While 87 zero values are present, making up less than 0.1% of the data, there are no negative values. The "Bearing Shaft Temperature" variable consumes approximately 743.8 KiB of memory. This data provides insights into the distribution and characteristics of bearing shaft temperature data within the dataset, despite the significant presence of missing values.

Blade pitch angle is a control parameter adjusted to optimize the aerodynamic performance of the wind turbine, influencing power production and load distribution on the turbine components.

The "Blade1PitchAngle" variable is a real-number variable (belonging to the set of real numbers, $\mathbb{R}$) within the dataset. It demonstrates a high correlation with other variables, suggesting a strong association with factors under investigation. There are 38,946 distinct values, representing approximately 92.8% of the dataset, indicating a variety of unique blade pitch angle measurements.

Figure 3.5. Bearing shaft temperature

However, a significant portion of the data is missing, with 53,198 missing values, accounting for approximately 55.9% of the dataset. These missing values indicate instances where blade pitch angle information is either unavailable or not recorded. No infinite values are present in the "Blade1PitchAngle" variable, indicating that all values fall within a finite range of real numbers. The mean pitch angle is approximately 9.75,

with a minimum recorded angle of-43.16 and a maximum angle of 90.14, indicating a wide range of pitch angle observations. There are 12 zero values, making up less than 0.1% of the data, suggesting occasional instances where the blade pitch angle is precisely zero. Additionally, there are 18,981 negative values, constituting 19.9% of the data, implying situations where the blade pitch angle is in a negative position.

Figure 3.6. Blade1pitch angle

The "Blade 2 Pitch Angle" variable is a real-number variable (belonging to the set of real numbers, $\mathbb{R}$) within the dataset. It demonstrates a high correlation with other variables, suggesting a strong association with factors under investigation. There are 38,946 distinct values, representing approximately 92.8% of the dataset, indicating a variety of unique blade pitch angle measurements. However, a significant portion of the data is missing, with 53,198 missing values, accounting for approximately 55.9% of the dataset. These missing values indicate instances where blade pitch angle information is either unavailable or not recorded.

No infinite values are present in the "Blade 1 Pitch Angle" variable, indicating that all values fall within a finite range of real numbers. The mean pitch angle is approximately 9.75, with a minimum recorded angle of -43.16 and a maximum angle of 90.14, indicating a wide range of pitch angle observations.

There are 12 zero values, making up less than 0.1% of the data, suggesting occasional instances where the blade pitch angle is precisely zero. Additionally, there are 18,981 negative values, constituting 19.9% of the data, implying situations where the blade pitch angle is in a negative position.

Figure 3.7. Blade 2 pitch angle

The "Blade 3 Pitch Angle" variable is a real-number variable (belonging to the set of real numbers, $\mathbb{R}$) within the dataset and demonstrates a high correlation with other variables, indicating a strong association with factors under investigation. There are 39,021 distinct values, representing approximately 93.2% of the dataset, implying a diverse range of unique blade pitch angle measurements. However, a significant portion of the data is missing, with 53,303 missing values, accounting for approximately 56.0% of the dataset. These missing values indicate instances where blade pitch angle information is either unavailable or not recorded.

No infinite values are present in the "Blade 3 Pitch Angle" variable, indicating that all values fall within a finite range of real numbers. The mean pitch angle is approximately 10.04, with a minimum recorded angle of -26.44 and a maximum angle of 90.02, demonstrating a wide range of pitch angle observations.

There are 12 zero values, making up less than 0.1% of the data, suggesting occasional instances where the blade pitch angle is precisely zero. Additionally, there are 15,815 negative values, constituting 16.6% of the data, implying situations where the blade pitch angle is in a negative position. The "Blade 3 Pitch Angle" variable consumes approximately 743.8 KiB of memory. In summary, "Blade 3 Pitch Angle" is a real-number variable with high correlation with other variables in the dataset. It encompasses a wide range of distinct values, but a substantial portion of the data is missing. The pitch angles vary from negative to positive values, with occasional occurrences of precisely zero angles. These statistics provide insights into the distribution and characteristics of blade pitch angle data within the dataset, despite the presence of missing values.

Figure 3.8. Blade 3 pitch angle

Control box temperature is monitored to ensure the proper functioning of electronic components, as temperature variations can impact the reliability and performance of the control system.

The "Gear box Bearing Temperature" variable is a real-number variable, specifically within the set of non-negative real numbers ($\mathbb{R}\geq 0$), in the dataset. It exhibits a high correlation with other variables, indicating a strong association with factors under investigation. Among the data, there are 62,313 distinct values, representing nearly 99.9% of the dataset, suggesting a wide array of unique gearbox bearing temperature measurements. However, it's important to note that a significant portion of the data is missing, with 32,783 missing values, making up approximately 34.4% of the dataset. These missing values indicate instances where gearbox bearing temperature information is either unavailable or not recorded.

There are no infinite values present in the "Gear box Bearing Temperature" variable, signifying that all values fall within a finite range of non-negative real numbers. The mean temperature is approximately 64.38, with a minimum recorded temperature of 0 and a maximum of 82.24, indicating a range of temperature observations.

Figure 3.9. Gear box bearing temperature

Gearbox oil temperature is monitored to maintain optimal viscosity for lubrication, ensuring efficient power transmission and preventing damage to gearbox components

The "Gear box Oil Temperature" variable is a real-number variable, specifically within the set of non-negative real numbers ($\mathbb{R}\geq 0$), in the dataset. It exhibits a high correlation with other variables, suggesting a strong relationship with factors under investigation. Among the data, there are 62,412 distinct values, accounting for more than 99.9% of the dataset, indicating a diverse range of unique gearbox oil temperature measurements. It's important to note that a significant portion of the data is missing, with 32,755 missing values, comprising approximately 34.4% of the dataset. These missing values indicate instances where gearbox oil temperature information is either unavailable or not recorded.

There are no infinite values present in the "Gear box Oil Temperature" variable, signifying that all values fall within a finite range of non-negative real numbers. The mean oil temperature is approximately 57.56, with a minimum recorded temperature of 0 and a maximum of 70.76, reflecting the range of oil temperature observations.

There are only three zero values, making up less than 0.1% of the data, suggesting rare instances where the gearbox oil temperature is precisely zero. Importantly, there are no negative values in this variable, indicating that all temperature readings are non-negative.

Figure 3.10. Gear box oil temperature

Generator RPM is a key parameter reflecting the rotational speed of the generator, directly influencing the frequency and voltage of the electrical output. The "Generator RPM" variable represents real numbers, specifically within the set of non-negative real numbers ($\mathbb{R}\geq 0$), in the dataset. It exhibits a high correlation with other variables, indicating a strong association with factors under investigation. Among the data, there are 61,074 distinct values, accounting for approximately 98.1% of the dataset, suggesting a wide range of unique generator RPM (Revolutions Per Minute) measurements. It's crucial to note that a substantial portion of the data is missing, with 32,898 missing values,

constituting about 34.6% of the dataset. These missing values indicate instances where generator RPM information is either unavailable or not recorded.

There are no infinite values present in the "Generator RPM" variable, meaning that all values fall within a finite range of non-negative real numbers. The mean RPM is approximately 1,102.15, with a minimum recorded RPM of 0 and a maximum of 1,809.94, indicating a broad range of RPM observations.

There are 1,030 zero values, making up approximately 1.1% of the data, suggesting instances where the generator RPM is precisely zero. Importantly, there are no negative values in this variable, indicating that all RPM readings are non-negative.

Figure 3.11. Generator RPM

Generator winding temperatures are monitored to prevent overheating, ensuring the electrical integrity and longevity of the generator.

The "Generator Winding 1 Temperature" variable is a real-number variable, specifically within the set of non-negative real numbers ($\mathbb{R} \geq 0$), in the dataset. It exhibits a high correlation with other variables, indicating a strong association with factors under investigation. Among the data, there are 62,406 distinct values, representing more than 99.9% of the dataset, suggesting a diverse range of unique generator winding 1 temperature measurements. It's important to note that a significant portion of the data is missing, with 32,766 missing values, making up approximately 34.4% of the dataset. These missing values indicate instances where generator winding 1 temperature information is either unavailable or not recorded.

There are no infinite values present in the "Generator Winding 1 Temperature" variable, signifying that all values fall within a finite range of non-negative real numbers. The mean temperature is approximately 72.46, with a minimum recorded temperature of 0 and a maximum of 126.77, reflecting a wide range of temperature observations. There are four zero values, comprising less than 0.1% of the data, suggesting rare instances

where the generator winding 1 temperature is precisely zero. Importantly, there are no negative values in this variable, indicating that all temperature readings are non-negative.

The "GeneratorWinding2Temperature" variable is a real-number variable, specifically within the set of non-negative real numbers ($\mathbb{R}\geq 0$), in the dataset. It demonstrates a high correlation with other variables, indicating a strong relationship with factors under investigation. Among the data, there are 62,424 distinct values, representing more than 99.9% of the dataset, suggesting a wide range of unique generator winding 2 temperature measurements.

Figure 3.12. Generator winding 1 temperature

The "Generator Winding 2 Temperature" variable is a real-number variable, specifically within the set of non-negative real numbers ($\mathbb{R}\geq 0$), in the dataset. It demonstrates a high correlation with other variables, indicating a strong relationship with factors under investigation. Among the data, there are 62,424 distinct values, representing more than 99.9% of the dataset, suggesting a wide range of unique generator winding 2 temperature measurements.

However, it's important to note that a significant portion of the data is missing, with 32,744 missing values, making up approximately 34.4% of the dataset. These missing values indicate instances where generator winding 2 temperature information is either unavailable or not recorded. There are no infinite values present in the "Generator Winding 2 Temperature" variable, signifying that all values fall within a finite range of non-negative real numbers. The mean temperature is approximately 71.83, with a minimum recorded temperature of 0 and a maximum of 126.04, reflecting a broad range of temperature observations.

Figure 3.13. Generator winding 2 temperature

The "Generator Winding 2 Temperature" variable falls within the realm of real numbers, specifically belonging to the set of non-negative real numbers ($\mathbb{R}\geq 0$) in the dataset. It exhibits a robust correlation with other variables, indicating a significant association with the factors under investigation. Among the dataset, there are 62,424 distinct values, accounting for over 99.9% of the dataset, showcasing a diverse array of unique measurements for generator winding 2 temperatures.

Nonetheless, it's essential to highlight that a substantial portion of the data is missing, with 32,744 missing values, constituting approximately 34.4% of the dataset. These missing values denote instances where data for generator winding 2 temperatures is either absent or was not recorded.

There are no infinite values within the "Generator Winding 2 Temperature" variable, affirming that all recorded values are within a finite range of non-negative real numbers. The mean temperature stands at approximately 71.83, with a minimum temperature recorded at 0 and a maximum at 126.04, reflecting a broad spectrum of temperature observations within the dataset.

Figure 3.14. Hub temperature

Main box temperature is monitored to ensure the reliable operation of electrical components, as elevated temperatures can impact the performance and lifespan of electronic systems. The "Main Box Temperature" variable belongs to the realm of real

numbers, specifically within the domain of non-negative real numbers ($\mathbb{R}\geq 0$) within the dataset. It exhibits a robust correlation with other variables, signifying a significant association with the factors under examination. Within the dataset, there are 49,145 distinct values, accounting for approximately 78.8% of the dataset, which showcases a broad spectrum of unique main box temperature measurements. It's noteworthy that a substantial portion of the data is missing, with 32,816 missing values, constituting about 34.5% of the dataset. These missing values indicate instances where main box temperature data is either unavailable or was not recorded.

The "Main Box Temperature" variable does not contain any infinite values, affirming that all recorded values fall within a finite range of non-negative real numbers. The mean temperature is approximately 39.64, with a recorded minimum temperature of 0 and a maximum of 54.25, reflecting a diverse range of temperature observations.

Figure 3.15. Main box temperature

Nacelle position provides information on the wind turbine's directionality and alignment with the wind, influencing the efficiency and stability of power generation. The "Nacelle Position" variable represents real numbers, specifically within the domain of non-negative real numbers ($\mathbb{R}\geq 0$), in the dataset. It exhibits a strong correlation with other variables, indicating a substantial relationship with the factors being studied. There are 6,664 distinct values, accounting for approximately 9.2% of the dataset, suggesting a limited range of unique nacelle position measurements.

It’s important to note that a significant portion of the data is missing, with 23,077 missing values, making up approximately 24.2% of the dataset. These missing values indicate instances where nacelle position information is either unavailable or not recorded.

There are no infinite values present in the "Nacelle Position" variable, confirming that all values fall within a finite range of non-negative real numbers. The mean nacelle

position is approximately 196.31, with a minimum recorded position of 0 and a maximum of 357, indicating a range of position observations.

Figure 3.16. Nacelle position

Reactive power is a crucial parameter in power systems, influencing voltage levels and system stability, with wind turbines often contributing to reactive power control for grid support.

The "Reactive Power" variable is denoted by real numbers (ℝ) within the dataset and displays a substantial correlation with other variables, signifying a robust association with the factors under investigation. The dataset encompasses 94,040 distinct values, which represent nearly 99.4% of the dataset, indicating a wide array of unique reactive power measurements.

