TECHNICAL UNIVERSITY OF CRETE

SCHOOL OF PRODUCTION ENGINEERING AND MANAGEMENT

Predictive Maintenance and Fault Detection on Onshore Windfarm Using Digital Twins

By

KAPENIS ANTONIS

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIRMENTS FOR THE DIPLOMA OF

PRODUCTION ENGINEERING AND MANAGEMENT

Chania, February 2024

THESIS COMMITTEE

Assistant Professor: Georgios A. Arampatzis, Supervisor

Laboratory Teaching Staff: Georgios Tsinarakis

Assistant Professor: Dimitris K. Ipsakis

ABSTRACT

This thesis places its focus on the development of a digital twin that faithfully embodies a physical wind farm located in Greece. The principal objective is to establish a virtual counterpart that emulates the real-world characteristics and dynamics of the wind farm. In order to accomplish this, the thesis presents algorithms that are specifically devised to facilitate three vital functionalities: power output prediction, predictive maintenance and fault detection. These algorithms are an integral part of the digital twin's operation, enabling it to forecast potential issues and identify existing problems in the wind turbines. An important characteristic of the digital twin devised in this thesis is its capability to regulate the operations of the wind turbines, per demand. This entails monitoring their performance and, crucially, taking appropriate measures in the event of a malfunction. When the system recognizes a malfunction, it possesses the capability to either temporarily or permanently suspend the operation of the affected turbines until the issue is completely resolved. This approach ensures that any problems are promptly addressed, minimizing downtime and potential harm. A crucial element of the wind farm digital twin centers around the incorporation of real-time data acquired from the wind farm. This data is essential in order to execute the algorithms, as it provides the vital input for the digital twin to successfully perform its functions of predictive maintenance and fault detection. Through the utilization of genuine operational data, the digital twin can generate more accurate predictions and diagnoses, ultimately resulting in a more effective and dependable management of the wind farm.

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my professor Georgios Arampatzis and Dr. Nikolaos Sifakis for their invaluable guidance and support throughout the development of my thesis. Their expertise and encouragement have been pivotal in navigating the complexities of my research, leading me to achieve exceptional results. Their step-by-step mentorship has not only enriched my academic journey but also my personal growth. Furthermore, I cannot overlook the unwavering support and love I've received from my family. Their sacrifices and belief in me have shaped me into the individual I am today, instilling in me the values and resilience needed to pursue my dreams. Lastly, a special thanks goes to my wonderful girlfriend, whose constant support and belief in me have been my cornerstone. Her presence has been a source of strength and inspiration, and without her, I would not be where I am now. Together, each of you has played a crucial role in my journey, and for that, I am eternally grateful.

TABLE OF CONTENTS

1. Background Information on Wind Energy

1.1.1. Overview of the Global Energy Landscape

The global energy landscape has undergone a profound transformation, transitioning from traditional, dominant sources like oil, coal, and natural gas toward a diversified mix inclusive of more sustainable and renewable sources. Throughout history, fossil fuels such as oil have held a significant position in fulfilling the global energy requirements. Nevertheless, due to their limited availability, detrimental effect on the environment, and geopolitical ramifications, there has been an emergence of a transition towards more sustainable alternatives. Renewable energy sources, encompassing solar, wind, hydroelectric, and geothermal power, have experienced a surge in popularity. These sources provide eco-friendly solutions that effectively mitigate carbon emissions and reduce reliance on finite resources. However, while these renewable sources have shown immense promise and have seen substantial growth, achieving complete replacement of traditional fuels remains a complex challenge. While renewables are increasingly competitive, they have yet to monopolize the energy market due to intermittency issues, storage constraints, and the need for further technological advancements and infrastructural developments. The vision of renewable sources entirely dominating the energy market as pollution-free alternatives is aspirational, with ongoing research and innovations aiming to optimize these sources for scalability, affordability, and reliability, positioning them as the primary pillars of a sustainable energy future.

1.1.2. Growing Importance of Renewable Energy Sources

The term Renewable Energy Sources(RES) refers to all those sources of energy that are extracted from the natural environment in order to be used for the production of electricity. Wind, solar, hydroelectric, geothermal, biomass and hydrogen energy are the most leading examples of RES. Renewable energy sources are justifiably characterized as mild forms of energy, with a key element being that they are environmentally friendly, as they are based on an ecological model of utilization with the aim of protecting the ecosystem and the planet. Besides, as far as their exploitation is concerned, they do not cause the release of harmful waste, toxic or radioactive and they do not contribute to the greenhouse effect and the intense climate changes Click or tap here to enter text.. Given that RES are based on the use of the sun, wind, heat, from the subsoil, lakes, rivers, the sea, they are considered to be inexhaustible. Although their exploitation relies on research and technologies that have a high cost, already with the example of solar and wind energy becoming more affordable and economically more accessible, it is becoming clear that they will prevail over non-renewable sources (*The Importance of Renewable Energies | ACCIONA | Business as Unusual*, n.d.). The expansion of clean energies is ever-increasing, as evidenced by the data published yearly by the International Energy Agency (IEA): according to the projections from the IEA, the proportion of renewables in the global electricity supply will surge from 28.7% in 2021 to 43% in 2030. Moreover, they will account for two-thirds of the rise in electricity demand observed during that timeframe, primarily driven by wind and photovoltaic technologies. Renewable energy sources according to research results have both advantages and disadvantages. On the one hand:

- Are environmentally friendly, as they leave little waste and help reduce the greenhouse effect by reducing gas emissions.
- They are inexhaustible sources of energy and are rightly considered as soft or green energy, as they come from natural sources such as wind, geothermal energy and water circulation.
- They do not require any active intervention, i.e. extraction, pumping, burning procedures.
- In addition, they rely on simple equipment in their construction and maintenance and have the financial backing of subsidies from international organisations.
- They are characterised as flexible applications, as they provide energy over long distances and produce energy according to the needs of each country without the need for huge power plants. On the other hand, the disadvantages of RES are real, and they are mentioned that:
- They cannot be implemented on a large scale at this time because of the high costs.
- At the same time, their performance is not always the maximum, because it depends on many variable factors, such as climate, weather conditions and the geographical limits of a region.
- Even many applications of RES, such as in the case of photovoltaics and wind turbines, require large storage spaces, which makes it difficult to use them.
- More specifically, wind turbines are often mentioned, with strong protests from environmental organizations, that they degrade the environment and are responsible for bird deaths and noise pollution. (Papanastasiou Dimitroula, 2022)

1.1.3. Role of Wind Turbines and Farms in Energy Generation

In the present day, wind farms have constitute a significant factor in the energy field, in order to generate clean energy and contribute towards the achievement of the net-zero emission target by the year 2050. One of the main reasons responsible for harmful gas emissions is the utilization of fossil fuels, such as coal and gas, for electricity generation. In order to maintain environmental cleanliness, wind energy has emerged as a plentiful and eco-friendly alternative for energy production(Fahim et al., 2022a). By harnessing the power of the wind, wind turbines facilitate the rotation of generators, thereby enabling the creation of energy. The performance of these wind turbines varies in accordance with the season and geographical location. Consequently, the same wind turbine can exhibit different levels of performance during different months and at different locations. The energy management team faces a significant challenge when it comes to managing uncertainties. To monitor performance, modern wind turbines are equipped with a supervisory control and data acquisition (SCADA) unit. Wind energy has emerged as a notable renewable energy source on a global scale over the past decade. The number of global installations has risen from 192 GW to 743 GW. The density of turbines in wind farms has steadily increased due to advancements in wind turbine technology and a decrease in the levelized cost of energy (LCOE). The average power densities of onshore and offshore turbines are approximately 20 MW/km2 and 7.2 MW/km2, respectively. The optimization of wind farm design in complex terrain and the examination of wind turbine characteristics in urban environments have garnered increased attention in the onshore wind farm sector. In the offshore wind farm sector, there is a focus on the integration of aerodynamics and hydrodynamics, as well as the utilization of limited space through an integrated offshore wind optimization approach (Z. Fan et al., 2023).

Image 1. Wind Farm (*Por-Que-Se-Paran-Aerogeneradores.Jpg (2400×1260)*, n.d.)

1.2. Introduction to DigitalTwins

1.2.1 Definition and Concept of Digital Twins

Firstly, in order to present a precise definition of a Digital twin, it is essential to differentiate between a digital model, a digital shadow, and a digital twin. A digital model is characterized as a virtual representation of a physical system that accurately portrays a predetermined set of behaviors exhibited by its physical counterpart. It should be noted that there is no automated exchange of information between the physical system and its digital model; any information transfer is done manually. Additionally, a digital model is not obliged to encompass all potential behaviors of the physical system, and there are no restrictions on the computational effort or real-time computation required. In simpler terms, the level of fidelity required for the model depends entirely on the specific use case. Model fidelity, in this context, refers to how realistically the behavior of the actual system is reproduced. The transformation of a digital model into a digital shadow is achieved by incorporating a unidirectional automated flow of data or information from the physical system to the virtual system. Conversely, a digital twin requires a bidirectional data flow. By transmitting carefully selected sensor data from the physical system to the virtual system, the digital twin can exhibit synchronous and identical behavior to the physical system. While real-time computation is not an absolute necessity, it is preferable for the virtual state of the digital twin to align with the state of the physical system at regular predefined intervals to be practically useful. The length of this interval depends on the computational effort required to model the desired behaviors. Digital Twins (DTs) are virtual replicas of physical systems and can be used to represent their behaviors and operations. The concept of DTs has rapidly spread in recent years and has become a valuable tool. Decisions based on DTs can be applied in various production systems and sectors, such as manufacturing, logistics, service, healthcare, and energy. The DT model can incorporate data from multiple sources, including sensors, machines, and other devices, to create a comprehensive and accurate representation of the physical system. Furthermore, DT models can be developed by utilizing and integrating various technologies, including simulation, machine learning, big data, cloud technology, and the internet of things (IoT). By combining these technologies, analysts and developers can create highly sophisticated DTs that can replicate the behavior of complex systems and processes. The utilization of DTs in the energy sector, specifically Energy Digital Twin (EDT), can bring innovation to the management of energy systems, leading to improved energy efficiency, reduced downtime, and lower maintenance costs. The application of EDTs is continuously expanding, with numerous studies and research projects being carried out in various domains, such as renewable energy, energy storage, energy distribution, and energy consumption and management (do Amaral et al., 2023). (De Kooning et al., 2021a).