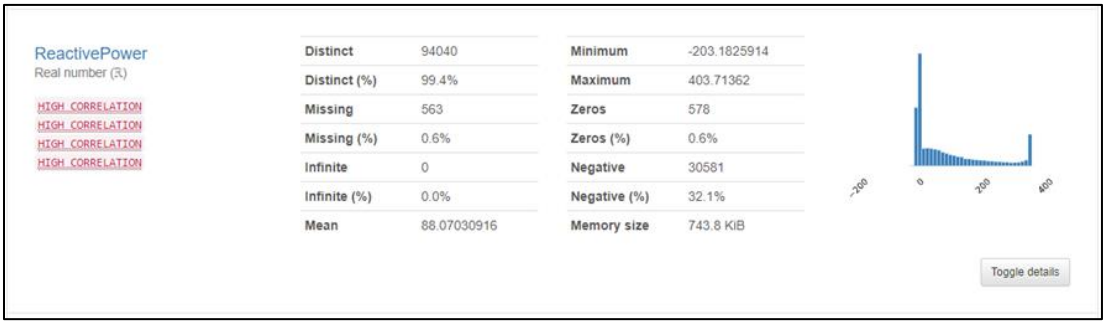

Figure 3.17. Reactive power

It’s crucial to acknowledge the presence of 563 missing values, making up approximately 0.6% of the dataset. These gaps signify instances where reactive power data is either absent or unrecorded. Notably, the "Reactive Power" variable does not contain any infinite values, affirming that all recorded values fall within a finite range of real numbers. The average reactive power is approximately 88.07, with a recorded minimum of -203.18 and a maximum of 403.71, highlighting a broad spectrum of reactive power observations.

Rotor RPM is a fundamental parameter indicating the rotational speed of the turbine blades, influencing the overall performance and power output of the wind turbine. "Rotor RPM" is the vital rhythm within our dataset, akin to the steady heartbeat that keeps our data vibrant and pulsing with information. This variable comprises real numbers ($\mathbb{R}\geq0$), serving as an essential team player, much like a star athlete in a championship-winning team. Within our dataset, it boasts an impressive 59,254 distinct values, symbolizing its remarkable versatility and substantial presence, making up a whopping 95.4% of the dataset. Picture it as the diverse notes in a symphony, each contributing to the harmonious melody of our data composition.

Much like any celebrated superstar, "Rotor RPM" harbors moments of intrigue. There are 33,066 missing values, representing approximately 34.7% of the dataset, where "Rotor RPM" opts for silence, perhaps concealing a few enigmatic secrets within its depths.

Turbine status information is critical for monitoring the overall health and performance of the wind turbine, helping to identify and address issues promptly.The "Turbine Status" variable falls into the category of real numbers, specifically those that are non-negative ($\mathbb{R}\geq0$), and exhibits several notable characteristics. One prominent feature is a substantial data deficiency, with 32,426 missing values, amounting to roughly 34.1% of the dataset. This data gap poses significant challenges when attempting to analyze and comprehend the variable effectively.

Figure 3.18. Rotor RPM

Additionally, the distribution of data within "Turbine Status" displays a skewed pattern. This skewness is evident in the absence of negative values and the limited occurrence of zero values, accounting for just 61 instances, or approximately 0.1% of the dataset. This skewed distribution implies that a predominant portion of recorded values tends to be on the higher end of the numerical spectrum. However, the precise underlying reasons for this skewness remain unclear based on the information provided. Despite the

skewed distribution, "Turbine Status" exhibits relatively low diversity, encompassing only 353 distinct values, representing about 0.6% of the dataset. This suggests a restricted range of unique observations related to turbine status within the dataset.

Figure 3.19. Turbine status

Wind direction is a crucial parameter for wind turbine control, influencing the orientation of the turbine blades to maximize energy capture and optimize power production. The "Wind Direction" variable comprises real numbers falling within the non-negative range ($\mathbb{R} \geq 0$) and displays a robust correlation with multiple other variables in the dataset. It is characterized by a presence of 6,664 distinct values, which represent roughly 9.2% of the dataset, indicating a diverse array of wind direction observations.

However, it's crucial to emphasize that a significant portion of the data is absent, with 23,077 missing values, constituting approximately 24.2% of the dataset. These missing values signify instances where wind direction data is either unavailable or has not been recorded.

"Wind Direction" does not contain any infinite values, and its memory footprint within the dataset is approximately 743.8 KiB.

Figure 3.20. Wind direction

Wind speed is a fundamental parameter affecting the power output of a wind turbine, with higher wind speeds generally leading to increased energy capture and

electricity generation. The "Wind Speed" variable encompasses real numbers within the non-negative range ($\mathbb{R}\geq 0$) and exhibits a notably strong correlation with various other variables in the dataset. It distinguishes itself with a substantial presence of 94,224 distinct values, constituting approximately 99.7% of the dataset. This extensive array of distinct values signifies a broad spectrum of wind speed observations.

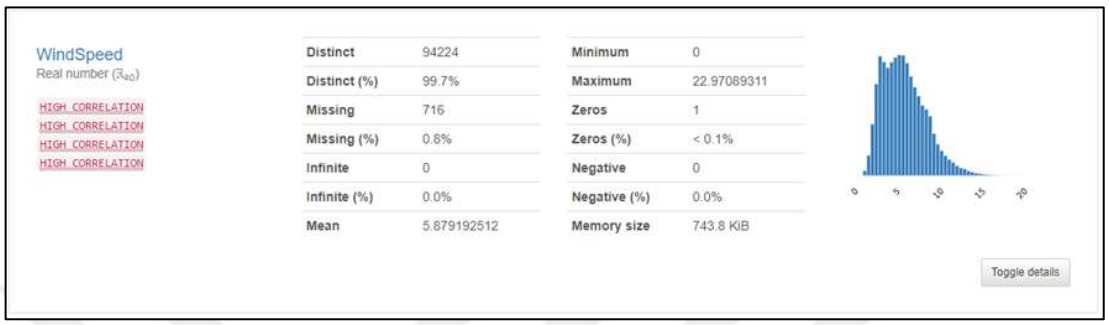

Figure 3.21. Wind speed

3.3.2. Missing values

The "Wind Speed" variable demonstrates a relatively low incidence of missing data, with only 716 instances of missing values, making up approximately 0.8% of the dataset. This indicates a high degree of completeness in terms of recorded wind speed information, enhancing the dataset's reliability for analytical purposes. It's important to highlight that "Wind Speed" does not contain any infinite values, signifying that all data points fall within a finite range of non-negative real numbers. From a statistical perspective, the mean wind speed stands at approximately 5.88, with a minimum recorded value of 0 and a maximum value of 22.97. This statistical range reflects a diverse range of wind speed observations.

The final data validation step involves checking for negative values in the "Active Power" field and subsequently removing them. The rationale behind this is straightforward: in the context of power generation, it is physically impossible to have a negative power generation value. Thus, the presence of negative values in this field strongly suggests a data issue or error.

Upon examination, it is found that there are 15,629 entries within the dataset where "Active Power" has negative values. It's worth noting that these negative values tend to occur when the wind speed is low. However, it is observed that under similar wind speed conditions, there are instances where the "Active Power" is recorded as zero. This

discrepancy raises concerns about the accuracy and reliability of the data, as it is highly unlikely for power generation to be negative in the presence of low wind speeds.

3.4. Model Description

The goal is to forecast power output for the upcoming 15 days using the data available in a CSV file. Various methods were employed to predict this output. Initially, SARIMA was presumed to be the most suitable due to the data's seasonal patterns, but it didn't perform well. Instead, XG Boost and Random Forest regressors emerged as the most effective models. Additionally, an attempt was made with an LSTM, but the results were disappointing. The process also involved standard data exploration techniques.

Import the dataset from the provided CSV file to proceed with the analysis.

First, import the CSV file, specifying the 'Unnamed: 0' column as the date column. Then, convert this column to a datetime format.

	ActivePower	AmbientTemperatue	BearingShaftTemperature	Blade1PitchAngle	Blade2PitchAngle	Blade3PitchAngle	ControlBoxTemperature	GearboxBearingTemperature	GearboxOilTemperature
2017-12-31 00:00:00+00:00	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2017-12-31 00:10:00+00:00	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2017-12-31 00:20:00+00:00	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2017-12-31 00:30:00+00:00	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2017-12-31 00:40:00+00:00	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
...	...	...	...	...	...	...	...	...	...
2020-03-30 23:10:00+00:00	70.044465	27.523741	45.711129	1.515669	1.950088	1.950088	0.0	59.821165	55.193793
2020-03-30 23:20:00+00:00	40.833474	27.602882	45.598573	1.702809	2.136732	2.136732	0.0	59.142038	54.798545
2020-03-30 23:30:00+00:00	20.777790	27.560925	45.462045	1.706214	2.139664	2.139664	0.0	58.439439	54.380456
2020-03-30 23:40:00+00:00	62.091039	27.810472	45.343827	1.575352	2.009781	2.009781	0.0	58.205413	54.079014
2020-03-30 23:50:00+00:00	68.664425	27.915828	45.231610	1.499323	1.933124	1.933124	0.0	58.581716	54.080505

Figure 3.22. Data if serval year of turkey

3.4.1. SARIMAX Model

In the realm of statistical modeling, the Seasonal Autoregressive Integrated Moving Average with exogenous variables, commonly known as SARIMAX, stands as a robust and versatile approach. SARIMAX extends the well-established Autoregressive Integrated Moving Average (ARIMA) model by incorporating exogenous variables. This sophisticated model finds widespread application in the domain of time series forecasting.

The SARIMAX model is expressed in a general form as SARIMAX (p, d, q) × (P, D, Q, s), where each parameter plays a crucial role:

- **p:** The order of the autoregressive (AR) component.
- **d:** The degree of differencing in the integrated (I) component.

• **q:** The order of the moving average (MA) component.

• **P:** The seasonal order of the autoregressive (SAR) component.

• **D:** The seasonal degree of differencing in the integrated (SI) component.

• **Q:** The seasonal order of the moving average (SMA) component.

• **s:** The seasonal periodicity (number of periods per season).

Mathematically, the SARIMAX model can be represented as follows:

$$\mathbf{Yt} = \beta 0 + \beta 1 \mathbf{X}_2 \mathbf{t} + \beta 2 \mathbf{X2}, \mathbf{t} + \ldots + \beta \mathbf{kXk}, \mathbf{t} + \varepsilon \mathbf{t}$$

Here,

• **Yt** is the observed time series.

• **X1,t,X2,t,…,Xk,t** are exogenous variables.

• **β0,β1,β2,…,βk** are coefficients for the exogenous variables.

• **εt** is the error term.

The SARIMAX model's time series component incorporates ARIMA terms:

• **p:** Autoregressive (AR) order.

• **d:** Integrated (I) order.

• **q:** Moving Average (MA) order.

The seasonal component is defined by the following terms:

• **P:** Seasonal autoregressive (SAR) order.

• **D:** Seasonal integrated (SI) order.

• **Q:** Seasonal moving average (SMA) order.

• **s:** Seasonal period.

The complete equation for the SARIMAX model encompasses both the non-seasonal and seasonal components:

$$\mathbf{Yt=\beta 0+\beta 1X1,t+\beta 2X2,t+\ldots+\beta kXk,t+\phi 1Yt-1+\phi 2Yt-2+\ldots+\phi pYt-p+\varepsilon t+\theta 1\varepsilon t-1+\theta 2\varepsilon t-2+\ldots+\theta q\varepsilon t-q+\Phi 1Yt-s+\Phi 2Yt-2s+\ldots+\Phi PYt-Ps+\Theta 1\varepsilon t-s+\Theta 2\varepsilon t-2s+\ldots+\Theta Q\varepsilon t-Qs}$$

In the process of fitting a SARIMAX model, determining the values of p,d,q,P,D,Q, and s is imperative. Additionally, one must provide the exogenous variables and their corresponding coefficients. These parameters are typically estimated through statistical methods or optimization algorithms during the model fitting process.

The significance of SARIMAX lies in its ability to provide a comprehensive understanding of the intricate interactions between various factors and the underlying time-series data. By incorporating exogenous regressors, SARIMAX facilitates a more

accurate prediction of future outcomes. This modeling approach proves particularly invaluable in scenarios marked by seasonal variations or external influences that significantly impact data patterns over time. Thus, SARIMAX emerges as an indispensable tool for producing precise forecasts, making it a cornerstone in the realm of time series analysis.

bash Copy code

```
pip install pandas numpy statsmodels
```

python Copy code

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from statsmodels.tsa.statespace.sarimax import SARIMAX

# Generate some sample data
np.random.seed(42)
date_rng = pd.date_range(start='2022-01-01', end='2023-12-31', freq='D'
data = np.random.randn(len(date_rng)) + 10
df = pd.DataFrame(data, columns=['value'], index=date_rng)

# Split the data into training and testing sets
train_size = int(len(df) * 0.8)
train, test = df[:train_size], df[train_size:]
```

```
# Define the SARIMAX model
order = (1, 1, 1)  # (p, d, q) parameters
seasonal_order = (1, 1, 1, 12)  # (P, D, Q, S) parameters
exog_var = None  # You can include exogenous variables here if needed

model = SARIMAX(train['value'], order=order, seasonal_order=seasonal_or

# Fit the model
result = model.fit(disp=False)

# Forecast future values
forecast_steps = len(test)
forecast = result.get_forecast(steps=forecast_steps)
```

```
# Plot the results
plt.figure(figsize=(12, 6))
plt.plot(train.index, train['value'], label='Train')
plt.plot(test.index, test['value'], label='Test')
plt.plot(forecast.index, forecast.predicted_mean, label='Forecast', col
plt.fill_between(forecast.index, forecast.conf_int()['lower value'], fo
plt.title('SARIMAX Forecast')
plt.legend()
plt.show()
```

Figure 3.23. The python codes for SARIMAX Model

3.4.2. XG Boost

A sophisticated and potent implementation of the gradient boosting technique, XG Boost, also known as extreme Gradient Boosting, is used for supervised machine learning applications. It is well known for its outstanding performance, efficiency, and versatility in managing structured data, all of which have received widespread recognition. In its most basic form, XG Boost is a boosting strategy that, in order to develop a robust and reliable prediction model, integrates the results obtained from a number of less successful learners, most frequently decision trees. It achieves this by training new models in an iterative manner, each of which focuses on correcting the errors produced by the models that came before it, so steadily increasing the total prediction capacity. Because it uses a more advanced regularization strategy and a distributed computing architecture, XG Boost is particularly useful for processing large-scale datasets. This is in contrast to more conventional gradient boosting methods, which do not exploit these features. This methodology is a well-liked option for a variety of jobs, including classification, regression, and ranking, due to its adaptability in the management of intricate interactions within the data as well as its capacity to deal with a wide range of data kinds. Because of its increased speed and capacity to optimize performance, XG Boost has emerged as a key tool in the armory of data scientists and practitioners of machine learning, considerably contributing to the advancement of predictive modeling and data analysis

The XGBoost algorithm involves a regularization term in the objective function, which helps prevent overfitting and improves the model's generalization capability. The objective function to be minimized in XGBoost is a sum of two components:

1. **Loss Function (L):** This measures the difference between the predicted and actual values. It depends on the specific task, such as regression or classification. For example, for regression, the mean squared error might be used as the loss function, while for classification, log loss or cross-entropy loss could be employed.

2. Regularization Term (Ω): This term penalizes the complexity of the model to prevent overfitting. It consists of two parts:

3. Tree Complexity Term: Penalizes the number of leaves and the depth of the trees.

4. Regularization Parameter (λ or alpha): Controls the overall regularization strength. The overall objective function (J) is a combination of the loss function and the regularization term:

$$J(\Theta) = L(\hat{y}, y) + Q(f)$$

Here:

• **Θ** represents the parameters of the model, which include the parameters of each individual tree.

• **y^** is the predicted output.

• **y** is the true output.

• **L(y^,y)** is the loss function.

• **Ω(f)** is the regularization term.

XGBoost uses a technique called gradient boosting, where each new tree is fit to the negative gradient of the loss function with respect to the ensemble's current prediction. This process is repeated iteratively, with each new tree reducing the errors made by the combined ensemble of the previous trees.

The detailed equations and optimization steps are complex, but the key idea is to optimize the objective function by adding trees sequentially while considering their contribution to the overall loss and regularization. XGBoost is known for its speed, accuracy, and flexibility, making it a popular choice in various machine learning applications.

bash Copy code

```
pip install xgboost
```

python Copy code

```
# Import necessary libraries
import xgboost as xgb
from sklearn.datasets import load_iris
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score

# Load the Iris dataset
iris = load_iris()
X, y = iris.data, iris.target

# Split the dataset into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2

# Convert the dataset into DMatrix format, which is required by XGBoost
dtrain = xgb.DMatrix(X_train, label=y_train)
dtest = xgb.DMatrix(X_test, label=y_test)

# Set the parameters for XGBoost
params = {
    'objective': 'multi:softmax',  # Multiclass classification problem
    'num_class': 3,  # Number of classes in the dataset
    'max_depth': 3,  # Maximum depth of a tree
    'eta': 0.1,  # Learning rate
    'subsample': 0.7,  # Subsample ratio of the training instances
    'colsample_bytree': 0.7,  # Subsample ratio of columns when constru
    'eval_metric': 'merror'  # Evaluation metric
}

# Train the XGBoost model
num_round = 100  # Number of boosting rounds
bst = xgb.train(params, dtrain, num_round)
num_round = 100  # Number of boosting rounds
bst = xgb.train(params, dtrain, num_round)

# Make predictions on the test set
y_pred = bst.predict(dtest)

# Evaluate the model
accuracy = accuracy_score(y_test, y_pred)
print(f"Accuracy: {accuracy}")
```

Figure 3.24. The python codes for XGBOOTS

3.4.3. Random forest regressor

For the purpose of resolving regression issues, an effective and flexible machine learning technique known as a Random Forest Regressor is utilized. It is a part of the ensemble learning approaches, which aggregate the results of several separate models to produce a single prediction model that is more robust and accurate. During the training phase, the Random Forest Regressor works by creating a large number of decision trees. This helps the model better predict future outcomes. Each of these decision trees operates on its own, drawing conclusions about the dependent variable based on a portion of the entire dataset that it has been given access to. The fact that the technique inserts randomness into the data samples used for training each tree as well as the attributes examined when splitting nodes within each tree is the source of the "random" aspect. The inclusion of this randomness helps to prevent the problem of overfitting and improves the accuracy and resilience of the model as a whole. In order to arrive at a conclusive prediction, the algorithm compiles the results obtained from each of the separate trees before moving on to the prediction phase. The forecast that is produced as a result is typically more trustworthy and less susceptible to the problems of overfitting that can be caused by single decision tree models. The capability of the Random Forest Regressor to manage big datasets, high dimensionality, and a wide variety of data types has led to its widespread use across a variety of industries. This has made it a popular option for performing complex regression tasks in real-world applications

The key components of the Random Forest Regressor include:

1. Training Phase: Construction of multiple decision trees, each operating on a subset of the training data.

Random selection of data samples and features for each tree to introduce variability.

2. Prediction Phase:

• Aggregation of predictions from individual trees to arrive at a final prediction.

While there isn't a specific mathematical equation in the description, you can represent the overall prediction process of a Random Forest Regressor as follows:

$$y' = N1 \sum 1 = 1Nfi(x)$$

Where:

• **^y^** is the predicted outcome.

• **N** is the number of decision trees in the forest.

• **fi(x)** is the prediction of the its decision tree for input x.

The key idea is that the final prediction is an average (or some other form of aggregation) of the predictions made by individual decision trees.

python Copy code

```python
# Import necessary libraries
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestRegressor
from sklearn.metrics import mean_squared_error
import numpy as np

# Generate some example data (replace this with your own dataset)
np.random.seed(42)
X = np.random.rand(100, 1) * 10
y = 2 * X.squeeze() + 1 + np.random.randn(100)