1.2.2. Relevance in Modern Engineering and Renewable Energy

The utilization of Digital Twin technology has been recognized of most of importance in the energy industry and the evolution of engineering in recent years. This technology proves beneficial as it enables the monitoring and optimization of asset performance, prediction of failures, as well as the planning of maintenance and replacement activities. The incorporation of digitalization contributes significantly to the enhancement of security, efficiency, and durability within energy systems. One of the most crucial and effective digital solutions employed in the energy sector is the implementation of Digital Twins (DTs). These DTs possess the capability to monitor and aid in the optimization of power generation, transmission, and distribution systems, along with simulating building energy management systems. By generating a DT for a power plant or building, it becomes feasible to simulate and analyze the behavior of the system in real-time. This allows for the identification of areas for improvement in order to enhance efficiency and decrease energy consumption. The application of DT technology in energy systems is transformative, as it facilitates the production, distribution, and consumption of energy in a more sustainable and efficient manner. In this context, DT-based decision approaches can be implemented within four primary domains of the energy industry:

• Energy supply/production systems (e.g., photovoltaic equipment, wind turbines, hydroelectric plants, and microgrids)

• Energy demand/consumption (e.g., building applications and industrial systems)

• Energy for Transportation (e.g., electric vehicles and engines)

• Energy storage (e.g., Batteries)

Energy DT applications can be categorized into three primary groups, each pertaining to a specific phase of application:

• Design phase, wherein analysis can be conducted to evaluate and validate new assets.

• Operation phase, wherein the monitoring and control of processes, prediction of behaviors, as well as optimization of system outcomes are achievable.

• Service phase, wherein DT can be utilized for maintenance planning and fault detection during process operation.

In relation to the Design phase, the development of a digital twin for a proposed power plant or transmission line allows for the simulation and analysis of system behavior under various operational circumstances. This facilitates the enhancement of design endeavors, with the aim of achieving optimal efficiency and cost reduction. Furthermore, digital twins can also be utilized for the modernization of existing processes and plants, with the objective of enhancing product quality, production rate, and energy efficiency. When contemplating the adoption of digital twins during the Energy Operation phase, these virtual counterparts can be employed to monitor, analyze, and predict system performance throughout its entire life cycle. This enables the optimization of energy production, reduction of downtime, and improvement of overall energy value chain efficiency. By virtue of the monitoring capabilities of digital twins regarding energy demand, it becomes feasible to forecast behaviors by taking into account numerous scenarios and factors that may influence the systems. In the Service phase, it is paramount to underscore the significance of digital twins in the realm of predictive maintenance for energy systems. In this particular scenario, through the analysis of real-time data acquired from sensors, IoT devices, databases, and various other sources, a digital twin possesses the ability to identify potential issues before they escalate into critical problems. Subsequently, it can recommend proactive maintenance activities that aim to prevent downtime and reduce costs. Furthermore, it is worth mentioning that the integration of digital twins during the service phase may lead to reduced costs over the entire lifespan of the systems. This is accomplished through the implementation of maintenance and fault detection support systems that are rooted in digital twins, ultimately resulting in saved time and money. In addition to these practical applications, it is imperative to highlight the potential of digital twins in facilitating collaboration and knowledge sharing across different segments of the energy value chain (do Amaral et al., 2023).

1.2.3. Applications in Wind Energy Management

The concept of a digital twin (DT) has achieved significant success in various domains such as physical and engineering systems, manufacturing, and energy, with particular emphasis on the wind industry. The implementation of a digital twin for a wind farm, which entails creating a digital replica of the real-time spatiotemporal wind field that includes the entire wind energy site throughout the lifecycle of the wind farm, presents unprecedented opportunities for all stages of wind farm development. These stages include wind assessment, planning, turbine-level control, farm-level control, maintenance, repowering, and grid integration(Zhang & Zhao, 2023a). The establishment of a digital twin specifically for wind farm flow offers numerous advantages for the following key areas:

- <u>Wind resource assessment</u>: The periodic nature of wind presents significant challenges when assessing the potential of a wind energy site, including the estimation of annual power production, operation and maintenance costs, and turbine lifespan. Accurately assessing the wind resource has a substantial impact on the decision-making process for wind energy planning and construction.
- <u>Wind turbine and farm control</u>: Real-world wind farms exhibit chaotic wind velocity fields with strong spatiotemporal variability. Furthermore, wake effects from wind turbines significantly influence both the overall power production of a wind plant and the performance of neighboring farms. Therefore, accurately quantifying spatiotemporal wind field information is crucial for efficiently controlling wind turbines and wind farms, aiming to enhance energy capture efficiency and mitigate structural loads.
- <u>Wind energy site monitoring</u>: Real-time monitoring of wind farms plays a critical role in preventing extreme events, reducing structural failures, and scheduling turbine maintenance.
- <u>Wind speed prediction</u>: Accurately predicting wind speed, and subsequently wind power through power curves, is essential for supporting grid integration efforts, as it contributes to stabilizing the electricity grid and enhancing its resilience.

Image 2. Digital Twin in the Wind Turbine Industry(*Digital Twin Image*, n.d.)

1.3. Challenges in Wind Energy Management

1.3.1 Operational Challenges of Wind Turbines/Farms

The generation of wind power is still subject to various environmental factors, such as wind speed, wind direction, and temperature. Consequently, wind power output demonstrates volatility, randomness, and intermittency, which can disrupt the power system's stability. Therefore, the accurate prediction of wind power is essential for the progress of new energy technologies and the reliability of the power system. This has emerged as a significant matter in need of attention. A primary hurdle in wind power prediction research is the limited timeliness and precision of prediction outcomes due to real-time fluctuations in environmental factors. Previous prediction models typically employed the mean value of meteorological data within a specific timeframe as input for wind power prediction. However, real-time variations in environmental factors may hinder the prompt updating of collected data, thereby impacting the accuracy of wind power prediction and diminishing the system's dependability. Consequently, ensuring timely and accurate wind power prediction poses a notable challenge (S. Liu et al., 2023a). Furthermore, both offshore and onshore wind farms encompass intricate systems comprising mechanical, electrical, and structural components. Such complexity gives rise to numerous challenges and potential failures, which can be classified into four categories: fatigue damages resulting from prolonged operations, disasters caused by unforeseen weather conditions and other factors, dimensional and positional deviations during installation, and electrical failures. The examination of wind farm failures and the optimization of maintenance confront persistent concerns, encompassing several unresolved key issues: the dearth of research on the interactions between failures and the underlying mechanisms of said interactions, the arduous task of visualizing the remaining lifespan of wind turbine components, the challenge in effectively integrating data from diverse monitoring systems and systematically analyzing wind farm failures, and the absence of a comprehensive tool that amalgamates all failure analysis methods, impeding the timely analysis of wind turbine failures. (Xia & Zou, 2023a).

1.3.2. Maintenance and Durability Issues

Due to the environmental and societal consequences associated with onshore wind farms, there has been a growing emphasis on the advancement of offshore wind farms. Over the past two decades, solutions involving buoyant wind turbines have been formulated for deep-water areas. Each turbine is affixed to a floating structure and linked to a mooring system. Although this technology allows for the generation of electricity in water depths where fixed-foundation turbines are not feasible, the challenges pertaining to operation and maintenance (O&M) have become substantial. In the context of floating offshore wind farms, the demands for installation and O&M are relatively expensive compared to onshore wind farms. Consequently, the utilization of service vehicles for installation, divers/remotely operated vehicles (ROVs) for inspection, and additional equipment for power distribution is necessary. To facilitate the O&M processes and reduce costs, advanced technologies related to the self-sufficiency of the offshore wind industry are imperative. This scenario presents numerous hurdles, necessitating the development of potential enabling technologies for the establishment of autonomous floating offshore wind farms. The O&M of floating wind farms is a major cost driver, accounting for a significant portion of the overall lifecycle costs of these offshore wind energy facilities. Several factors contribute to the high costs and challenges associated with the O&M of floating wind farms. For instance, the unpredictable weather and harsh conditions of the offshore environment pose significant obstacles for maintenance and repairs. Furthermore, the remote location of these wind farms, situated far from the shore, adds to the difficulties and costs of accessing the turbines for maintenance and repairs. Additionally, the maintenance requirements of the turbines, mooring systems, and other components of floating wind farms can be extensive, necessitating specialized equipment and expertise for efficient execution. Given the hazardous conditions in which these wind farms are situated, ensuring operational safety is also of utmost importance.(Ambarita et al., 2023).

1.3.3. Efficiency in Energy Conversion and Management

Wind energy management poses several challenges, particularly concerning the efficiency of energy conversion and management within wind farms. One of the key hurdles involves maximizing the efficiency of energy conversion from wind to electrical power. Variations in wind speeds and directions present complexities in consistently harnessing optimal energy, demanding sophisticated turbine designs and control systems. Additionally, ensuring efficient energy management across a wind farm, where multiple turbines operate in diverse conditions, poses challenges. Coordinating these turbines to operate at their peak efficiency collectively, considering variables such as wake effects (where downstream turbines receive less wind due to the upstream turbines), requires advanced control strategies and predictive models. Furthermore, integrating energy storage systems to mitigate intermittency and grid integration challenges also remains an obstacle. Enhancing the efficiency of energy conversion and management in wind energy necessitates continued advancements in turbine technology, control systems, predictive analytics, and grid infrastructure to optimize energy output and ensure a stable, reliable renewable energy supply (Cantore, 2017; Sifakis et al., 2020).

1.4. The Role of Digital Twins in Addressing These Challenges

1.4.1. Real-time Monitoring and Prediction Using Digital Twins

The matter of monitoring wind farms and predicting power generation is a complex issue due to the unpredictable nature of wind speed. As a result, it hampers the decision-making abilities of the management team in effectively planning energy consumption. This challenge is addressed by utilizing digital twins to virtually oversee wind turbines and create a predictive model for forecasting wind speed and power generation. The predictive modeling of digital twins is based on a deep learning approach, which consists of two components. Firstly, it analyzes the univariate time series data of wind to anticipate its speed. Secondly, it estimates the power generation for various time intervals ranging from a week to a month. This model can assist the management team in remotely monitoring the wind farms and predicting power generation in advance. The utilization of digital twins' technology allows for real-time data exploration and a feedback loop to the wind farms, providing state-of-the-art computer-oriented solutions. It enables the creation of a digital replica of wind farms connected to physical wind turbines, granting access to supervisory control and data streams for analysis and prediction. The digital twin technology in wind turbines permits fault diagnosis and condition monitoring. A digital twin model facilitates timely monitoring and analysis of wind turbines, while also enabling visualization of construction plans, early detection of structural abnormalities, and accurate identification of wind turbine posture. Furthermore, digital twins are a collection of adaptive models that simulate the behavior of a physical system in a virtual space. These models update themselves with real-time monitoring data throughout the life cycle of the physical system, evaluating and predicting its performance. Specifically, during the installation or construction stage of the physical system, digital twins can assist in real-time evaluation and optimization of the construction scheme, benefiting the coordination of multiple stakeholders in terms of progress, cost, and quality. During the Operation & Maintenance (O&M) stage, digital twins can anticipate the remaining lifespan of the physical system based on state monitoring and virtual operation, such as maintenance, and provide recommendations for O&M. The real-time monitored data also serve as a basis for anomaly diagnosis and fault identification. Therefore, digital twins are a powerful tool for describing, diagnosing, predicting, and decision-making with regards to the installation and O&M of wind turbines (Fahim et al., 2022b; Y. Liu et al., 2023a).

1.4.2. Predictive Maintenance Through Digital Twinning

Thanks to recent advances in digital twins (DT) and their facilitation of predictive maintenance (PdM), companies can significantly optimize their maintenance schedules, minimize downtime and increase profitability and competitiveness. DT enables accurate equipment status detection and proactive failure prediction, thereby improving reliability. Moving from reactive to proactive services can enable PdM to achieve better results. Exploiting real-time data and advanced analytics, the potential to improve wind turbine predictive maintenance through the digital twin concept becomes apparent. Creating digital replicas of turbines can improve the accuracy of performance predictions and estimates of maintenance costs and production losses. Using digital twins can also improve operations by understanding the health of turbines and adapting to changes. The digital twin offers a host of helpful applications such as planning, design, construction, and analysis. By integrating physics-based models, it is capable of assessing wind speed, determining aerodynamic loads, and projecting section loads throughout the tower. This allows for better estimation of fatigue life, which can inform maintenance decisions going forward. Additionally, through the use of weather forecasts and machine learning, the digital twin can aid in both predictive and prescriptive maintenance. Achieving peak performance in a wind turbine entails precise alignment and synchronizing of numerous components. The bearing, blade, and gear are especially susceptible to failure, posing a significant challenge. Quick detection and diagnosis of faults are essential for offshore wind farms, where costly repairs are more likely. Timely intervention is key to addressing high-cost issues that can arise in remote offshore locations. Employing algorithms designed to anticipate and prevent problems is integral to efficiently managing maintenance activities, while minimizing downtime and defect costs. A prototype capable of detecting, supervising, and anticipating failures through distinguishing features from existing systems, is crucial. In current times, potential failures in wind turbines are primarily monitored and predicted through vibrations, shaft speed, noise, and overheating of certain components. Research into establishing a rapid and accurate prediction system for wind turbines using digital and artificial intelligence technologies, such as machine learning (ML) and digital twins (DT), is crucial. The subject of various research endeavors has been digital solutions and machine learning for rotary machines. This includes analyzing system vibrations, detecting bearing defects, and predicting and controlling fast and slow axes (Chen et al., 2023; Vives et al., 2022).

1.4.3. Optimizing Performance and Energy Output

Digital twins play a crucial role in the energy sector by addressing challenges related to optimizing performance and maximizing energy output. These virtual replicas of physical assets, such as power plants or distribution systems, enable real-time monitoring, analysis, and simulation. In power generation, digital twins offer insights into equipment behavior, allowing for predictive maintenance that mitigates downtime and enhances operational efficiency. For instance, in the context of fossil fuel-based power plants, digital twins help optimize combustion processes and equipment performance, leading to improved energy efficiency. Similarly, in renewable energy sources like solar and wind, digital twins enable precise monitoring of conditions affecting energy generation, facilitating adjustments for optimal output. By leveraging data analytics and simulations, digital twins provide a platform for testing scenarios, optimizing operations, and implementing predictive strategies, ultimately contributing to more reliable and efficient energy production across the board.

1.5. Research Gap and StudyJustification

1.5.1. Identification of Research Gaps in Current Studies

Current studies on digital twins in wind energy have made significant strides in enhancing operational efficiency and predictive maintenance. However, several research gaps persist in this domain. Firstly, there's a need for further exploration into the development of more accurate and sophisticated digital twin models specifically tailored for wind turbines. These models should incorporate complex aerodynamic interactions, structural dynamics, and control strategies to better simulate real-world scenarios. Additionally, the integration of data from various sensors and sources to improve the accuracy and reliability of these digital twins remains an area requiring focused research. Furthermore, while digital twins offer predictive capabilities, refining these models to accurately forecast performance under dynamic wind conditions and their long-term reliability is an ongoing challenge. Bridging these gaps will involve interdisciplinary research involving data science, mechanical engineering, meteorology, and advanced control systems to enhance the precision and applicability of digital twins in optimizing wind energy production.

1.5.2. Importance and Necessity of the Current Research

The current research on the role of digital twins in the wind energy field holds paramount importance and necessity in shaping the future of renewable energy production. Digital twins offer a transformative approach to optimize wind turbine operations, enhance energy output, and ensure the reliability of renewable energy sources. Advancements in digital twin technology facilitate real-time monitoring, predictive maintenance, and precise simulations, enabling a deeper understanding of turbine behavior under varying wind conditions. This research is essential for improving the accuracy of digital twin models, enabling better predictions of turbine performance, reducing maintenance costs, and increasing overall operational efficiency. Moreover, as the demand for clean energy rises, leveraging digital twins becomes imperative for scaling up wind energy while ensuring grid stability and reliability. Continued research in this realm is crucial for refining these virtual replicas, integrating cutting-edge data analytics, and advancing control strategies to drive innovation and establish wind energy as a more reliable, cost-effective, and sustainable power source.

1.5.3. Potential Contributions and Impacts of the Study

The comprehensive review of scientific articles investigating the integration of digital twins within the context of wind turbines and wind farms presents multifaceted contributions and profound implications. This study delves into the transformative potential of digital twins, elucidating their role in optimizing wind turbine performance and enhancing overall wind farm operations. By examining various scholarly works, this analysis highlights how digital twins offer a sophisticated means to simulate, monitor, and predict the behavior of wind turbines, enabling proactive maintenance, fault detection, and performance optimization. Furthermore, from this study it is shed light on the potential economic benefits, such as increased energy output, reduced downtime, and extended equipment lifespan, fostering a more sustainable and efficient wind energy ecosystem. The implications of this research reverberate across renewable energy domains, underscoring the pivotal role of digital twins in shaping the future of wind energy technology and its pivotal role in mitigating climate change.

1.6. Objectives of the Thesis

1.6.1. Primary Goals and Aims of the Research

The primary goals of this research on digital twins within the wind energy sector encompass a multifaceted exploration aiming to define the concept of digital twins and their pivotal role within renewable energy, particularly in wind turbine technologies. This research seeks to evaluate the utility of digital twins in addressing critical challenges faced by wind energy systems, focusing on optimizing performance, enabling predictive maintenance strategies, and facilitating remote monitoring. Through this investigation, it is aimed to elucidate the extent to which digital twins can revolutionize wind turbine

technologies by offering real-time simulations, data-driven insights, and predictive capabilities, ultimately enhancing operational efficiency, minimizing downtime, and maximizing energy output. The overarching objective of this research is to underscore the potential of digital twins as a transformative tool in the renewable energy sector, specifically within the realm of wind energy, and to provide valuable insights into their practical applications for optimizing system performance and advancing sustainable energy solutions.

1.6.2. Specific Objectives Related to Digital Twins and Wind Energy

The specific objectives linked to digital twins within the domain of wind energy encompass a multifaceted approach. Firstly, this research aims to comprehensively assess the feasibility and efficacy of implementing digital twins in wind turbine systems, focusing on their role in optimizing operational efficiency and performance. Secondly, it seeks to develop and refine predictive maintenance models using digital twin technology, enabling proactive strategies to mitigate potential faults and enhance reliability. Additionally, the research endeavors to explore the integration of advanced sensor technologies with digital twins to enable precise and real-time monitoring of turbine components, ensuring optimal functionality and reducing maintenance costs. Moreover, the objectives involve evaluating the scalability and adaptability of digital twins across wind farms, aiming to facilitate seamless integration and management of multiple turbines within a unified framework. Overall, these specific objectives converge to unlock the transformative potential of digital twins, aiming to revolutionize wind energy systems by improving reliability, efficiency, and overall sustainability.

1.7. Outline of the Thesis Structure

1.7.1. Overview of the Thesis Chapters and Their Contents

The thesis consists of two extensive chapters. Chapter 1,the introduction, provides a comprehensive overview of wind energy and explores its importance in the global energy landscape, the growing importance of renewable energy, and the important role that wind turbines and power plants play in energy production. Additionally, it introduces the concept of digital twins, explaining its definition, relevance in contemporary engineering and renewable energy contexts, and its specific applications in wind energy management. This chapter also explores the various challenges faced in wind energy management, including operational obstacles, maintenance issues, and the necessity for efficient energy conversion and management. It also explains how digital twins can address these challenges by enabling real-time monitoring, predictive maintenance, and performance optimization. Moreover it establishes the research gaps, justifies the significance of the study, outlines the primary goals and specific objectives, presents the structure of the thesis, and describes the contributions of the study to wind energy technology and digital twin research, as well as its implications for sustainable energy practices. In Chapter 2, titled "State of the Art," a comprehensive analysis is conducted on various aspects related to wind energy and digital twin research. This includes an in-depth exploration of the history, development, and technological advancements in wind energy, as well as the evolution and applications of digital twins. The chapter also delves into the specific utilization of digital twins in renewable energy and wind energy contexts, highlighting the challenges faced in wind turbine operation, maintenance, and integration with digital twin systems. Furthermore, energy management strategies, economic and environmental considerations, methodological approaches and tools used in research, critical assessment of literature, identification of research gaps, and the discussion and adaptation of relevant theoretical frameworks and models are thoroughly discussed within the realm of wind energy and digital twin research. Each section within these chapters critically evaluates and contributes to the comprehensive understanding and advancement of wind energy management using digital twin technology.

1.8. Significance of the Study

1.8.1. Contribution to Wind Energy Technology and Digital Twin Research

The significance of this study examining digital twins within the context of the wind energy sector and renewable energy at large lies in its potential to revolutionize the efficiency, reliability, and sustainability of wind turbine technologies. By defining the essence of digital twins and investigating their applicability in wind energy systems, this research addresses critical gaps in understanding how these sophisticated digital replicas can transform the sector. Moreover, this study focus on assessing the utility of digital twins in optimizing wind turbine systems, facilitating predictive maintenance, and enabling remote monitoring highlights their pivotal role in mitigating challenges faced by these technologies. By bridging the gap between theoretical concepts and practical applications, this research significantly contributes to wind energy technology by offering insights into enhancing operational efficiency, minimizing downtime, and maximizing energy output. Furthermore, this study's contributions to digital twin research extend by providing a nuanced understanding of their implementation within the renewable energy landscape, potentially influencing future developments and innovations in this burgeoning field.