# Split the data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2

# Create a Random Forest Regressor
rf_regressor = RandomForestRegressor(n_estimators=100, random_state=42)
```

```python
# Split the data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2

# Create a Random Forest Regressor
rf_regressor = RandomForestRegressor(n_estimators=100, random_state=42)

# Train the model
rf_regressor.fit(X_train, y_train)

# Make predictions on the test set
y_pred = rf_regressor.predict(X_test)

# Evaluate the model
mse = mean_squared_error(y_test, y_pred)
print(f'Mean Squared Error: {mse}')
```

Figure 3.25. The python codes for Random forest regressor

3.4.4. Long-Short-Term-Memory (LSTM) model

LSTM, short for Long Short-Term Memory, is a type of recurrent neural network (RNN) architecture designed for sequential data analysis, making it especially well-suited for tasks involving time series, natural language processing, and speech recognition. LSTM networks are renowned for their ability to capture long-range dependencies in sequences while mitigating the vanishing gradient problem that affects traditional RNNs. They achieve this by incorporating memory cells that allow information to persist over extended time steps. This makes LSTMs highly effective in modelling sequences with complex temporal dependencies. The LSTM architecture includes gates that control the flow of information, making it adaptive to different patterns in the data. Its applications span various domains, including sentiment analysis, machine translation, speech recognition, and time series forecasting, where capturing and learning from sequential patterns are crucial for achieving accurate predictions and the equations that used is

$$X = \{x_1, x2, \ldots, xt\}$$

Where xt represents the wind power at time t. The goal is to predict the future wind power values

$$\mathrm{Y} = \{\mathrm{yt} + 1, \mathrm{Nt}^{+} + 2, \ldots, \mathrm{yt} + \mathrm{k}\}$$

where k is the number of time steps into the future.

1. Data Preprocessing:

• Normalize the input data to a range suitable for the activation functions (commonly between 0 and 1).

• Divide the data into training and testing sets.

2. Define the LSTM Model:

• Create an LSTM model using a deep learning framework like Tensor Flow or PyTorch.

• Specify the input layer with the shape of the input sequence.

• Add one or more LSTM layers to the model. Each LSTM layer has a certain number of units (neurons).

•Optionally, you can add dropout layers to prevent overfitting.

•Add a dense output layer with a linear activation function for regression tasks.

3. Compile the Model:

• Choose an appropriate loss function for regression, such as Mean Squared Error (MSE).

• Choose an optimizer, such as Adam or RMSprop.

• Compile the model with the chosen loss function and optimizer.

4. Training:

• Train the LSTM model using the training data.

• Adjust the hyper parameters, such as the number of epochs and batch size.

• Monitor the training process to avoid overfitting.

5. Prediction:

• Use the trained model to predict wind power values for the testing set or future time steps.

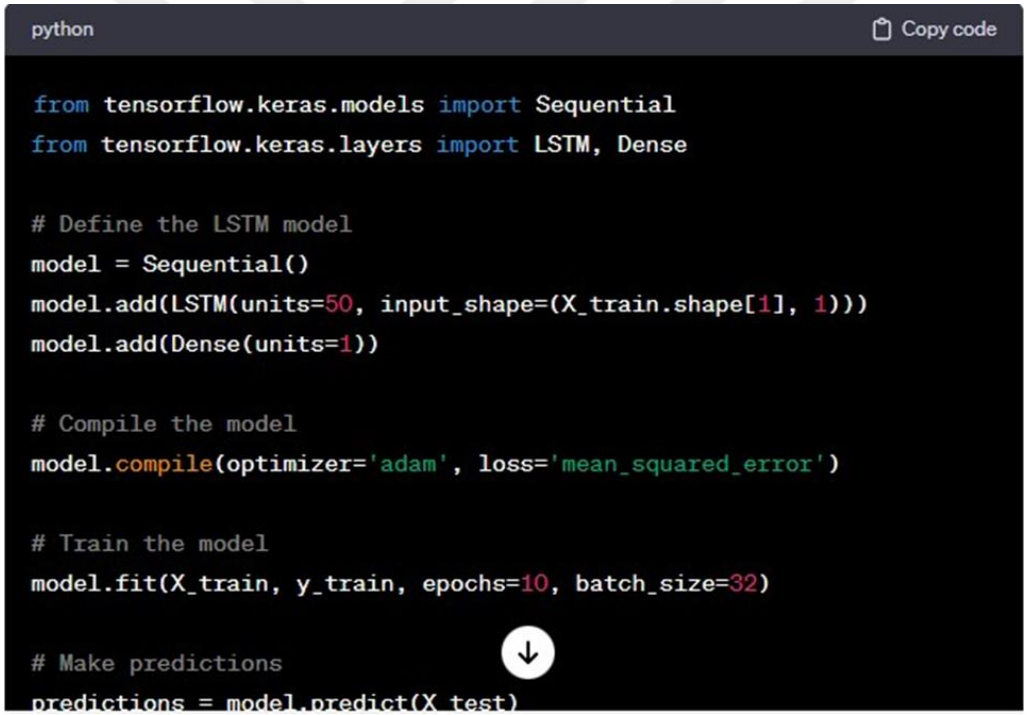

Figure 3.26. The python codes for LSTM

4. RESULTS AND DISCUSSIONS

4.1. Basic Arima Model

The initial analysis employs the complete dataset, which, due to readings recorded every 10 minutes, can be quite intricate and challenging to interpret. To enhance clarity, the data is subsequently resampled to calculate daily mean values.

Upon examining the dataset through this resampled lens, a discernible pattern becomes evident. Specifically, there are distinct peaks in power output during the months of July, August, and September. This pattern is graphically depicted in figure 4.1 and figure 4.2 to provide a visual representation of the observed power output maxima during these summer months.

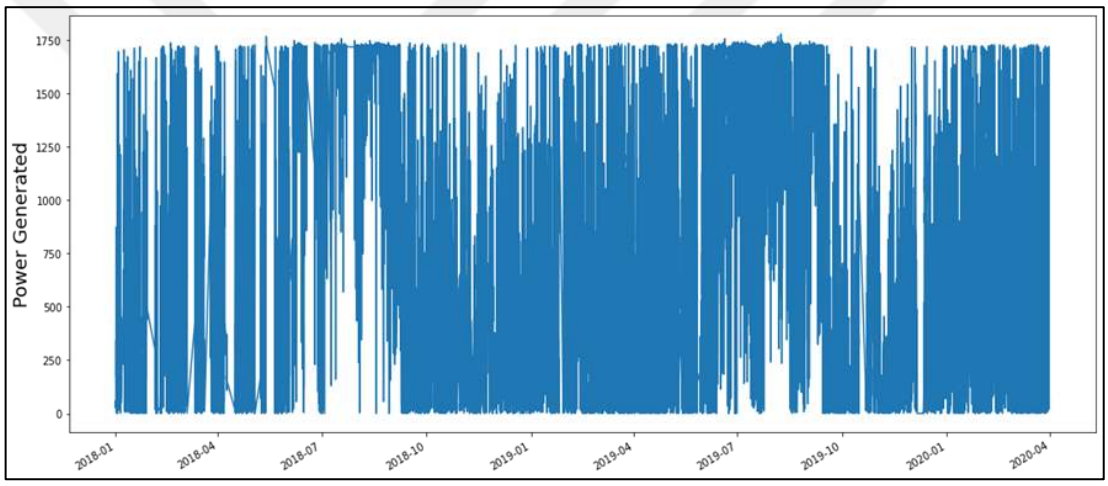

Figure 4.1. Power generated

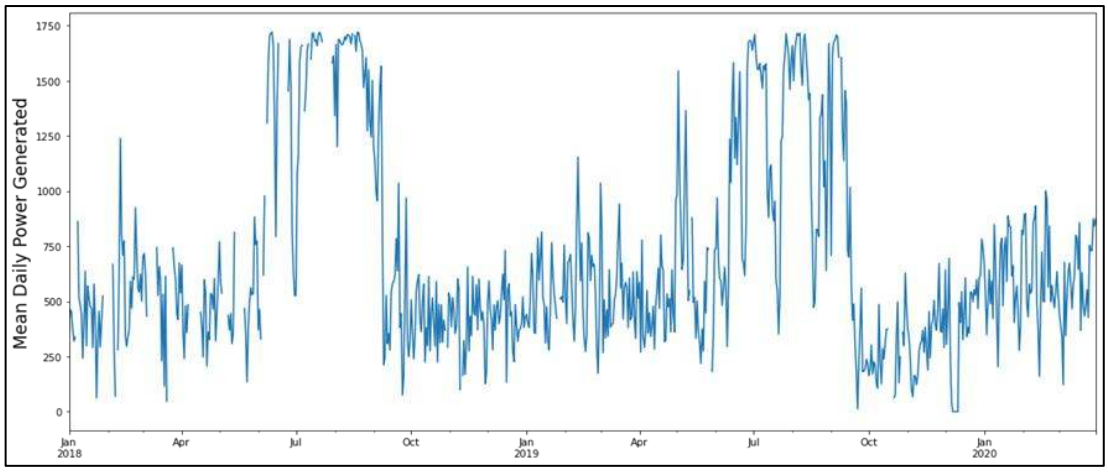

Figure 4.2. Mean daily power generated

Naturally, it's expected that a wind turbine would produce the highest power output when exposed to strong wind conditions. Therefore, the next step is to create a

graphical representation of the daily mean wind data, taking into account the visible gaps caused by missing values. In Figure 4.3 graphical analysis aims to determine whether there exists a parallel pattern in the wind data that corresponds to the observed peaks in power output, particularly during the months of July, August, and September.

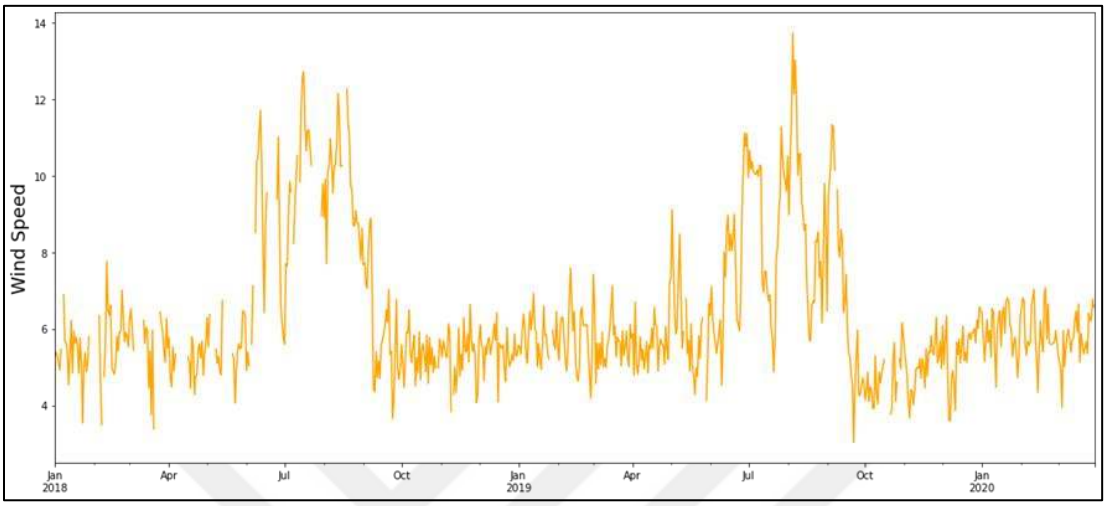

Figure 4.3. Wind speed

To gain a comprehensive understanding of the relationship between wind speed and power output, it's beneficial to graph both datasets together. However, to ensure that they are on the same scale for meaningful comparison, the wind speed values are multiplied by 100.

By overlaying both graphs, one depicting wind speed and the other power output, with the wind speed values scaled up, it becomes remarkably clear how these two variables are related. This visual representation helps illustrate the correlation between wind speed and power generation effectively.

To emphasize this correlation, the data is not only graphed at a daily level but also aggregated on a monthly basis. This allows for a more comprehensive view, highlighting the consistency of the observed pattern. The visual evidence strongly supports the notion that wind speed significantly influences power output, particularly when daily and monthly resampling values are considered together. In essence, the graphs provide a compelling visual representation of the clear and evident correlation between wind speed and power generation.

The columns that lack correlation with the Active Power column were removed, as well as certain columns that exhibit high correlation with each other. Figure 4.4 and Figure 4.5 shows the only features unrelated to the wind turbine system, such as Gear

Box Oil Temperature, are the ambient temperature, wind speed, and wind direction – all of which are relevant as this is a wind turbine designed to harness wind for power generation. It's worth noting that the correlation between wind direction and ambient temperature is relatively insignificant.

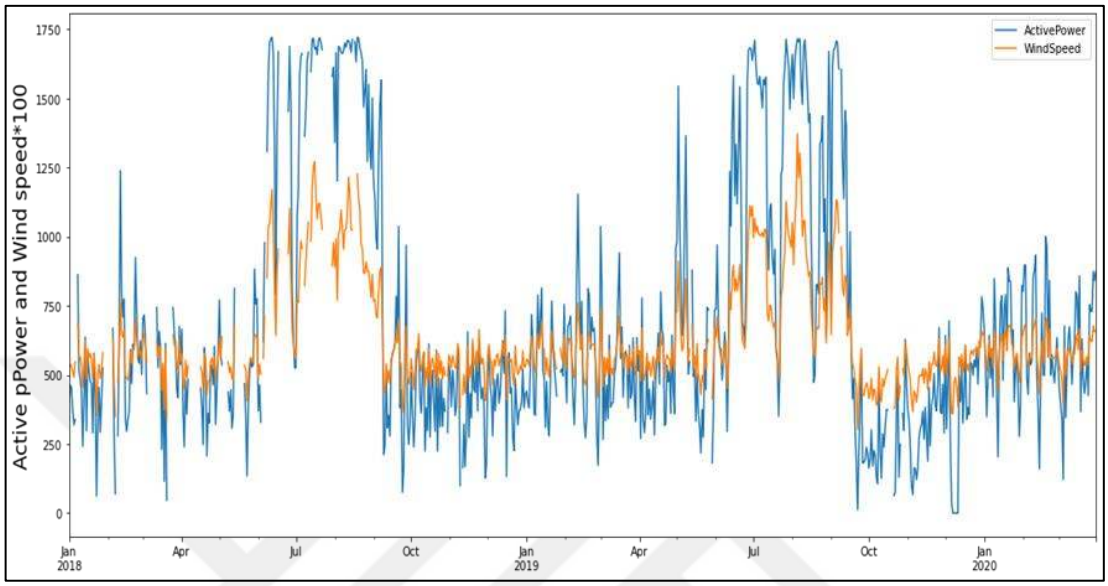

Figure 4.4. Active power and wind speed *100

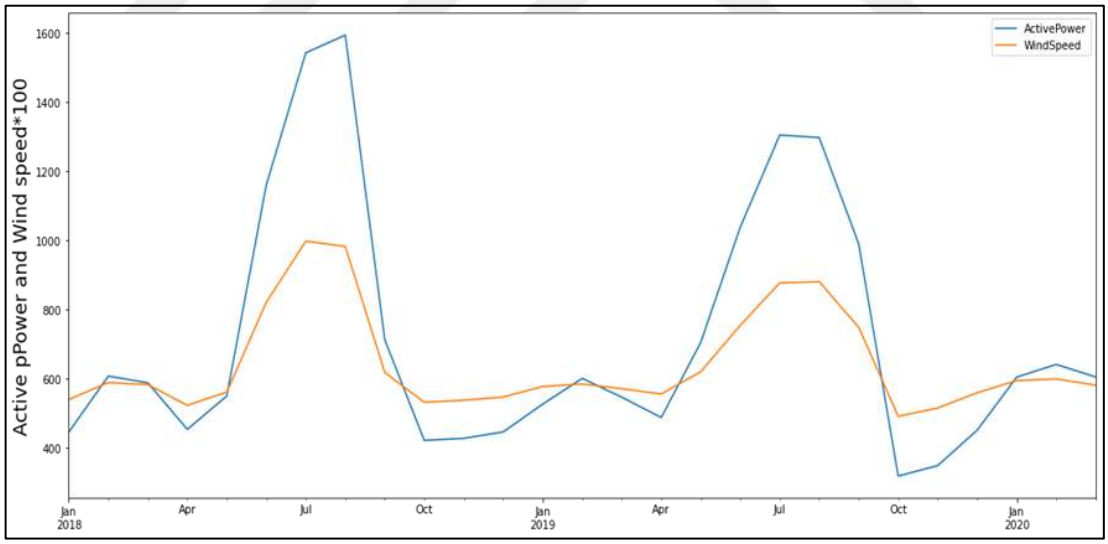

Figure 4.5. Power output and wind velocity *100

Consequently, the decision is made to retain only the wind speed column for further analysis. This choice is also facilitated by the fact that Wind Speed has the fewest missing values compared to other fields, making it a more reliable and representative parameter for examining its impact on power generation.

4.2. Pattern of Power Generation Versus Wind Speed

In Figure 4.6 and Figure 4.7, it becomes evident that the wind speed needs to reach approximately 2.5 meters per second (m/s) to initiate the wind turbine's operation and commence power generation. The maximum power generation occurs at around 8 m/s, and wind speeds exceeding this threshold do not contribute to additional power output, which likely stems from turbine-specific operational characteristics.

While there is some inherent variability or noise in the graph, the data can be reasonably described by a relatively straightforward mathematical function. Although the relationship appears nearly linear, a sigmoid function (exponential) is better suited to capture the nuances associated with both the minimum activation and maximum power generation points. This choice of function provides a more accurate representation of the observed data trends.

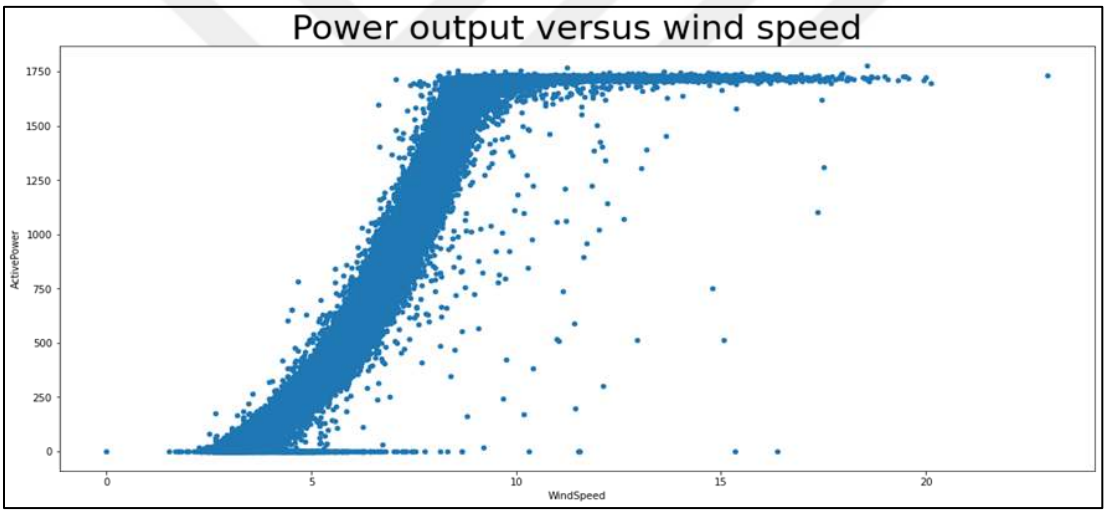

Figure 4.6. Power output versus wind speed.

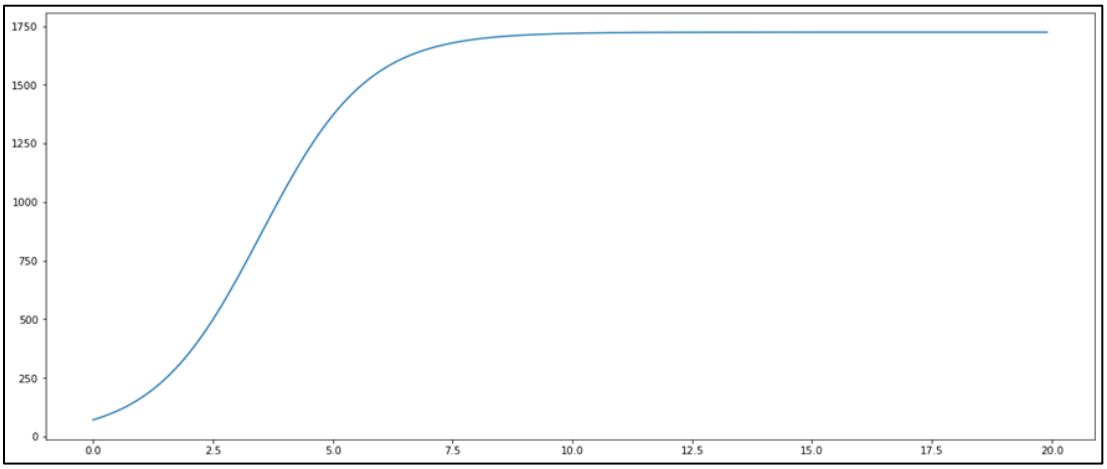

Figure 4.7. Power generated versus wind velocity.

Create a function to generate this graph and then employ the curve fit function from scipy to determine the optimal parameters. Subsequently, generate a graph that overlays the curve against the actual measured values. To enhance visibility, the transparency of the measured values is deliberately set to a very low value, making the outliers appear as if they have been attenuated.

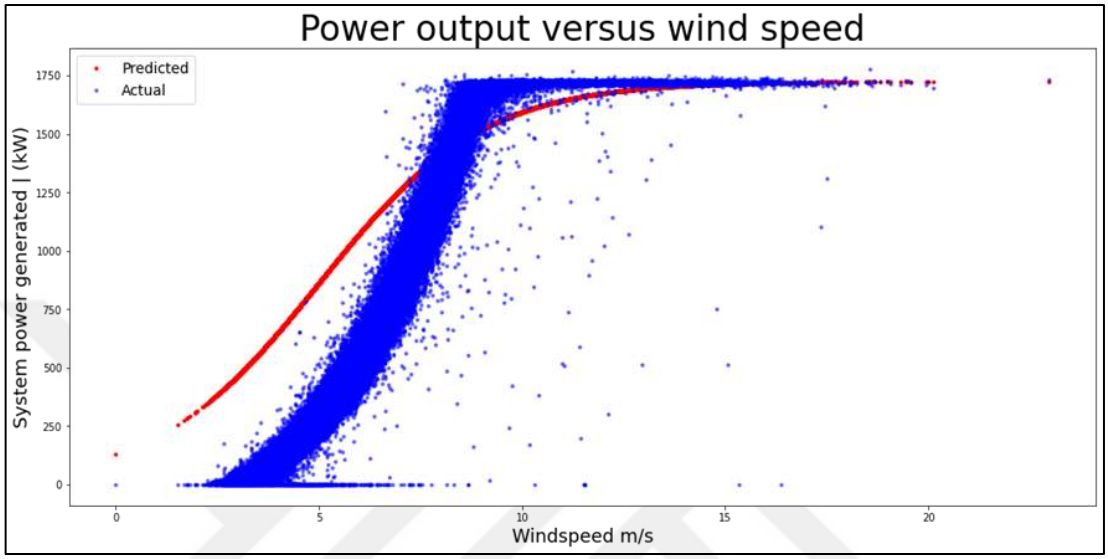

Figure 4.8. Power output versus wind speed

Accuracy metrics reveal outstanding performance, particularly highlighted by an excellent R-squared value in this approach. With the wind speed data available for the upcoming 15 days, power output can be forecasted with an impressive accuracy rate of 97%. However, it's essential to clarify that the primary objective outlined in the dataset context was focused on the development of a long-term wind forecasting technique.

Considering the dataset spans over two years and a fraction, It provides a valuable asset for predicting wind patterns and, consequently, the potential wind power generation for the subsequent 15 days. Given the inherent time-dependent nature of this analysis, marked by distinct seasonal trends illustrated in the monthly graphs (notably peaking in June, July, and August for both power generation and wind speeds), the application of specialized time series techniques becomes imperative.