1.8.2. Implications for Sustainable and Efficient Energy Practices

This study holds profound implications for sustainable and efficient energy practices within the realms of wind and renewable energy. By exploring the utilization of digital twins in wind turbine technologies, this research underscores potential pathways to significantly enhance the performance, reliability, and longevity of these systems. The integration of digital twins offers a strategic approach to optimize operational efficiency, enable predictive maintenance strategies, and facilitate remote monitoring, thereby mitigating critical challenges faced by wind turbines. These implications extend beyond just individual turbine efficiency, promising a more sustainable renewable energy landscape by minimizing downtime, reducing maintenance costs, and maximizing energy output. The incorporation of digital twins in wind energy not only improves the dependability and efficiency of wind farms, but also plays a crucial role in advancing the larger objective of promoting cleaner and more sustainable energy practices, aligning with international endeavors to address climate change and transition towards a more environmentally friendly future

2.1. Overview of Wind Energy Technologies

2.1.1. History and Development of Wind Turbines and Wind Farms

Wind turbines, as they are now called, differ from classical windmills in their primary function of generating electricity rather than extracting mechanical energy for pumping water or grinding. Since its inception, electricity's popularity has soared as a result of its versatility and transportability. It can be transformed into a variety of energy types, making it a highly valuable source of power. Over time, mechanical systems have become a common source for power generation, particularly after the invention of the alternating current generator. Professor James Blyth was the first individual to utilize a wind-powered machine for electricity generation. His successful efforts to power his vacation home with a wind turbine that stood at a height of 10 meters and had blades secured with sail cloth, date back to 1887, and the machine was developed while he was associated with the Anderson College in Glasgow. Just a year after Professor Blyth's attempt in 1888, Charles Bush made the most noteworthy debut of wind turbines in the USA. With a radius of roughly 8.5m, the turbine produced 12 kW of power and contained a rotor made up of 144 blades. Despite its size, the rotor turned at a sluggish pace and yielded a relatively low output. Marcellus Jacobs subsequently took the lead on the modernizing of wind turbines in America, ultimately resulting in the turbines that we know today. Equipped with battery storage, these machines had three blades and utilized real airfoil shapes. Tailored for residential areas, they were distinguishable by these features. The development of wind turbines grew steadily in response to high oil prices and a rising demand for wind power. Larger turbines emerged gradually. Wind turbine development mostly dominated in Denmark throughout the 20th century. Poul La Cour worked with Danish manufacturers Danish Lykkegaard and Ferritslev to create a total of 100 turbines, each with a rotor power of 20-35 kW. These turbines were used for generating direct currents that were supplied to small grids and batteries. At the Aerodynamic Experimental Institute in Göttingen in Germany, Albert Betz had a revolutionary moment in 1920 when he introduced a mathematical analysis to measure the utmost efficiency of wind turbines. By using this idea, we can approximate the maximum limit at which a wind-powered mechanism can work, around 59.3%. The abundant supply of fossil fuels caused the hype around wind turbine technology to decline post the First World War with steam and diesel engines taking over the limelight. But after the Second World War, the limited supply of fossil fuels revived the interest in wind turbines. It's clear that the success of the technology heavily relies on the accessibility of fossil fuels and the dedication towards funding. The fast-growing wind turbine industry owes its success to various national initiatives geared towards reducing the reliance on fossil fuels. These efforts have significantly matured the technology behind wind turbines. Nowadays, wind turbines are a major player in the energy sector. As the need for cleaner, renewable energy sources grows, wind turbines have emerged as a popular solution. This is due in part to concerns over high carbon emissions and the detrimental impact they have on the environment. To address this issue, many countries have established goals aimed at increasing renewable energy in their energy mixes, including the United Nations Sustainable Development Goals. Wind energy has been singled out as a key component in reaching these goals, with its rapid development and growing demand propelling efforts to improve capacity and efficiency. Starting in 2016, wind energy surpassed coal as the second largest power generation capacity, according to Wind Europe. Wind power is quickly transitioning into a commercialized, non-subsidized technology that can compete with fossil and nuclear sources. In Germany, wind turbines produce a significant majority of the country's renewable energy. Onshore wind energy is the leading renewable market for electricity sales, thanks to its lower operational, transport, and maintenance costs. The increasing demand for wind energy is hard to meet as there aren't many areas with high wind potential. However, to expand wind energy, erecting wind farms closer to residents and in the mountains may be necessary. With wind development taking place more and more near communities, it's crucial to grasp why communities accept wind facilities. Offshore wind farms are expanding in size to generate more electricity, and it won't be long before floating wind turbines become the norm. Another necessary measure is to systematically replace smaller wind power plants with larger ones at high-potential locations - a process known as re-powering. The old era of small wind turbines is a far cry from what we see today. The transformation is evident as the rotor on the modern wind turbine has been enlarged to a massive scale, making it one of the most colossal machines on earth. With an increased rotor size comes an entirely new set of challenges to overcome. The blades, for example, have become thinner and more flexible than ever. Additionally, the changes require a new set of safety standards to be implemented in order to keep up with the technical implications of the gigantic rotor size. Approaching the tip area, large turbines can attain extreme rotor speed, making high Reynolds number flow field interaction quite the challenge for present-day modeling strategies. The aerodynamics of small wind turbines from the 1980s can no longer hold for these machines. To assess wind turbine loads, holistic approaches must begin during the design phase and continue throughout planning, manufacturing, operation, and maintenance. With the increasing difficulty of experimental campaigns, these machines require numerical strategies to consistently meet high standards for accuracy and robustness. At the same time, it's important to balance computational cost and ensure that it remains reasonable(Bangga, 2022a; Herrmann & Bangga, 2019; Schaffarczyk, 2020).

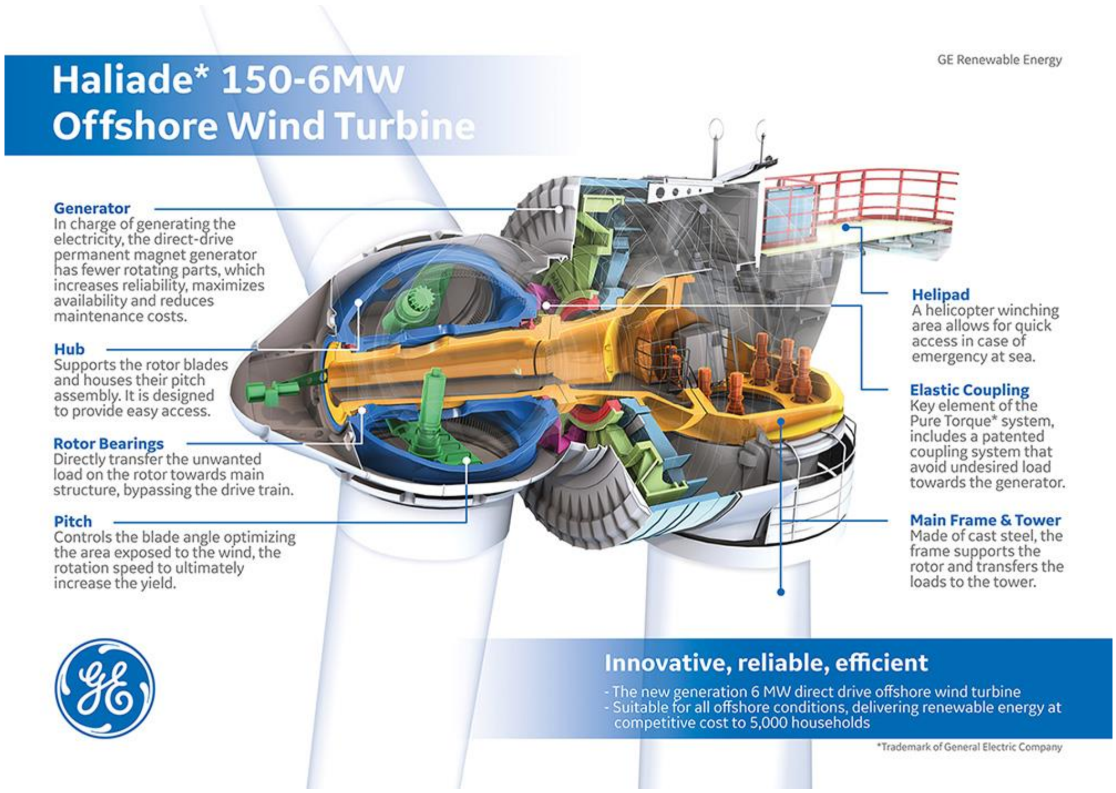

Image 3. Anatomy of the Modern Wind Turbine(*Anatomy of the Modern Wind Turbine*, n.d.)

2.1.2. Technological Advancements in Wind Energy

Through the conversion of kinetic energy into mechanical torque, generators driven by wind energy harvest the kinetic energy via an electromagnetic effect in wind turbines. The electricity resulting from this process is a result of the energy conversion described. Wind energy output is proportional to the square of the rotor radius and directly related to the blade sweep area and wind velocity cubed. In order to harness the most wind power possible, progress has been made in the expansion of rotor size. Vestas has produced the world's biggest wind turbine, which boasts an impressive 43,742 m2 sweep area and a 15 MW capacity, making it the leader in the industry. With the widespread acceptance of renewable energy technologies RETs, new strategies have been developed and tested to enhance wind energy conversions. Among these advancements are improvements to the working structure of radial wind turbines. Investments in wind energy production are promoted through lower operation costs, which are made possible by bladeless turbine exteriors that increase power coefficient. This design allows for online diagnosis and prognosis, improving operation and maintenance. Though similar to photovoltaic systems, wind energy is limited by the intermittency of available winds. Additional research efforts are required to effectively extract wind power and operate conversion systems due to the unpredictable changes in wind speed. As such, recent advancements in wind energy systems have been thoroughly examined through various literature works to identify tangible benefits that can positively impact the RETs industry as a whole. Several milestones were reached during the late 80's to the 2000's regarding power electronics and wind energy conversion system control. These advancements enabled significant increases in wind turbine size. Studies on the fully synchronous or asynchronous generators and doubly fed induction generators (DFIG) led to innovative solutions that fully control the rotational speed of generators. The extraction of energy from the wind and its conversion into electrical energy is facilitated by conventional wind turbines via the electromagnetic effect. The benefits of this process include reduced mechanical stress on the turbines, increased efficiency across a range of wind speeds, and improved integration into the power grid. As a result of these benefits, innovative methods such as wind energy harvesting and scavenging have emerged in urban areas. The mass adoption of conventional wind farms in urban areas is hindered by limitations. The construction is expensive due to the creation of huge turbine blades, and this limits where the wind farms can be installed(Mossa et al., 2021; Tan et al., 2022).