Hourly power generation graph indicates the anticipated trend where windier conditions predominantly occur in the afternoon, resulting in higher power generation during those hours. However, it's important to note that power generation is feasible at any time, regardless of the specific hour.

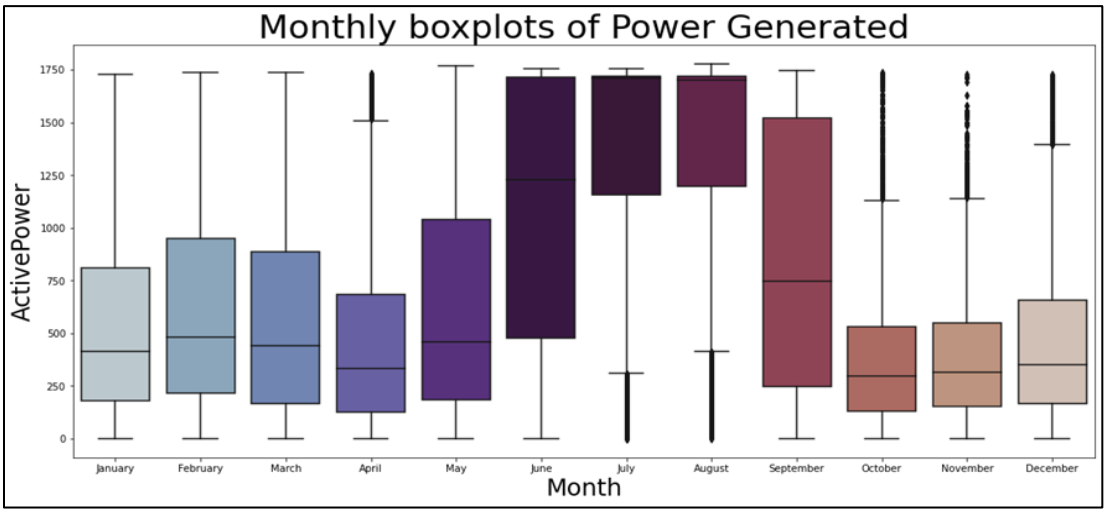

Figure 4.9. Monthly boxplots of power generated

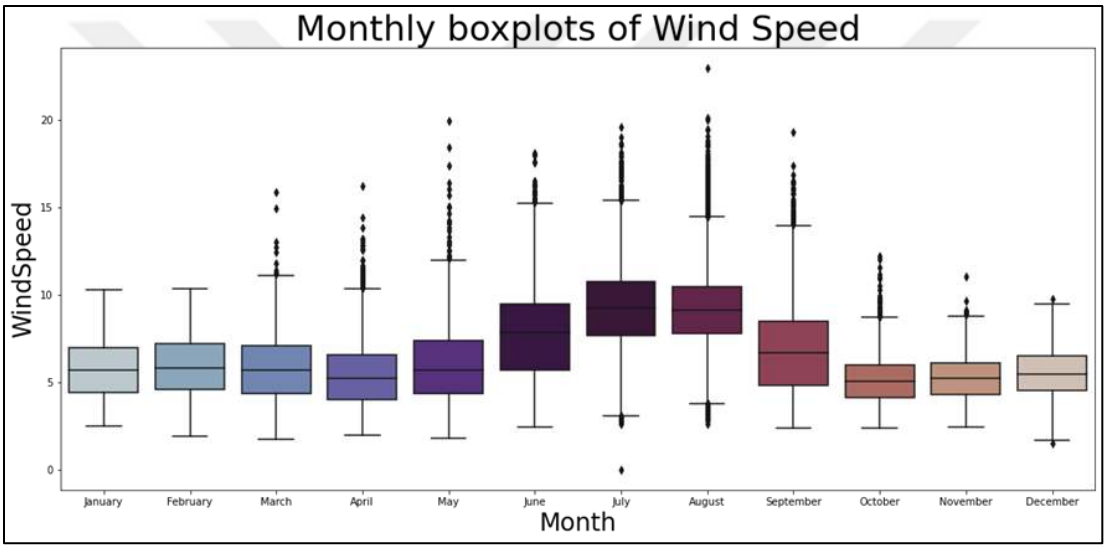

Figure 4.10. Monthly boxplots of wind Speed

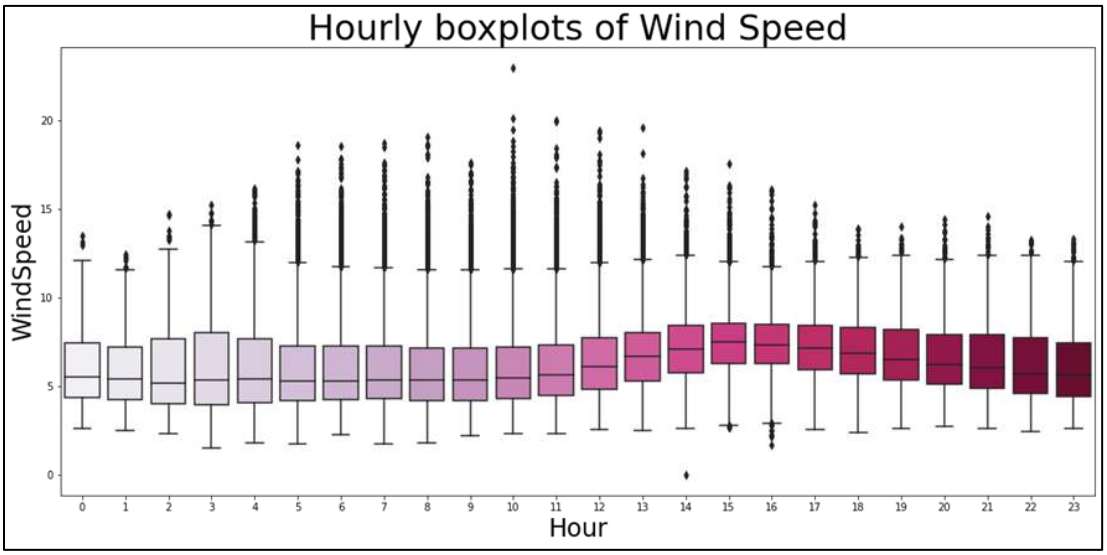

Figure 4.11. Hourly boxplots of wind speed.

4.2.1. Seasonal ARIMA (SARIMA) model

Given the evident presence of a seasonal component in the aforementioned graphs, a basic ARIMA model is unsuitable, making it necessary to employ a Seasonal ARIMA (SARIMA) model. It's worth mentioning that the SARIMA model adheres to the requirement of being univariate, which aligns with the approach of solely considering wind speed for power output analysis. The SARIMA model encompasses the core three components of ARIMA while also incorporating four supplementary seasonal variables to account for the seasonal variations.

By applying the Dicker-Fuller test, we are able to assess stationary. In this case, the test yielded an exceptionally low p-value of 1.04x10^ (-28), significantly smaller than the typical significance level of 0.05. This result indicates that despite the data displaying seasonality, it can be considered stationary for analytical purposes.

The auto_arima function can help determine the appropriate values for each of the required parameters: p, d, q, P, D, Q, and m. However, given the stationary of the data, the d value is set to 0. To begin, the data is resampled to a daily frequency, aligning with the objective of forecasting the power generated for the next 15 days.

4.2.2. Whole dataset on same plane

Figure 4.12 results obtained from training and testing on the entire dataset are impressive, which is expected since the model is essentially predicting data it has already learned from.

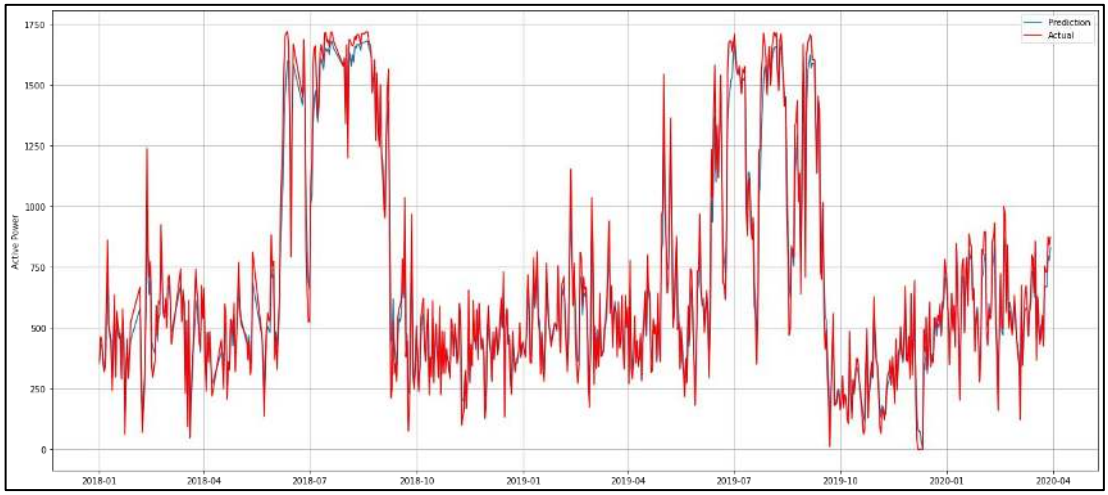

Figure 4.12. Active power

The next challenge is to forecast power generation for the upcoming 15 days. To accomplish this, three distinct groups for training and testing data are set up:

1. The entire dataset except for the last 15 days, which serves as the test set.

2. 80% of the data for training, followed by the subsequent 15 days as the test set.

3. Two-thirds of the data for training, with the next 15 days as the test set.

However, the outcomes of these forecasts are not as promising. To address this issue, further analysis will be based on the number of observations, which totals 748.

4.2.3. Graph of predicted versus actual for last 15 days of the dataset

Figure 4.13 presented here illustrates a comparison between the predicted values and the actual power generation data for the last 15 days of the dataset. This comparison provides a visual representation of how well the forecasting model aligns with real-world observations during this specific period. By plotting the predicted and actual values on the same graph, we can assess the accuracy and performance of the forecasting model. The graph visually highlights any discrepancies or areas where the predictions deviate from the actual data, helping to evaluate the model's effectiveness in forecasting power generation for this critical timeframe.

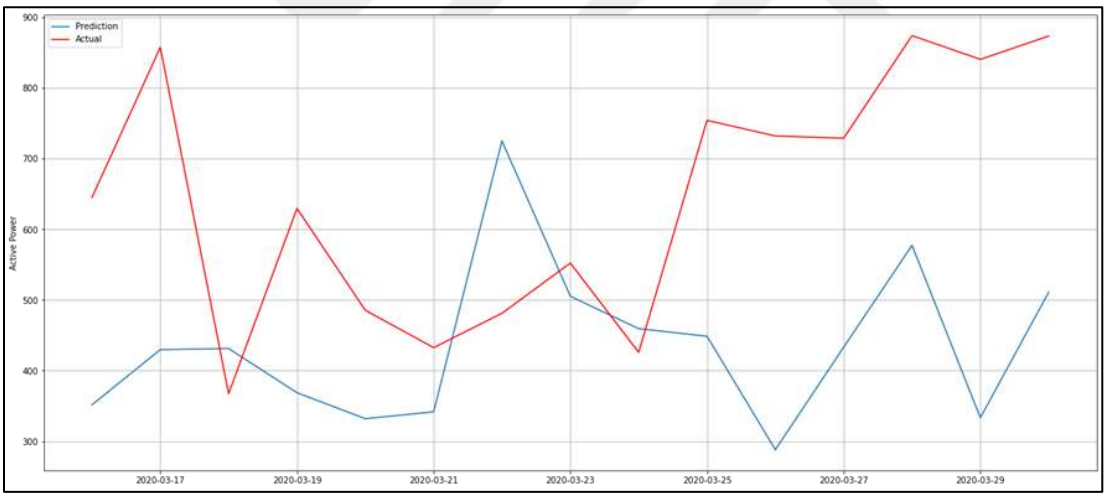

Figure 4.13. Graph of predicted versus actual for last 15 days of the dataset.

4.2.4. Graph using 80% of the dataset for training

Figure 4.14 presented here displays the results of a predictive model trained on 80% of the dataset to forecast power generation for the subsequent 15 days. This approach involves using a significant portion of the available data to train the model, enabling it to learn patterns and relationships within the dataset. The trained model is then applied to predict power generation values for the 15-day period immediately following the training data.

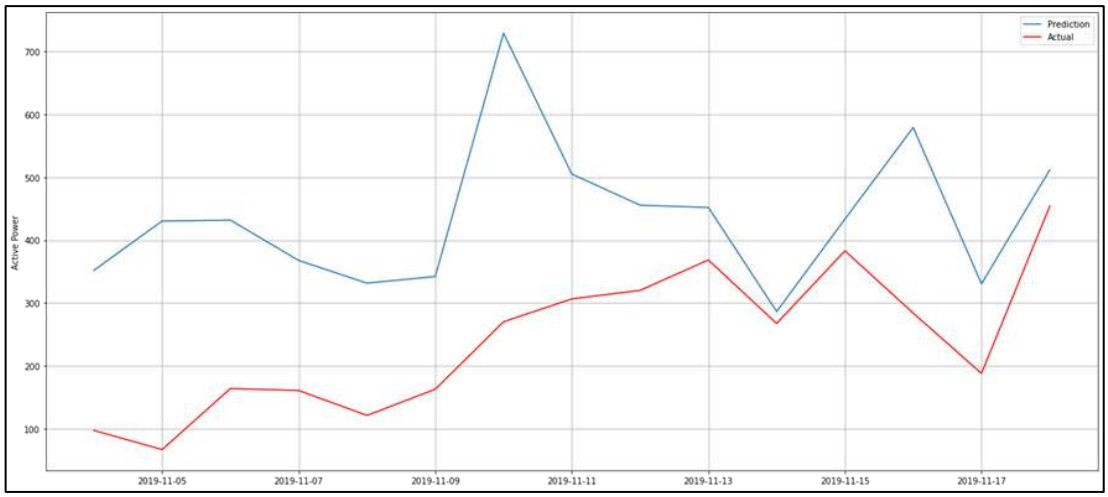

Figure 4.14. Graph using 80% of the dataset for training.

Figure 4.14 graph visually compares the predicted power generation values against the actual values observed during those 15 days. This visual representation allows us to assess the model's accuracy and its ability to make reliable forecasts for this specific time frame. By analyzing the graph, we can identify any disparities between the predicted and actual values, which provide insights into the model's performance during this particular period of interest.

The results obtained from the previous two steps, specifically for prediction intervals of 733 days and 600 days, did not yield satisfactory outcomes. Consequently, there was no further exploration conducted with the approach of using two-thirds of the training data to forecast future power generation.

In the initial two steps, it appears that the predictive model struggled to provide accurate forecasts for these extended time frames. These less favourable outcomes might be attributed to the inherent complexity of predicting power generation, which can be influenced by a multitude of variables, including weather conditions and seasonal patterns. The decision not to proceed with the 2/3 training data approach was likely based on the observed limitations in predictive accuracy during the earlier steps.

4.3. Extreme Gradient Boost (XGBOOST)

In an attempt to improve predictive accuracy, a different modelling approach was explored. XG Boost, a powerful machine learning algorithm, was employed for this purpose. In this approach, the input values (denoted as 'x') were set as Wind Speed values, while the output (denoted as 'y') was set as the corresponding Active Power values. The model was trained on the training data (X_train and y_train), which included all data

points except for the last 15 days. The testing data (X_test and y_test) consisted of the values from the final 15 days of the dataset, as described previously.

The XG Boost algorithm demonstrated notably superior performance compared to the previous modelling attempts. The predictions generated by the XG Boost model closely aligned with the actual values, as illustrated in the accompanying graph. This enhanced predictive accuracy was further substantiated by an impressive R-squared (R2) score of 0.91, which is indicative of a strong correlation between predicted and actual values. Additionally, the model yielded lower values for both Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE), signifying reduced prediction errors and enhanced accuracy.

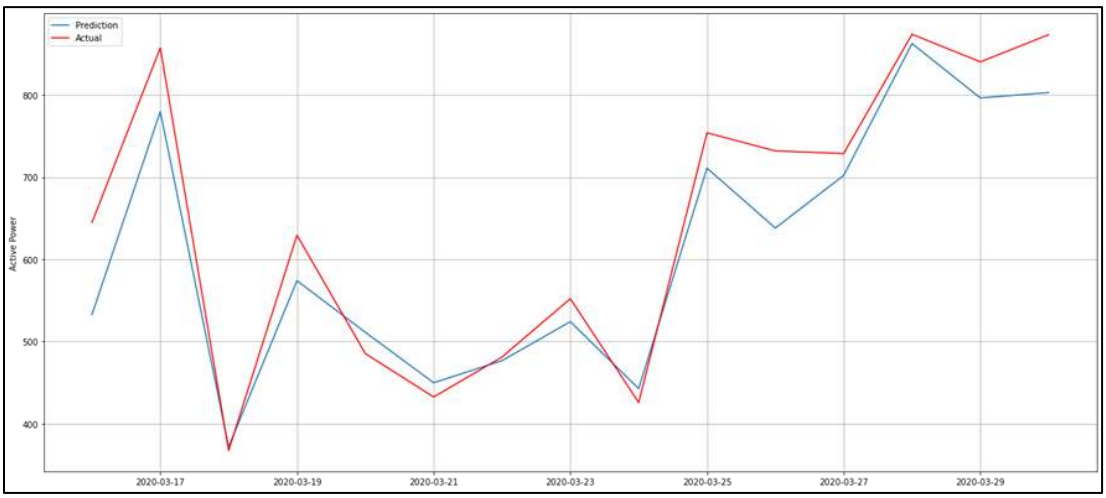

Figure 4.15. Extreme gradient boost.

4.3.1. Long Short-Term Memory (LSTM)

The next approach considered for enhancing predictive accuracy involved the use of Long Short-Term Memory (LSTM), a type of recurrent neural network (RNN). LSTM networks are well-suited for handling sequential data, making them a potential candidate for improving predictions in this context.

To prepare the data for LSTM modelling, the same training data as used in previous attempts was employed. However, there was a fundamental change in how the data was formatted. Specifically, the values for X_train and y_train were structured differently. Instead of using single rows of data to predict the next Active Power value, a sequence of multiple previous rows was utilized as input to forecast the subsequent 15 values. In this case, 35 rows of past data were used as a guide for forecasting the next 15 values. This choice was informed by the recognition of a recurring monthly pattern in the

data, which suggested that considering a longer history of values might yield improved predictions.

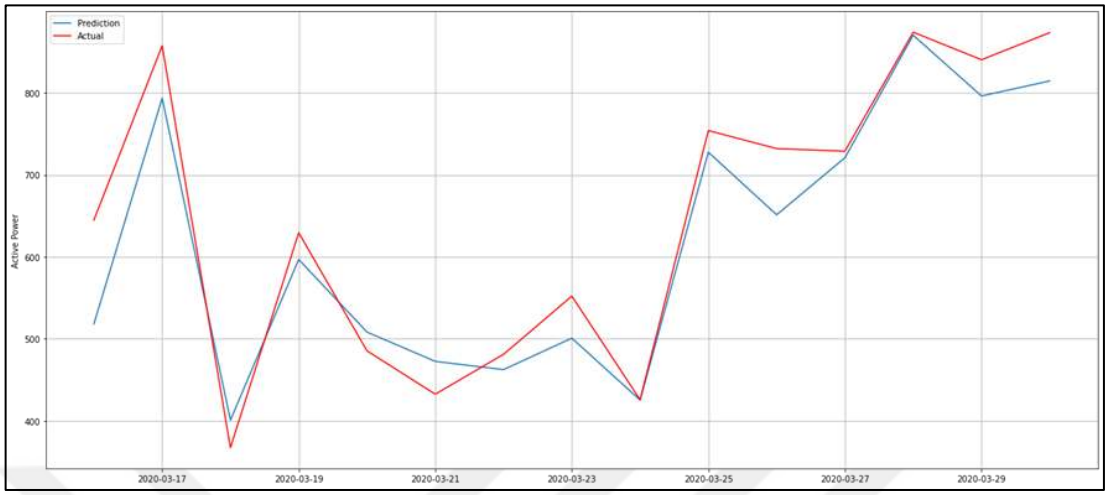

Figure 4.16. Long Short-Term Memory (LSTM).

The dataset was divided into two segments: the first 698 values were allocated for training, and the remaining 50 values were reserved for testing the LSTM model.

Despite diligent experimentation with varying the number of neurons and batch size, the LSTM model did not yield satisfactory results. It appeared challenging to achieve the desired predictive accuracy using this architecture, despite the promising nature of LSTM networks for sequential data analysis.

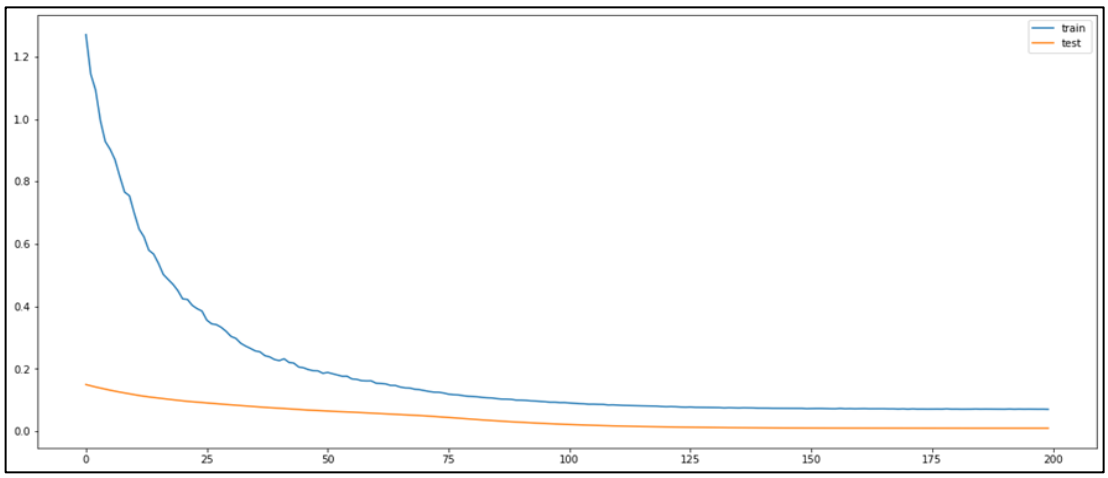

Figure 4.17. LSTM modelling

The Coefficient of determination (R-squared) = 0.08

The mean absolute error (MAE) = 144.05

The RMSE error (RMSE) = 165.32

The Mean absolute percentage error (MAPE) = 0.26

The Coefficient of Determination (R-squared), often denoted as R^2, is a statistical measure that assesses the goodness of fit of a regression model to the actual data points. It quantifies the proportion of the variance in the dependent variable (in this case, likely Active Power) that can be explained by the independent variables (e.g., Wind Speed). An R^2 value of 0.08 suggests that only 8% of the variability in Active Power can be accounted for by Wind Speed, which indicates a relatively weak linear relationship between these variables.

The Mean Absolute Error (MAE) is a metric that calculates the average absolute differences between the predicted values and the actual values. In this context, with an MAE of 144.05, it means that, on average, the model's predictions for Active Power are off by approximately 144.05 units. MAE is a useful measure because it provides insight into the magnitude of errors without considering their direction.

The Root Mean Squared Error (RMSE) is another error metric that calculates the square root of the mean of the squared differences between predicted and actual values. It is closely related to the MAE but tends to give more weight to larger errors. With an RMSE of 165.32, it indicates the typical error in the predictions is around 165.32 units, which is slightly larger than the MAE.

The Mean Absolute Percentage Error (MAPE) expresses the average relative error as a percentage of the actual values. An MAPE of 0.26 means that, on average, the model's predictions for Active Power deviate from the actual values by approximately 0.26% of the true values. This metric is valuable for understanding the percentage error in the predictions, making it easier to interpret the practical significance of errors.

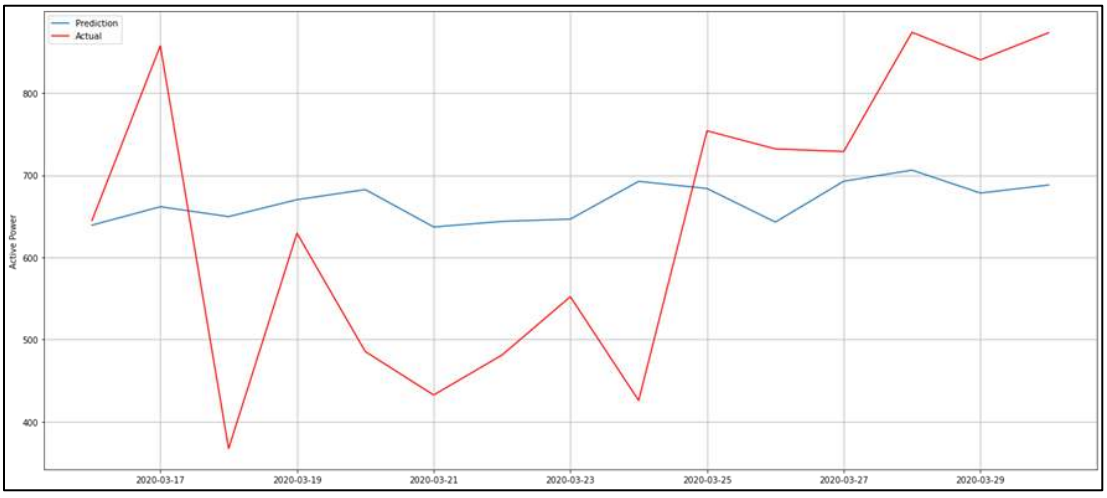

Figure 4.18. Performance of a predictive model

In Figure 4.18 these metrics collectively assess the performance of a predictive model. In this case, the low R-squared and relatively high MAE and RMSE suggest that the model's predictions for Active Power based on Wind Speed have limitations and may not be highly accurate.

4.4. Evaluation of The Performance of The Forecasting Models

Initially, the research aimed to find the most suitable forecasting method for predicting power output based on the available data, which showed a clear seasonal pattern. The SARIMA model, designed for time series data with seasonal components, was considered a promising approach. However, this expectation was not met, suggesting that the seasonal nature of the data was not adequately captured by this method.

Similarly, the research explored the use of Long Short-Term Memory (LSTM) neural networks, which are well-suited for sequential data like time series. However, this approach also fell short of expectations, indicating that more complex neural networks might not necessarily yield better results for this specific dataset.

Surprisingly, it was the simpler machine learning methods, specifically the Random Forest and XG Boost Regressor that proved to be the most effective for forecasting power output. Among these, the Random Forest Regressor exhibited a slightly better R-squared value, while XG Boost had a slightly higher Mean Absolute Percentage Error (MAPE). These differences were minor and required detailed analysis to distinguish.

The research findings suggest that either the XG Boost or Random Forest Regressor can predict power output for the next 15 days with an accuracy of approximately 94% or a mean average percentage error of 6%. Furthermore, roughly 91% of the variance in the data is explained by either method. Importantly, both of these approaches achieved these results in a much more time-efficient manner compared to the initially considered models, demonstrating their practical suitability for this forecasting task.

Not specifying a "random state" for each method can lead to variations in the results obtained when running the same machine learning algorithms. In the submitted notebook, these variations may because discrepancies compared to the figures presented. When working with machine learning algorithms, randomness often plays a role in processes like data splitting, initialization, or hyper parameter tuning. Without setting a specific "random state," these random components can yield different outcomes each time the algorithm is executed.

By setting a "random state," you essentially fix the initial conditions of the random processes, ensuring that the same sequence of random events occurs consistently across runs. This is crucial for reproducibility and allows others to obtain the same results when working with your code.

Table 4.1. Model performance comparison for time series forecasting.

Method	R squared	RMSE	MAE	MAPE
SARIMAX (733 days training)	-1.890	225.06	96.86	0.340
SARIMAX (600 days training)	-3.260	195.04	227.71	1.260
Extreme gradient boost (XG Boost)	0.906	42.07	52.95	0.062
Random Forest Regressor	0.915	40.54	50.99	0.064
LSTM	-0.190	162.71	187.76	0.300

The table provides a comparison of different methods used for forecasting power output, measured by several evaluation metrics. Each method is assessed based on its ability to predict power output, with higher values for R-squared, and lower values for RMSE, MAE, and MAPE indicating better predictive performance.

1. SARIMAX (733 days training): This method uses a Seasonal Autoregressive Integrated Moving Average model with exogenous variables. However, it performs poorly, as evidenced by the negative R-squared and high RMSE and MAE values. The negative R-squared suggests that the model explains less variance in the data than a horizontal line. Additionally, the relatively high MAPE indicates that the predicted values have a moderate percentage error compared to the actual values.

2. SARIMAX (600 days training): Similar to the previous method, this SARIMAX model also struggles, with a negative R-squared and high RMSE, MAE, and MAPE values. The decrease in training data duration did not lead to improved forecasting accuracy; instead, the model's performance deteriorated.

3. Extreme Gradient Boost (XG Boost): XG Boost, a gradient boosting algorithm, significantly outperforms the SARIMAX models. It achieves a high R-squared value of 0.906, indicating that it explains a substantial portion of the variance in the data. The low RMSE, MAE, and MAPE values suggest that the predicted values closely align with the actual values, with a minimal percentage error.

4. Random Forest Regressor: Similar to XG Boost, the Random Forest Regressor performs well, with a high R-squared value of 0.915. It also demonstrates low RMSE, MAE, and MAPE values, indicating accurate and reliable power output predictions.

5. LSTM (Long Short-Term Memory): The LSTM neural network performs the worst among the methods evaluated. It exhibits a negative R-squared value and relatively high RMSE, MAE, and MAPE values. This suggests that the LSTM model struggles to capture the underlying patterns in the data and provides less accurate predictions.

The results of wind energy prediction using the Long Short-Term Memory (LSTM) model present an intriguing contrast with the other methods employed in the study. The LSTM model, known for its prowess in sequence data modelling, appears to struggle when applied to this specific wind energy prediction task.The R-squared value for the LSTM model is negative (-0.19), indicating that the model's predictions are worse than simply using a horizontal line as a prediction. This suggests that the LSTM fails to capture the underlying patterns and relationships within the wind energy data effectively. The LSTM model exhibits relatively high RMSE (Root Mean Squared Error) and MAE (Mean Absolute Error) values, which suggest significant prediction errors. This is in stark contrast to models like Extreme Gradient Boosting (XG Boost) and Random Forest Regressor, which yielded substantially lower RMSE and MAE values, indicating superior predictive accuracy. Although the MAPE (Mean Absolute Percentage Error) value for the LSTM model is relatively low compared to the SARIMAX models, it is considerably higher than that of XG Boost and Random Forest Regressor. This indicates that while the percentage error may be smaller, the absolute error remains relatively high for LSTM predictions. In conclusion, the LSTM model appears to be ill-suited for this wind energy prediction task based on these results. Its inability to capture the complex relationships in the data and its relatively high prediction errors suggest that other methods, such as XG Boost and Random Forest Regressor, are more appropriate choices for accurate wind energy forecasting in this context. These findings underscore the importance of carefully selecting the modelling approach and highlight that LSTM's strengths in sequential data may not always translate into superior performance for all-time series forecasting problems.

5. CONCLUSION AND RECOMMENDATION

5.1. Initial Data Examination and Processing

The initial phase involved exploring different modeling approaches after the SARIMA model failed. The alternatives included XG Boost, Random Forest Regressor, and LSTM. This section also delved into the dataset's properties using visualization and statistical analysis. It highlighted the prevalence of missing cells, necessitating careful data handling.

5.2. Data Overview and Preprocessing

This phase detailed the process of data importation, recognizing the date column and handling duplicate entries. It mentioned the exploration through Pandas profiling and boxplots, identifying redundant variables like "Control Box Temperature" and "WTG" for potential exclusion. Moreover, it outlined the key variables such as "Active Power," "Ambient Temperature," and others, discussing their correlations, distributions, and the challenge of missing values.

5.3. Methodology and Modeling Comparison