2.1.3. Current Trends and Future Outlook in Wind Energy Technology

Generating energy from the wind involves numerous components and devices that integrate to make a wind turbine. Improvement and expansion of the wind turbine capacity was necessary due to a growing demand for clean energy and this has been recorded from the late 1990s to present. The commissioning of offshore turbines is going to reach a nominal power of about 15 MW by 2025. As there is still a need for wind farms, the goal is less visual impact, increased efficiency of energy and affordability in operation and maintenance. Longer and lighter blades, taller towers, improved control systems, and stronger electrical transmission grids have been developed to enhance the efficiency of high-tech machines as explored. One must also consider the durability of large-scale turbines in harsh climates and the use of environmentally-friendly materials. The energy transition has a critical reliance on the consolidation of hybrid farms and wind energy storage systems. Ongoing efforts are being made to improve the quality and productivity of rotor blade manufacturing processes. Generating more energy on an annual basis is possible by lowering the cut-in speed and increasing the cut-out speed of wind turbines. With the development of wind turbines generating energy at low wind speeds, regions with lower average annual wind speeds can now have wind farms constructed. To determine at what electricity price such a wind turbine, such as one with a capacity of 3.4 MW and a rotor diameter of 208 m, would be practical in the European electricity system is currently being analyzed. This technology is currently priced at 45% more than a traditional onshore turbine with the same hub height. However, it can produce more than twice the electricity in regions where lower wind speeds prevail, making it a wise investment for the future European electricity system. In order to increase energy production, it is advantageous to create wind turbines that are capable of operating in high wind conditions, shifting the limit further to the right. Additionally, updating the control system for current turbines can lead to a better power curve. Although this can cause extreme vibrating and added stress, the benefits of increased energy output are substantial. Wind turbines face a difficult challenge in controlling overspeed during extreme wind speeds. They currently use mechanical and aerodynamic braking systems to do so.

However, a new technique has been developed that is more unique in its approach. It involves placing openings, or chord slots, in the blades to alter the pressure distribution on the surface. This causes wind turbine speed to reduce to proper limits, providing a novel way to control overspeed aerodynamically. Without impacting energy creation, the turbine rotor's fast pace is reduced through a successful approach. The impact of multiple slot parameters on chord, including slot position and length, was examined for power generation, and the optimizing of slot parameters was conducted through experimentation and computation. Due to the escalating prices of conventional energy sources and the reliance on their import by numerous countries, the necessity for a secure energy supply has prompted substantial investments in the expansion of wind power capacity. Despite wind power generation being an established technology with low levelized electricity costs, there is still potential for enhancement. An examination of existing literature has revealed that future advancements in wind turbine development will focus on scaling up turbine size and implementing minor design refinements. These refinements encompass further enhancements in rotor blade aerodynamics, the implementation of active control systems for rotor blade rotation, and the integration of aerodynamic brakes to maximize power generation efficiency. Additionally, improvements in system maintenance and the early detection of transmission and power-related faults, as well as blade surface damage, will reduce turbine downtime and enhance overall system reliability and availability. The transportation and assembly challenges associated with the production of larger wind turbines are being addressed by adopting a segmented approach for manufacturing the blades. In the analysis of wind turbine efficiency, as well as stress and vibration, numerical methods are increasingly employed. The use of direct drive is gaining competitiveness over traditional gearbox power transmission. In the realm of offshore wind farms, the prevailing trend is to increase the size of wind turbines and position them further from the coastline, necessitating the development of innovative floating foundations. Optimization techniques are currently being developed to meet the unique demands and challenging conditions of marine environments when constructing offshore substructures. Plans are underway to replace the current 33-kV cables with 66-kV cables for power transmission from offshore wind farms. The integration of offshore wind farms can play a significant role in facilitating the transition towards a hydrogen-based economy.Plans are underway to generate a substantial amount of environmentally friendly hydrogen through the process of electrolysis using water. As the initial generation of wind turbines approaches the end of their operational lifespan, efforts are being made to devise strategies that involve repowering, prolonging their usage, or dismantling and recycling them (Bangga, 2022b; Bošnjaković et al., 2022; Lucena, 2021).

2.2. Digital Twins in Engineering

2.2.1. Evolution of the Digital Twin Concept

The concept of a Digital Twin (DT), which is a virtual representation of the characteristics of physical assets, has been present for a considerable period. Its origins can be traced back to 2002, when it was initially introduced in the field of product lifecycle management (PLM) as a "conceptual ideal for PLM." Initially, it found its primary application in engineering and simulation tasks within the aerospace and astronautics industries. Despite the conception of an ideal representation, the main objective was to provide technical insights, visualization tools, and virtual or augmented reality capabilities with relevant data. However, from an architectural perspective, these tools were self-contained and lacked integration with other phases of the lifecycle or IT systems, such as those found in operational industrial plants. The achievement of such integration required significant effort and resources. The landscape is on the verge of transformation with the emergence of the Industrial Internet of Things (IIoT) and Industry 4.0. The proliferation of sensors in industrial environments, combined with networked machines, has revolutionized the availability of data. The cumbersome task of collecting and processing data for digital twins has been simplified. Thanks to interconnected edge/cloud environments, these valuable data can now be accessed globally, surpassing company boundaries. To avoid the dominance of proprietary solutions or major IT players, the concept of "dataspaces" was introduced. Dataspaces act as a "data middleware," facilitating controlled sharing and utilization of data among partners within a federation, while upholding the principles of data sovereignty. Consequently, the development and potential of digital twins must be closely intertwined with the evolution of dataspaces. The concept of the digital twin has undergone changes since its introduction in 2002 by Michael Grieves, who was teaching a course on product lifecycle management at the Florida Institute of Technology. Grieves defined the digital twin as a digital representation of a physical system that exists independently. This representation encompasses all the information about the physical system and is continuously connected to it throughout the lifecycle of the product. Additionally, Grieves outlined the key components of the digital twin approach, which include a physical space, a virtual space, mechanisms for data flow between the two spaces, and virtual sub-spaces. Despite the absence of the necessary technology at the time Michael Grieves introduced the term "digital twin," his description was already quite advanced. It is important to note that Grieves primarily focused on digital twins for products, given his background in product lifecycle management. However, as we will explore later, the concept of a digital twin can be applied to a much broader range of domains. In 2012, the National Aeronautics and Space Administration (NASA) provided their own definition of a digital twin, referring to it as a simulation that encompasses multiple physics, scales, probabilities, and high-fidelity representations. This simulation accurately reflects the state of its corresponding twin based on historical data, real-time sensor data, and physical models, all in a timely manner. Over the course of time, numerous enterprises and scholars have embraced the notion of digital twins; however, a current examination of the existing literature fails to identify a universally accepted definition. It transpires that as the Internet of Things (IoT) and, more recently, dataspaces have come into existence, digital twins have transcended their original purpose of simulation and have instead evolved into pivotal conceptual and architectural components within distributed environments, catering to a multitude of usage scenarios (Jeong et al., 2022a, 2022b; Usländer et al., 2022).

2.2.2. Key Technologies Underpinning Digital Twins (e.g.,IoT,AI,Big Data)

DT models can integrate data from a variety of sources, including sensors, machines, and other devices, to create comprehensive and accurate physical models. In addition, DT models can be developed using and integrating various technologies, including simulation, machine learning, big data, cloud technology, and the Internet of Things (IoT). By combining these techniques, analysts and developers can create complex DTs that reflect the behavior of complex systems and processes. Digital twins have become an integral part of Industry 4.0, playing a crucial role in helping businesses gain a deep understanding of their data, optimize complex processes, and enhance operational efficiency. Industry 4.0, which marks the transition from embedded systems to cyber-physical systems, relies heavily on these cyber-physical systems (CPS) that merge real-time workflows with digital technologies. By incorporating embedded systems and sensors, production technologies and smart production processes enable optimal performance, revolutionizing industries such as production value chains and business models. The integration of Internet of Things (IoT) sensors with assets provides businesses with access to vast amounts of data, which forms the foundation of digital twins. Through IoT, machines and devices can seamlessly connect and interact with each other, further enhancing the capabilities of digital twins. Connected products, utilizing the Internet of Things (IoT) and a digital thread, form the foundation of Digital Twins. This interconnected system enables seamless connectivity throughout the lifecycle of the products. The Digital Twin gathers data from its physical counterpart, continually updating its models. Real-time supervision and improved communication between Cyber-Physical Systems (CPS) and users are facilitated by the IoT. By harnessing the Volume, Veracity, Velocity, and Value of captured data, the Internet of Services (IoS) effectively provides services through the internet. Initially introduced to address the growing volume of data impacting the IoT, Digital Twins merge with IoT to provide essential insights into the behavior and performance of physical twins in operational environments. Integration of multiple Digital Twins into the IoT expands the benefits, allowing for centralized monitoring of maintenance schedules and cycles (Adjei & Montasari, 2022; do Amaral et al., 2023).

2.2.3. Applications of Digital Twins in Various Engineering Fields

By using DT to reflect and simulate real-world scenarios, companies can gain insights into how their products and services will perform under different conditions. Therefore, the overall operation can be optimized. DT can be used to support decision-making in three approaches:

- diagnostic, which aims to evaluate past decisions
- surveillance, which aims to monitor and control processes and
- prognostic, whose goal is to predict and predict behavior

DT is widely used in the manufacturing industry to simulate manufacturing processes, identify potential bottlenecks and issues, and help decision makers optimize production procedures, reduce costs, and improve product quality. On the other hand, in healthcare and service applications, DT simulates the behavior of individuals such as patients and customers to help decision makers make informed decisions. You can use DT to model logistics processes to simulate traffic patterns and warehouse operations. For example, decisions related to vehicle routes and replenishment management can be optimized. Over the course of the previous years, the energy industry has witnessed a surge in the adoption of DT technology. This technology has become increasingly popular due to its ability to effectively monitor and enhance the performance of assets, anticipate potential failures, and strategize maintenance and replacement operations. The utilization of DTs allows for the monitoring and enhancement of power generation, transmission, and distribution systems, as well as the simulation of building energy management systems. Through the creation of a DT for a power plant or a building, one can simulate and analyze the system's performance in real or near-real time, pinpointing areas for improvement to enhance efficiency and decrease energy usage. The integration of DT technology in energy systems has the potential to revolutionize energy operations, encompassing production, distribution, and consumption, ultimately fostering sustainability and optimizing efficiency (do Amaral et al., 2023; Mohamed et al., 2023).

2.3. DigitalTwins in Renewable Energy

2.3.1. Specific Applications of Digital Twins in Renewable Energy

The innovative concept of the digital twin for renewable energy sources within power grids has the potential to completely transform energy management. This concept involves the utilization of digital models to replicate and simulate the actions and behaviors of renewable energy sources, such as wind and solar power, with the ultimate goal of enhancing performance and understanding. By implementing digital twins, operators of energy systems can acquire a deeper understanding of their grid's operations and effectively manage their energy resources. Initially introduced within the automotive industry, the concept of digital twins has now been adapted for application within power grids and renewable energy sources. By utilizing a digital twin, energy system operators can gain valuable insights into the performance of their grid and identify opportunities for optimization. This includes identifying areas for improvement, such as enhancing energy efficiency and reducing energy costs. The real-time monitoring of renewable energy sources is made possible through the utilization of digital twins. Energy system operators can obtain a comprehensive overview of their grid's performance and detect any potential issues in advance by gathering data from various sensors and other sources. This proactive approach enables them to take necessary measures to ensure grid reliability and enhance energy efficiency. Consequently, the incorporation of digital twins can lead to a reduction in energy management expenses. By gaining insights into the performance of renewable energy sources, operators can pinpoint areas where cost reduction is feasible. This may involve minimizing the necessity for costly infrastructure upgrades or optimizing the utilization of existing resources (X. Fan & Li, 2023; Kamyabi et al., 2022; Sleiti et al., 2022a).