The final phase encapsulated the methodology used, emphasizing the importance of addressing missing data points and anomalies for accurate analysis. It discussed the rigorous cleaning process, model selection (SARIMA, XG Boost, Random Forest Regressor, LSTM), and their respective performance. Additionally, it underscored the significance of data accuracy, the impact of wind speed on power output, and the necessity for varied modeling methods to capture wind energy dynamics effectively.

These three subsections offer a comprehensive view of the data examination, preprocessing, and modeling approaches taken to forecast wind energy output, laying the groundwork for informed decision-making and sustainable energy management.

5.4. Recommendation

1. Use more advanced data pre-processing methods to address missing values, outliers, and noise to improve dataset quality and dependability for accurate modelling.

2. Explore hybrid modeming technologies that combine classic statistical methodologies and machine learning algorithms for more accurate wind energy predictions.

3. Add meteorological and geographical elements that affect wind energy generation to feature engineering methodologies to better understand power output.

4. Develop more interpretable models to better comprehend the relationship between relevant variables and wind energy generation, enabling more informed renewable energy policy and decision-making.

5. Ensemble learning methods like model averaging and stacking can use the diversity of forecasting models to increase prediction accuracy by minimizing model flaws, resulting in more accurate and consistent forecasts.

6. Use real-time data streams and advanced monitoring systems to capture dynamic weather patterns and environmental conditions to help forecasting models adjust quickly and provide accurate predictions for energy management.

7. Conduct a rigorous sensitivity study to determine the forecasting models' robustness to input parameter adjustments and identify the most relevant variables to better understand their effects on wind energy generation and forecasting accuracy.

8. To ensure the reliability and generalizability of forecasting models across different geographical locations and environmental conditions, prioritize rigorous model validation and verification, including comprehensive back-testing and validation on diverse datasets.

9. Long-term wind energy generation forecasting studies can help plan and build infrastructure for sustainable energy production and consumption in the face of changing climate dynamics and global energy demands.

Encourage academic institutions, industry stakeholders, and government agencies to collaborate on research projects to share knowledge, data, and innovative solutions for wind energy forecasting technologies and sustainable energy practices worldwide.

6. REFERENCE

Abdoos, A.A. (2016). A new intelligent method based on combination of VMD and ELM for short term wind power forecasting. *Neurocomputing, 203*, 111-120.

Aguilar, S., Souza, R.C. & Pensanha, J.F. (2014, October). *Predicting probabilistic wind power generation using nonparametric techniques* [Conference presentation]. In 2014 International Conference on Renewable Energy Research and Application (ICRERA)*,* 709-712.

Amjady, N., Keynia, F. & Zareipour, H. (2011). Short-term wind power forecasting using ridgelet neural network. *Electric Power Systems Research, 81*(12), 2099-2107.

Andrew, K., & Zhe, S. (2010). Design of wind farm layout for maximum wind energy capture. *Renewable Energy, 35*, 685-694.

Andrew, K., Haiyang, Z., & Zhe, S. (2009). Wind Farm Power Prediction: A Data-Mining Approach. *Wind Energy*, (12), 275–293.

Andrew, K., Haiyang, Z., & Zhe, S. Power optimization of wind turbines with data mining and evolutionary computation. *ELSEVIER*.

Asis, S., & Dhiren, K. B. (2012). Wind Turbine Blade Efficiency and Power Calculation with Electrical Analogy. *International Journal of Scientific and Research Publications, 2*(2).

Asis, S., Dhiren, K. B. (2012). Wind Turbine Blade Efficiency and Power Calculation with Electrical Analogy. *International Journal of Scientific and Research Publications, 2* (2).

Ayadi, F., Colak, I., Garip, I. & Bulbul, H.I. (2020, June). *Impacts of Renewable Energy Resources in Smart Grid* [Conference presentation]. 8th International Conference on Smart Grid (ICSmartGrid), 183-188.

Banna, H.U., Luna, A., Ying, S., Ghorbani, H. & Rodriguez, P. (2014, October). *Impacts of wind energy in-feed on power system small signal stability* [Conference presentation]. International Conference on Renewable Energy Research and Application (ICRERA), 615-622.

Barthelmie, R. J., Folkerts, L., Larsen, G. C., Rados, K., Pryor, S. C., Frandsen, S. T., Lange, B., & Schepers, G. (2006). Comparison of Wake Model Simulations with Offshore Wind Turbine Wake Profiles Measured with Sodar. *Journal of Atmospheric and Oceanic Technology*, (7), 888-901.

Barthelmie, R. J., Frandsen, S. T., Nielsen, M. N., Pryor, S. C., Rethore, P.E., & Jørgensen, H. E. (2007). Modelling and Measurements of Power Losses and Turbulence Intensity in Wind Turbine Wakes at Middelgrunden Offshore Wind Farm. *Wind energy*.

Carpinone, A., Giorgio, M., Langella, R. & Testa, A. (2015). Markov chain modeling for very-short-term wind power forecasting. *Electric Power Systems Research, 122*, 152-158.

Catalão, J. P. S., Pousinho, H. M. I., & Mendes, V. M. F. (2011). Hybrid Wavelet-PSO-ANFIS Approach for Short-Term Wind Power Forecasting in Portugal. *Transactions on Sustainable Energy, 2* (1), 50-59.

Cleve, J., Greiner, M., Envoldsen, P., Birkemose, B., & Jensen, L. (2009). Model based Analysis of Wake-flow Data in the Nysted Offshore Wind Farm. *Wind Energy,* (2), 125-135.

Crasto, G., & Gravdahl, A. R. (2008). *CFD wake modeling using a porous disc* [Conference presentation]. European Wind Energy Conference and Exhibition. Brusssels, Belgium.

Crespo, A., Hernandez, J., & Frandsen, S. (1999). Survey of Modelling Methods for Wind Turbine Wakes and Wind Farms. *Wind Energy*. 1-24.

Diaconu, S., Onea, F., & Rusu, E. (2013). Evaluation of The Nearshore Impact of a Hybrid Wave-Wind Energy Farm. *International Journal of Education and Research, 1*(2).

Diaconu, S., Onea, F., Rusu, E. (2012, February). Evaluation of The Nearshore Impact of a Hybrid Wave-Wind Energy Farm. *International Journal of Education and Research, 1* (2).

Dolara, A., Gandelli, A., Grimaccia, F., Leva, S. & Mussetta, M. (2017, November). *Weather-based machine learning technique for Day-Ahead wind power forecasting* [Conference presentation]. 6th international conference on renewable energy research and applications (ICRERA), 206-209.

Duan, J., Wang, P., Ma, W., Tian, X., Fang, S., Cheng, Y., Chang, Y. & Liu, H. (2021). Short-term wind power forecasting using the hybrid model of improved variational mode decomposition and Correntropy Long Short-term memory neural network. *Energy, 214*, 118980.

Erik, L. P. *The new generation of tools for prediction of wind power potential and site selection*. Technical University of Denmark, DTU Wind Energy.

EWEA (2009). *Wind Energy – The Facts*. Earthscan

Fadare, D. A. (2008). A Statistical Analysis of Wind Energy Potential in Ibadan, Nigeria, Based on Weibull Distribution Function. *The Pacific Journal of Science and Technology, 9*(1).

Frandsen, S. (1992). On the wind speed reduction in the center of large clusters of wind turbines. *Journal of Wind Engineering and Industrial Aerodynamics*, (39), 251-265.

Frandsen, S., Barthelmie, R., Pryor, S., Rathmann, O., Larsen, S., Hojstrup, J., & Thogersen, M. (2006). Analytical Modelling of Wind Speed Deficit in Large Offshore Wind Farms. *Wind Energy.* (9), 39-53

Gao, B., Huang, X., Shi, J., Tai, Y., & Zhang, J. (2020). Hourly forecasting of solar irradiance based on CEEMDAN and multi-strategy CNN-LSTM neural networks. *Renewable Energy*, *162*, 1665-1683.

Gonzalez, F., Longatt, P., Wall, & Terzija, V. Wake effect in wind farm performance study- state and dynamic behavior. *ELSEVIER*.

Harrouz, A., Colak, I. and Kayisli, K. (2019, November). *Energy Modeling Output of Wind System based on Wind Speed* [Conference presentation]. 8th International Conference on Renewable Energy Research and Applications (ICRERA), 63-68.

Hau, E. (2013). *Wind turbines: fundamentals, technologies, application, economics*. Springer Science & Business Media.

Ho Lip, W., Shaharin, I., Sutarji, K., Che Musa C. O., & Ahmad, M. A. Review of Offshore Wind Energy Assessment and Siting Methodologies for Offshore Wind Energy Planning in Malaysia. *American International Journal of Contemporary Research, 2*(12).

Ishihara, T., Yamaguchi, A., & Fujino, Y. (2004). Development of a New Wake Model Based on a Wind Tunnel Experiment. *Global Wind Power*.

Izelu, C. O., Agberegha, O. L., Oguntuberu, S., & Olusola, B. (2013). Wind Resource Assessment for Wind Energy Utilization in Port Harcourt, River State, Nigeria, Based on Weibull Probability Distribution Function. *International Journal of Renewable Energy Research, 3*(1).

Jaesung, J., & Robert, P. B. *Current Status and Future Advances for Wind Speed and Power Forecasting*.

Jensen NO (1983). *A note on Wind Generator Interaction*. Riso National Laboratory, Roskilde, Denmark

Katic, I., Hojstrup, J., & Jensen, N.O. (1986). *A simple model for cluster efficiency* [Conference presentation]. Proceedings of the European Wind Energy Conference and Exhibition, 407- 410.

Kunal, K. B., Souvnik, R., & Harinarayana, T. (2013). Optimization in Site Selection of Wind Turbine for Energy Using Fuzzy Logic System and GIS— A Case Study for Gujarat. *Open Journal of Optimization*, (2),116- 122.

Kusiak, A., & Song, Z. (2010, March). Design of wind farm layout for maximum wind energy capture. *Renewable Energy*, *35*(3), 685–694. https://doi.org/10.1016/j.renene.2009.08.019.

Lionel, F., J´er´emie, J., & George, K. (2008). *Data mining for wind power forecasting* [Conference presentation]. European Wind Energy Conference, Brussels, Belgium.

Lionel, F., Jeremie, J., & George, K. (2008). *Data mining for wind power forecasting* [Conference presentation]. European Wind Energy Conference. Brussels, Belgium.

Louka, P., Galanis, G., Siebert, N., Kariniotakis, G., Katsafados, P., Pytharoulis, I., & Kallos, G. (2008). Improvements in wind speed forecasts for wind power prediction purposes using Kalman filtering. *Journal of Wind Engineering and Industrial Aerodynamics*, *96*(12), 2348-2362.

Luo, W., Liu, W., & Gao, S. (2017, July). *Remembering history with convolutional LSTM for anomaly detection* [Conference presentation]. IEEE International conference on multimedia and expo (ICME),7, 439-444.

Markus, W., Kalyan, V., Frank, N., & Una-May, O. R. *Optimizing the Layout of 1000 Wind Turbines*.

Mechali. M., Barthelmie, R., Frandsen, S., Jensen, L., & Rethore, P. E. (2006). [Conference presentation]. EWEC 2006, 10, Athens, Greece.

Memarzadeh, G. & Keynia, F. (2020). A new shortterm wind speed forecasting method based on finetuned LSTM neural network and optimal input sets. *Energy Conversion and Management, 213*, 112824.

Munir, A. N. & Balwois, M. C. (2010). Wind Speed Prediction by Different Computing Techniques, 25, 29.

Murali, R. M., Vidya, P. J., Poonam, M. & Seelam, J. K. (2014). Site selection for offshore wind farms along the Indian coast. *Indian Journal of Marine Sciences, 43*(7).

Murali, R. M., Vidya, P. J., Poonam, M., & Seelam, J. K. (2014). Site selection for offshore wind farms along the Indian coast. *Indian Journal of Marine Sciences, 43*(7).

Nielsen, T.S., Madsen, H., Nielsen, H.A., Pinson, P., Kariniotakis, G., Siebert, N., Marti, I., Lange, M., Focken, U., Bremen, L.V. & Louka, G. (2006, February). Short-term wind power forecasting using advanced statistical methods.

O'Boyle, L., Elsäßer, B., & Whittaker, T. (2017). Experimental measurement of wave field variations around wave energy converter arrays. *Sustainability*, *9*(1), 70.

Oliver, K., Fabian, G., Justin, H., Jendrik, P., & Nils Andre, T. *A Framework for Data Mining in Wind Power Time Series*. www.windml.org.

Osório, G.J., Matias, J.C. & Catalão, J.P. (2014, August). *Hybrid evolutionary-adaptive approach to predict electricity prices and wind power in the shortterm* [Conference presentation]. In 2014 Power Systems Computation Conference, 1-7.

Pinson, P., Ranchin, T., & Kariniotakis, G. (2004). *Short-term Wind Power Prediction for Offshore Wind Farms - Evaluation of Fuzzy-Neural Network Based Models* [Conference presentation]. in Proc. of the 2004 Global Wind Power Conference, 28-31, Chicago.

Rakeshchandra, D., Sailaja Kumari, M., & Sydulu, M. (2013). *A Detailed Literature Review on Wind Forecasting* [Conference presentation]. International Conference on Power, Energy and Control (ICPEC).

Rakeshchandra, D., Sailaja, M. K., & Sydulu, M. (2013). *A Detailed Literature Review on Wind Forecasting* [Conference presentation]. International Conference on Power, Energy and Control (ICPEC).

René, J., Bernhard, L., & Kurt, R. *Wind Power Prediction with Optimization and Clustering Techniques*.

Righter, R.W. (1996). *Wind Energy in America: A History*. University of Oklahoma Press

RWE npower renewables Mechanical and Electrical Engineering Power Industry (2007). *Wind Turbine Power Calculations*. published by Elsevier Ltd.

Saidi, A., Harrouz, A., Colak, I., Kayisli, K. & Bayindir, R. (2019, December). *Performance Enhancement of Hybrid Solar PV-Wind System Based on Fuzzy Power Management Strategy: A Case Study* [Conference presentation]. In 2019 7th International Conference on Smart Grid (icSmartGrid), 126-131).

Senthil, K., Daphne, L. A Survey on Data Mining Techniques Used for Wind Power Forecasting. *International Journal of Wind and Renewable Energy. 1*(2), 105-107.

Shahid, F., Zameer, A., Mehmood, A. & Raja, M. A. Z. (2020). A novel wavenets long short-term memory paradigm for wind power prediction. *Applied Energy, 269*, 115098.

Shahriar, A. *A Survey on Recent Off-Shore Wind Farm Layout Optimization Methods*.

Shivam, K., Tzou, J. C., & Wu, S. C. (2020). Multi-step short-term wind speed prediction using a residual dilated causal convolutional network with nonlinear attention. *Energies*, *13*(7), 1772.

Simon, G., & Bruce, S. (2012). Wind Turbine Condition Assessment Through Power Curve Copula Modeling. *IEEE TRANSACTIONS ON SUSTAINABLE ENERGY, 3*(1).

Simon, G., Bruce, S. (2012). Wind Turbine Condition Assessment Through Power Curve Copula Modeling. I*EEE TRANSACTIONS ON SUSTAINABLE ENERGY, 3*(1).

Srivastava, T., Vedanshu, S., & Tripathi, M.M. (2020). Predictive analysis of RNN, GBM and LSTM network for short-term wind power forecasting. *Journal of Statistics and Management Systems, 23*(1), 33-47.

Sun, Z. & Zhao, M. (2020). Short-term wind power forecasting based on VMD decomposition, ConvLSTM networks and error analysis. *IEEE Access, 8*, 134422-134434.

Syu, Y.D., Wang, J.C., Chou, C.Y., Lin, M.J., Liang, W.C., Wu, L.C. & Jiang, J.A. (2020, March). *UltraShort-Term Wind Speed Forecasting for Wind Power Based on Gated Recurrent Unit* [Conference presentation]. In 2020 8th International Electrical Engineering Congress (iEECON), 1-4.

Toubeau, J.F., Dapoz, P.D., Bottieau, J., Wautier, A., De Grève, Z. & Vallée, F. (2021). Recalibration of recurrent neural networks for short-term wind power forecasting. *Electric Power Systems Research, 190*, 106639.

Venayagamoorthy, G.K., Rohrig, K. & Erlich, I. (2012). One step ahead: short-term wind power forecasting and intelligent predictive control based on data analytics. *IEEE Power and Energy Magazine, 10*(5), 70-78.

Vermeer, L. J., Sørensen, J. N., & Crespo, A. (2003). Wind turbine wake aerodynamics. *Progress in aerospace sciences*, *39*(6-7), 467-510.

Vinhoza, A., & Schaeffer, R. (2021). Brazil's offshore wind energy potential assessment based on a Spatial Multi-Criteria Decision Analysis. *Renewable and Sustainable Energy Reviews*, *146*, 111185.

Wang, J., Hu, J., & Ma, K. (2016). Wind speed probability distribution estimation and wind energy assessment. *Renewable and sustainable energy Reviews*, *60*, 881-899.

Wang, X., Guo, P., & Huang, X. (2011). A review of wind power forecasting models. *Energy procedia*, *12*, 770-778.

Werle, M. J. (2008). *A New Analytical Model for Wind Turbine Wakes*. Flo Design Inc., Wilbraham, MA.

Wing, Y. K., Peter, Y. Z., David, R., Joaquin, M., Michael, M. & Cristina, A. (2012). *Wind Farm Layout Optimization Considering Energy Generation and Noise Propagation* [Conference presentation]. Proceedings of the ASME 2012 International Design Engineering Technical Conferences & Computers and Information in Engineering Conference IDETC/CIE, 12-15, Chicago, IL, USA.

Wing, Y. K., Peter, Y. Z., David, R., Joaquin, M., Michael, M., & Cristina, A. (2012). *Wind Farm Layout Optimization Considering Energy Generation and Noise Propagation* [Conference presentation]. Proceedings of the ASME 2012 International Design Engineering Technical Conferences & Computers and Information in Engineering Conference IDETC/CIE, 12-15, Chicago, IL, USA.

Xu, Q., He, D., Zhang, N., Kang, C., Xia, Q., Bai, J. & Huang, J. (2015). A short-term wind power forecasting approach with adjustment of numerical weather prediction input by data mining. *IEEE Transactions on sustainable energy, 6*(4), 1283- 1291.

Yoon, S. J., & Kun, H. H. (2013). A Study of Time Predication Algorithm for Wind Power Generation Estimation. *Journal of Next Generation Information Technology (JNIT), 4* (8).

Yoon, S. J., & Kun, H. H. A Study of Time Predication Algorithm for Wind Power Generation Estimation. *Journal of Next Generation Information Technology(JNIT), 4*(8).

Zeng, J. & Qiao, W. (2011, March). *Support vector machine-based short-term wind power forecasting* [Conference presentation]. In 2011 IEEE/PES Power Systems Conference and Exposition.1-8.

CURRICULUM VITAE

STUDENT INFORMATION	
Name/Surname:	Ali Abdulrahman Hussein Salihi
Nationality:	Iraq
Orcid No:	0000-0002-3428-0376

SCHOOL INFORMATION	
Undergraduate Study	
University:	University Of Kirkuk
Faculty:	Faculty Of Engineering
Department:	Mechincal Engineerig
Graduation Year:	2016-2017
Graduate Study	
University:	Kırşehir Ahi Evran University
Institute:	Institute Of Natural And Applied Sciences
Department:	Mechincal Engineering
Graduation Year:	2024

Tezden Üretilen Makaleler ve Bildiriler
Salihi, A.A. & Danışmaz, M. (2023). A Comparative Study on Wind Power Forecasting Models Based on the Use of LSTM. *Tuijin Jishu/Journal of Propulsion Technology,* 44(*5*)*:* 1866-1877. https://doi.org/10.52783/tjjpt.v44.i5.2874