2.3.2. Case Studies or Examples of Digital Twins in Wind Energy

With the entry of the digital twins(DT) in the energy sector and especially wind energy, many issues that were hindering its development and its prevalence as the most basic source of energy have been resolved. There are many examples and studies on the use of digital twins. With a review of their involvement and their unreserved help in the wind energy sector, each example is presented separately. One study has shown that using DT, a model is created that can predict the wind farms' electricity production, wind speed and can act as a monitor to control the whole farm remotely (Kamyabi et al., 2022). Another example is the use of DT to implement a model that will prevent any failures in the systems of offshore wind farm installations (possible damage to wind turbines) while finding more innovative and economic solutions for maintenance and repairs of the installations(Xia & Zou, 2023b). Utilize a DT to model a wind turbine turbine taking into account the aerodynamics, structure and mechanics of the drive system, the synchronous permanent magnet generator, the electronic power converter and the tilt and rotation systems in order to avoid any anomalies in the operation of the turbine (De Kooning et al., 2021b). In another research, a new approach of merging data and knowledge is proposed to create the first DT for onshore and offshore wind farm flow system which can predict the in situ spatio-temporal wind field covering the whole wind farm. DT is developed by integrating Lidar measurements, Navier Stokes equations and actuator disk modeling of the vortex through neural networks and laws of physics (Zhang & Zhao, 2023b). A digital twin (DT) approach to monitoring the status of drive systems in floating offshore wind turbines. The focus of this research is a DT solution for estimating the remaining lifetime of drive system components based on real-time estimation of the equivalent drive system model and subsequent monitoring of stress concentration changes in the various components and applying the estimated stresses to probabilistic and stochastic degradation models that can indicate component fatigue failure (Moghadam & Nejad, 2022a). The role of DT in a different research is to extend the utility of the simulation model to the operational phase of offshore wind turbine (OWT) construction. It reflects the actual performance of a system by correlating simulations with observed data. Addresses the challenges in quantifying uncertainty in the predicted lifetime of fatigue. It supports the in-service evaluation of OWT structures by defining a framework to address uncertainty. It allows for efficient simulation of computationally expensive numerical simulators. It informs decision-making on operation, inspection and maintenance (Jorgensen et al., 2023). The aim of another research is to optimize the energy production of a smart island, ensuring the efficient distribution of energy from RES such as Photovoltaic (PV) systems, Wind Turbines (WT) taking into account the environmental and economic impacts. To predict the output power of these RES accurately using a deep learning model with a recurrent neural network (RNN). DT digital twins are used to model the optimal performance of the Smart Island(SI) in grid-connected operation. They facilitate power shifting for intermittent loads using storage batteries in a digital twin environment. The proposed model is implementable in Matlab, which can be used in a digital twin environment (Jorgensen et al., 2023). In a subsequent paper, a simulation environment based on digital twin technology is proposed for modeling and simulation of the energy system of the Hywind Tampen wind farm (Equinor-Norway). It demonstrates the ability of the wind farm to consistently provide one third of the electricity needed by oil and gas platforms, reducing CO2 and NOx emissions. At the same time, the whole process serves as a laboratory for the development of future floating wind turbine designs (Qaiser et al., 2023). One study proposed a predictive maintenance method for wind turbines using DT digital twin technology to accurately predict wind power in real time.Used the BP (back propagation neural network) neural network for the initial wind power forecast and weighted it with older meteorological data for increased accuracy. Implemented sensors to collect information from equipment in real time, including weather data, to model forecasts. Developed a level of data fusion for preprocessing, ensuring reliable baseline data for forecasting. Improving the power accuracy of the forecasting system by integrating DT digital twin technology with meteorological information (S. Liu et al., 2023b). A digital twin (DT) method was proposed for rapid prediction and visualization of the airflow distribution behind a wind turbine, structural deformation and stresses in stationary offshore wind turbines (OWTs). Reducing the high costs associated with on-site monitoring, operation and maintenance of offshore wind turbines due to the harsh offshore environment. Provide a reference for the safety assessment of the flow field distribution and the structural strength of the turbine. Development of a method that combines digital modelling, computational fluid dynamics (CFD) simulation and machine learning for intelligent real-time performance evaluation of stable OWTs (Cao et al., 2023). A case in point of construction a cost-benefit digital twin (DT) for managing the operation of an offshore wind turbine on a monopile using a component-based Reduced-Order Modelling (ROM) approach. To provide real-time information on structural health conditions and future projections for an offshore wind turbine system. To achieve faster computation speeds and high accuracy in predicting the structural responses of wind turbines under wind and wave loadings(Zhao et al., 2023). A case study to optimize the layout of floating offshore wind farms during operation using a digital twin-driven approach. The utility of the DT is to increase the total energy production by minimizing the wake effect through dynamic repositioning of wind turbines. To provide a visualization tool(DT) for wind farm designers that is intuitive and aids in decision-making.It provides an intuitive platform for assessing the wake effect and optimizing wind farm layouts. Enables designers to visualize different configurations and their impact on energy production.Facilitates faster and cost-effective simulation compared to traditional experiments (Kandemir et al., 2023). Other research developed a smart integrated energy management system for decarbonizing offshore oil and gas fields through the optimization of microgrid operations. It integrated renewable energy sources, such as floating wind turbines, into offshore microgrids, boosting energy efficiency while reducing carbon dioxide emissions. It demonstrated the effectiveness of the optimization method developed on real offshore platforms, showing a significant reduction in operating costs and emissions. DT digital twins were used for accurate predictions and high-speed processing of various scenarios in the microgrid optimization process. They enabled the use of artificial intelligence techniques to create powerful forecasting units for dynamic factors such as wind availability and demand patterns. Digital twin models, in particular ANN Artificial Neural Networks, have helped to accurately represent the complex dynamics of gas turbines and wind turbines (Banihabib & Assadi, 2023).

2.3.3. Benefits and Limitations Observed in Existing Applications

Existing applications of digital twins in renewable energy have demonstrated significant energy benefits alongside certain limitations. Digital twins offer precise monitoring, predictive analytics, and operational optimization, leading to increased energy efficiency in renewable energy systems such as solar and wind farms. Through real-time data analysis, these digital replicas enable proactive maintenance, thereby reducing downtime and enhancing overall system performance. Furthermore, digital twins facilitate accurate simulations and scenario modeling, enabling the fine-tuning of renewable energy systems for maximum output. However, limitations exist in the complexity of integrating various data sources and ensuring the accuracy of predictive models. Challenges related to data security, privacy, and the need for specialized expertise also pose hurdles. Nevertheless, despite these limitations, the energy benefits observed in current applications of digital twins in renewable energy underline their potential to significantly enhance the efficiency and reliability of sustainable energy systems. Continued research and development aim to address these limitations and further leverage the energy benefits offered by digital twin technology in the renewable energy sector.

2.4. Challenges in WindTurbine Operation and Maintenance

2.4.1. Detailed Review of Operational Challenges (e.g., Variability of Wind, Efficiency Issues)

Future wind turbine technology development should prioritize four main goals as the main challenges on the field that must to be solved. The first objective is to optimize the economic aspects, which entails reducing the levelized cost of energy (LCOE) and enhancing economic value through strategic decisions regarding reuse, repurposing, and decommissioning. This comprehensive approach considers the entire life cycle of wind turbines. The second goal is to design wind energy projects in a way that minimizes environmental impact and maximizes their value to the community. This involves striving for good neighbor status by mitigating any negative effects on the natural and built environments. The third objective is to achieve seamless and cost-effective integration with other energy generation sources and storage systems. To accomplish this, the utilization of wind resources should be optimized, taking into account specific power requirements in regions with varying wind potentials. By aiming to shorten the design cycle, the fourth objective seeks to decrease expenses and hasten the integration of new products in response to evolving needs. The utilization of precise and user-friendly simulation capabilities facilitates swift assessment of design options. Additionally, this could potentially diminish the need for excessively cautious safety measures, resulting in enhanced safety and reduced costs linked to insurance, liability, and instances of failure. Throughout the growth of wind turbines, careful measures have been taken to ensure that they stay within familiar boundaries and maintain sufficient safety margins. However, as turbine sizes have gradually exceeded their original design limits, the prevailing design practices have continued as if this change has minimal impact. Although high safety factors have provided a level of security, the constant drive to reduce costs and weight has diminished these margins of safety. As we strive to tailor designs to specific conditions, uncertainties multiply and gaps in scientific knowledge widen. Consequently, the introduction of large and flexible turbines into the market has ventured into uncharted territory with significant unknowns. The size, complexity, and cost pressures associated with these turbines have hindered innovation in wind turbine design by introducing uncertainties. The resolution of these issues is crucial for substantial improvements in the overall system(Veers et al., 2023).

2.4.2. Maintenance Challenges (e.g., Wear and Tear, Remote Locations)

In recent years, there has been a noticeable rise in both the quantity and intricacy of wind turbines. This increase in complexity has posed a growing challenge in maintaining a consistent level of reliability for wind turbine systems. Additionally, the size and intricacy of wind turbines, even at the subcomponent level, have contributed to a rise in maintenance costs. The faults that have the greatest impact on downtime and productivity loss are the most critical ones to address. While there have been studies conducted on specific faults that affect wind turbine operation, it is important to note that these studies are contingent upon factors such as the wind turbine model, geographic location, and environmental conditions. The act of prioritizing maintenance tasks based on available resources, such as personnel, equipment, and spare parts, is known as maintenance planning. This planning process encompasses all maintenance activities and can lead to significant cost savings through optimization. These savings are primarily associated with current assets, including fuel, mobilization costs, production losses, and logistics expenses. Managing operation and maintenance activities to minimize operating costs is a crucial challenge for offshore wind farms, as maintenance needs change over time based on the performance of individual wind turbines and their components, as well as weather conditions. To effectively determine maintenance activities, project managers must possess a comprehensive understanding of the history, background, performance, and weather patterns related to the sub-assemblies. Typically, maintenance activities are triggered by component failures or predetermined time intervals based on operational service principles. An excellent example of a better understanding of the challenges in the field of wind turbine maintenance is the example of the maintenance taking place on offshore wind turbines.The maintenance of Offshore Wind Turbines (OWTs) is widely recognized as a crucial undertaking, but it comes with its fair share of challenges. These challenges arise from various factors. Firstly, the distance between offshore wind farms and ports or shores poses a significant obstacle, as it hampers accessibility and leads to increased periods of downtime. Furthermore, the expenses associated with owning or hiring a maintenance fleet, as well as the need for a larger number of technicians, can be quite burdensome. Additionally, the complexity of OWTs is amplified by the utilization of both bottom-fixed and floating foundations. Moreover, adverse weather conditions, particularly high wave heights and strong wind speeds, further impede access to OWTs for service vessels and hinder the transfer of personnel from the vessel to the OWT. Over the past ten years, the use of motion-compensated gangways in conjunction with service operation vessels has become common practice for offshore access systems. However, these devices still pose challenges as they are both heavy and expensive. In the event of unfavorable weather conditions, maintenance tasks may need to be delayed, resulting in longer waiting periods and a greater loss of power generation during downtime. Even without taking weather into account, the costs associated with offshore wind turbine (OWT) maintenance are higher compared to equivalent onshore tasks due to the need for specialized equipment. Additionally, the harsh working conditions offshore, characterized by higher wind speeds, wave-induced motions, and structural vibrations, contribute to increased failure rates of OWT components. Furthermore, the trend towards larger OWTs in recent years, aimed at improving power generation efficiency, necessitates the use of larger and more specific devices for offshore maintenance and repairs (Márquez & Pinar Pérez, 2020; Ren et al., 2021a; V. Taboada et al., 2021; Zhou & Yin, 2019).

2.4.3. Approaches to Addressing These Challenges

Measures need to be taken to address the challenges mentioned above. As a continuation of the challenges that arise in the field of maintenance and especially in the field of maintenance of offshore wind turbines, a strategy that will actively contribute to addressing these challenges is outlined below(Ren et al., 2021b). More specifically a preventive strategy, in most cases, involves scheduled maintenance that occurs either at predetermined intervals or when a certain level of power generation is reached. The decision on which intervention to implement is based on the reliability of each component and the overall cost. If a failure occurs between two scheduled visits, the wind turbine will remain inactive until the next planned maintenance(McMorland, Flannigan, et al., 2022). This allows for repairs and regular maintenance to be conducted during this period, optimizing the use of resources. By improving reliability and reducing the need for costly maintenance tasks, the overall maintenance cost can be minimized. The number of scheduled maintenance intervals per year is determined by considering factors such as capacity, weather conditions, and the levelized production cost at each site. The maintenance strategy for power generation takes into account the impact of the rate of power generation on the level of wear and tear experienced by the turbine, thus influencing the preventive maintenance approach. The objective of implementing a preventive maintenance strategy is to maximize the efficiency of both the production plan and the economic maintenance plan. This strategy offers several benefits, including:

- The elimination of unscheduled maintenance
- The ability to plan for maintenance during favorable weather conditions
- Reducing the impact of unpredictable weather
- Optimizing the utilization of service vessels
- Avoiding excessive spare stock
- Combining maintenance and repairs
- Optimizing maintenance tasks and
- Contributing to an effective asset maintenance plan(McMorland, Collu, et al., 2022).

2.5. Integration of Digital Twins with Wind Energy Systems

2.5.1. Studies on the Integration of Digital Twins with Wind Turbine Technology

As mentioned in the previous section, there are many applications of digital twins in the field of wind turbines. For this reason, several studies have focused on the analysis of digital twins and their integration into wind turbine technology. Therefore, the research on how digital twins are used in wind energy is discussed below. The first two studies focus on offshore wind turbines as the most widespread and innovative form of wind energy. One study proposes a digital twin (DT) framework for acquiring and integrating different types of information used throughout the life cycle of floating wind turbines (FWTs). A digital 3D model serves as a means to allow real-time synchronization and inversion of sensor data, facilitating simulation and analysis of the overall FWT state. The proposed framework is evaluated through a case study which includes a simulation process, mechanical analysis and anomaly detection(Y. Liu et al., 2023b). A second study supports that digital twins are used as representations of real floating offshore wind turbines (FOWT) systems based on models and developed using real data. They serve to optimize and simplify the development of offshore wind power plants, including design, planning, installation, operation and management. Through their development stages digital twins are capable of detecting early fault and so reducing maintenance costs(Ciuriuc et al., 2022). Another research states that the use of a DT model enhances the management of the entire life cycle and value chain of wind power plants. It provides a two-way link between virtual models and their physical counterparts for real-time data integration and analysis. It offers predictive maintenance capabilities, reducing costs and improving productivity. It supports decision making by accurately reproducing system behavior and scenario analysis. It facilitates the planning, monitoring and control of wind power plants, contributing to the objectives of sustainable development(*(PDF) Digital Twin for the Management of Wind Power Plants*, n.d.).

2.5.2. Innovations in Monitoring, Predictive Maintenance, and Performance Optimization

In order to reduce operational expenditures and overall energy costs for both offshore and onshore wind turbines, the primary strategy is to improve turbine availability through predictive maintenance of critical components. This approach helps to prevent unexpected maintenance and expensive offshore transportation and operation costs. The achievement of these goals is facilitated by the use of online monitoring, which relies on computationally inexpensive digital twin models. By dynamically optimizing the turbine overhaul plan and scheduled maintenance intervals, the availability of wind turbines is maximized. The focus of predictive maintenance is on the components that pose the greatest risk to turbine availability. The power train system, which includes the rotor, main bearings, gearbox, generator, and power converter, is responsible for the majority of turbine failures and downtime. For offshore systems, particularly floating offshore wind turbines, the overall impact is expected to be greater due to their higher power ranges, larger components, and a wider range of excitation sources. The powertrain system consists of the rotor, gearbox, generator, main and high-speed shafts, and main bearings, which collectively account for most of the turbine downtime. It is anticipated that real-time monitoring of the lifespan of critical powertrain components in large floating offshore wind turbines will be available in the near future. Condition-based maintenance includes predictive maintenance, which involves assessing the remaining useful lifetime of components. In the case of wind turbines, this approach can be used to set alarms for critical components based on their severity and deviation from the expected lifespan, prompting the operator to take appropriate actions. These actions can then be integrated with scheduled maintenance for further investigation. Alternatively, the outputs of predictive maintenance can be incorporated into the decision support system at the wind farm level to determine the operating point of the turbines based on their condition. Depending on the risk analysis, the actions can also be integrated into the protection system. Recent literature suggests the use of digital twin models for both predictive and condition-based maintenance in various application domains. Computational models of system components that update themselves based on operational measurements enable cost-effective real-time monitoring of critical components. Additionally, to overcome the limitations of current simulation and monitoring technologies, digital twin technology has been adopted in both the installation and operation stages of wind turbines. By combining comprehensive monitoring and multi-level simulation technology, digital twin provides a more accurate representation of structures in complex environments. Serving as a digital representation of the asset, digital twin allows for information accumulation and integration throughout the entire life cycle. Digital twin has already been successfully applied in various areas, especially in relation to offshore wind turbines. For example, one study integrated digital twin with computational fluid dynamics methods to accurately predict forces in mooring systems during extreme wave conditions. Another research endeavor employed sensor data as input within the framework of the Decision Tree (DT) to estimate observed point loads and stresses, while also incorporating a stochastic degradation model to forecast the remaining service life. The utilization of the DT methodology in the realm of support structures for offshore wind turbines (FWT) has received extensive scrutiny in relation to real-time monitoring, fault diagnosis, and operational optimization. Nevertheless, the majority of investigations have predominantly focused on real-time monitoring during the operation and maintenance (O&M) phase, thereby overlooking other analytical technologies and stages (Katsidoniotaki et al., 2022; Moghadam & Nejad, 2022b, 2022c; Wang et al., 2021).

2.5.3. Gaps and Opportunities in Current Research

The industry of digital twins in wind turbine technology is currently in development and is growing more and more over time in conjunction with the development of technology. Today, digital twins cover a wide range in the field of wind turbines but do not fully meet the needs for optimization of these systems. All research from time to time focuses on the role of digital twins in terms of the operational part of wind turbines and how they can contribute to this by finding optimal scenarios for their better functionality. At the same time, as is well known, digital twins are a representation of physical systems in digital form, enabling two-way communication between them. Through this feature many issues will be solved such as the maintenance of systems: preventive or at the same time when a fault occurs, as well as the management and monitoring of the whole system remotely. In addition, by collecting and using data obtained from the wind turbines, the digital twins will be able to predict their expected lifetime and optimize them to perform better in electricity generation. As can be seen the whole utilization of digital twins is in an early or even experimental stage because they are not fully effective to use. For this reason, many researches have been developed in the field of digital twins from which engineers are trying to yield the best possible and functional model which will be the ultimate tool. But the obstacles and challenges are many and it takes time and further research to overcome them. So far the models that have been presented cover some of the capabilities that digital twins can provide but not all. The main concern of researchers is therefore to complete a model that will provide all the possibilities. The purpose of this paper is to review several research articles that elaborate on the capabilities of digital twins so that a general picture of their concept and role can be obtained. Therefore, for a better understanding we collected every function that a digital twin can provide in order to understand the complete form of the desired final model that is to be used in the field of wind turbines in the near future on a permanent basis from the moment of their installation until the final phase of their life.

2.6. Energy Management and Optimization

2.6.1. Review of Techniques and Strategies in Energy Management for Wind Farms

The wind farm industry is experiencing steady growth in terms of quantity, scale, and complexity, driven by the objective of increasing the proportion of electricity generated from wind energy. This expansion has resulted in significant advancements in the design of large-scale wind energy conversion systems, leading to improved economies of scale. Effective planning of wind farms holds substantial economic advantages for investors and developers. However, this task is not only technologically intricate but also plagued by a significant level of uncertainty, which directly impacts the potential profitability of the project throughout its operational lifespan. The inherent risks associated with wind farm development can make the necessary investments and payback period so significant that they become economically unviable. Therefore, detailed planning, operational strategies, modeling, and optimization are essential to ensure both technical feasibility and financial competitiveness when compared to other conventional forms of energy conversion. In recent years, there has been significant interest in developing operational strategies that treat wind farms as integrated systems in order to achieve various objectives, such as increasing power generation and reducing maintenance requirements. Achieving these operational goals requires estimating the available energy and wind conditions that affect each turbine. The significance of the wind turbine's aerodynamic interaction with dynamic atmospheric resources means that the wake (the decrease in momentum due to power extraction) and its interaction through the wind turbine have the greatest influence on available energy. Therefore, predicting the impact of eddy currents and their interactions forms the foundation of wind farm control strategies aimed at reducing power production losses, tracking power signals, mitigating structural loads, or compensating for wind turbine wear in order to decrease operating and maintenance costs (Hamilton et al., 2022; Herbert-Acero et al., 2014).

2.6.2. Role of Digital Technologies in Energy Optimization

The potential of energy efficiency to prompt economic growth and decrease greenhouse gas emissions is a very crucial issue. Digital technologies play an important role in enhancing energy efficiency through the provision of analytics, efficiency measures, and improved control. The effects of digitalization are likely to be claimed as innovation, the emergence of new business strategies and models, as well as the development of energy-efficient products and services. By accepting digital transformation, we can create smart homes, buildings, energy systems, cities, and low-carbon systems that yield substantial energy savings. Moreover exploiting the power of digitalization, we can maximize energy efficiency and make substantial progress towards achieving energy optimization and sustainable development goals(Sleiti et al., 2022b). DT models are a reliable tool for power plants, which can also be used for other similarly complex mechanical systems. These models can be applied to new and existing power plants to provide the design limits of power plants under different operating conditions, such as changes in weather data, ambient temperature, humidity, variable load, fuel mix, etc(Lamagna et al., 2021). Combined with advanced forecasting, control and optimization techniques, the results of these DT models can improve the performance, reliability, availability, maintainability and flexible operation of power plants. More generally, the role of DT in various areas can be listed as follows:

- To optimize the energy network by regulating the voltage stability which determines the reliability of the system and adding new RES (PV-OFWT) without risk of voltage fluctuation.
- To build a model that will provide the ability to manage and increase the adequacy of an energy system.
- For the management of energy that can be produced from buildings, which is also characterized as a form of RES.
- To find solutions in the transport system to connect to energy management systems, where the energy source is converted into motion.
- DTs can also face challenges such as energy storage, transport and consumption(Subramanian, 2023).

2.6.3. Studies Focusing on Maximizing Energy Output and Efficiency

Digital twins can be used to maximize energy output and efficiency on wind turbines in several ways. Firstly, they provide a virtual representation of the turbine and its operating environment, allowing for planning, design, and construction phases. Secondly, by enhancing the standalone digital twin with real-time data from the turbine, a descriptive digital twin is created. This allows for informed decision-making through data visualization and forms the basis for diagnostic, predictive, prescriptive, and autonomous tools. Additionally, predictive digital twins can be developed using weather forecasts, neural networks, and transfer learning, enabling accurate power prediction and maintenance models for wind farms. Finally, wind turbine digital twins can contribute to grid digitalization by addressing uncertainties in renewable power supply and demand, and by proposing open and closed loop scenarios for future digital grids. Numerous investigations have been undertaken to optimize the energy yield and effectiveness of wind turbines. One particular study endeavored to ascertain the optimal angular velocity and quantity of rotations for the generator winding to maximize energy potential for the purpose of charging batteries. Another study introduced an adaptable hybrid multi-criteria decision-making approach to select the most suitable wind turbine, considering technical, economic, environmental, and customer service factors. Furthermore, research has been conducted on wind power conversion systems with counter-rotating mechanisms, revealing that a configuration featuring a smaller front rotor diameter and a larger rear rotor contributes to enhanced system efficiency. Additionally, an experiment probed the influence of varying wind speeds on wind turbine efficiency and identified an optimum range of wind speeds for attaining peak efficiency. These studies considerably enrich our comprehension of how to optimize energy yield and effectiveness in wind turbines.

2.7. Economic and Environmental Aspects of Wind Energy

2.7.1. Cost-Benefit Analysis of Wind Energy Systems

In light of global population growth and the industrial revolution, there is a pressing need to expand the capacity for power generation worldwide. The current conventional power plants are insufficient to meet the escalating demand for electricity. However, by strategically designing and sizing clusters of renewable energy sources, particularly wind power, microgrid operators can offer an economically and environmentally sustainable solution to augment the power supply. By harnessing wind power, the strain on power plants can be alleviated, effectively reducing peak demands in constrained distribution networks. The advantages of wind power are manifold, including heightened energy revenue, improved system reliability, deferred investments, decreased power losses, and reduced environmental pollution. These benefits not only enhance the performance of the power system but also generate economic value for society. Nevertheless, the integration of wind power into the distribution system poses various challenges that must be carefully considered. These include protection device miscoordination, fundamental changes in the network topology, transmission congestion, bidirectional power flow, and harmonic current injections. In terms of conventional power sources, offshore wind power offers the same advantages as onshore wind power. One of the most significant benefits is its minimal carbon emissions throughout its lifespan, along with negligible emissions of mercury, nitrous oxides, and sulfur oxides. Unlike other forms of electricity generation that rely on fuel, wind power is not subject to price volatility associated with oil, natural gas, biomass, nuclear, and coal. Additionally, wind power does not require large amounts of freshwater, unlike conventional power sources. Although offshore wind power may initially be more expensive than onshore wind power, it offers unique advantages that onshore wind power does not possess. These advantages pertain to location, power generation, transportation, construction, and design. The additional costs associated with offshore wind power may or may not be justified by these benefits. The estimates for offshore wind costs vary greatly, depending on the assumptions made by analysts and the year in which the estimates were conducted. In recent years, there has been a significant increase in commodity prices, as well as the costs associated with turbine construction and installation, both onshore and offshore. Moreover, the methodology used to estimate costs and their potential applications can differ substantially. What is evident is that onshore wind power costs are comparable to those of conventional power sources, whereas offshore wind power costs are higher, surpassing both onshore and conventional electricity by a factor of 2-3. The specific premium price depends on the time and location, but it could reach up to $50/MWh. Given that onshore wind is cost competitive with conventional electricity, the premium is similar for both energy sources, and it may even be higher for onshore wind compared to conventional power (Snyder & Kaiser, 2009; Zietsman et al., 2022).

2.7.2. Environmental Impact Assessments

Wind turbines are an increasingly prevalent form of sustainable energy that possesses the capability to diminish our reliance on non-renewable resources and contribute to the alleviation of climate change. Nevertheless, akin to any developmental venture, the establishment and operation of wind turbines can yield both advantageous and detrimental repercussions on the environment. Before the commencement of a wind farm, it is customary to execute an Environmental Impact Assessment (EIA) in order to assess and manage the potential consequences. The EIA constitutes a comprehensive investigation that meticulously scrutinizes the environmental ramifications of the undertaking and proposes strategies to mitigate or alleviate any unfavorable outcomes. The process encompasses the collection of data pertaining to the site, the analysis of possible repercussions, the engagement with stakeholders, and the formulation of measures to rectify the issues. By conducting an EIA, developers can ensure that wind turbines are implemented and operated in a manner that is both ecologically sustainable and socially conscientious. When evaluating a planning application, the local planning authority will meticulously assess the effect of the project on the surrounding environment, as well as any notable species or habitats present on the premises. Typically, an Environmental Impact Assessment (EIA) becomes necessary when a wind turbine project has the potential to induce substantial environmental, social, or economic ramifications. Numerous factors determine the specific criteria that trigger an EIA. These factors encompass the size of the wind farm, the sensitivity of the location, the potential impact on wildlife or habitats, and the potential impact on local communities. In certain regions, the requirement for an EIA may be predicated on the capacity of the wind farm, the quantity of turbines, or the extent of land area affected. Ultimately, the necessity of an EIA hinges on the specific regulations and guidelines of the region, as well as the distinctive characteristics of the proposed wind turbine project. As part of the planning and wind development consent process for wind energy projects, an Environmental Impact Assessment (EIA) is frequently mandated by law. Developers must adhere to these legal obligations in order to obtain the requisite permits and approvals for their projects. The objective of an EIA is to enable developers to identify and evaluate the potential environmental consequences of their wind energy projects, encompassing impacts on wildlife, habitats, water resources, air quality, noise, and other pertinent factors. By comprehending these potential repercussions, developers can devise their projects in a manner that minimizes harm to the environment and guarantees the acquisition of planning permission. Through the EIA process, developers can obtain a more comprehensive understanding of the environmental context in which their wind energy project will be situated. This comprehension can inform more knowledgeable project planning and design decisions that take into account the environmental considerations of the project site (*Environmental Impact Assessment of Wind Turbines - CWE*, n.d.).

2.7.3. Policy and Regulatory Framework Influencing Wind Energy

Key policy and regulatory barriers to wind energy development include financial barriers, limited government subsidies, adequate funding, legal and regulatory frameworks, coordination among stakeholders, and talent development systems. Financial barriers such as limited government subsidies and adequate funding are considered to be major constraints for the Indian wind energy industry. In China, obstacles to the development of the wind energy industry are divided into institutional factors, economic and financial factors, social factors, technical factors and market factors. The United States faces administrative hurdles, including permitting, zoning and siting processes, that could hamper the expansion of wind capacity. These barriers may hinder the growth of wind energy and require the attention of policymakers to overcome them and promote wind energy development (Cai et al., 2022; Painuly & Wohlgemuth, 2022; Venkatareddy et al., 2022).

2.8. Review of Methodologies and Tools

2.8.1. Tools and Technologies Utilized in the Development and Implementation of Digital Twins

Various tools and technologies are utilized in the development and implementation of digital twins. High-fidelity simulations, which serve as virtual representations connected to real assets, are one such tool. Motion sensors, biological sensors, computational intelligence, simulation and visualization tools are also employed in this process. Furthermore, the creation of digital twins often necessitates the integration of technologies and paradigms such as machine learning, the Internet of Things (IoT), and 3D visualization. These technologies facilitate the seamless alignment of digital twins with real-time machine-learning predictions, IoT data streams, and 3D connected visualizations. Additionally, the use of extended reality is mentioned as a means to enable natural and effective training scenarios for robotic operators. In summary, the development and implementation of digital twins require the combination of various tools and technologies to produce reliable and efficient virtual representations of physical assets (Asad et al., 2023; Robles et al., 2023).

2.9. Critical Assessment and Identification of Gaps

2.9.1. Critical Analysis of the Reviewed Literature

The extensive examination of literature pertaining to digital twins for wind turbines offers a comprehensive assessment of the immense potential of this technology within the renewable energy industry. These studies emphasize the intricate relationship between physical assets and their digital counterparts, highlighting how digital twins facilitate real-time monitoring, predictive maintenance, and optimization of wind turbine performance. The literature specifically delves into the effectiveness of various modeling techniques, integration of sensors, and utilization of data analytics in the creation of these digital replicas. It scrutinizes the challenges associated with accuracy, scalability, and cyber security, while recognizing the promising outcomes in terms of cost reduction, enhanced operational efficiency, and prolonged lifespan of assets. Furthermore, the literature underscores the dynamic nature of digital twins and calls for further research and development to fully harness their capabilities in shaping a sustainable energy future.

2.9.2. Identification of Research Gaps and Unexplored Areas

Following on from the previous paragraph:Upon conducting a thorough examination of scientific literature pertaining to digital twins in wind turbine operations, encompassing both onshore and offshore applications within the renewable energy sector, several crucial insights emerge. Collectively, these articles underscore the remarkable progress achieved in harnessing digital twin technology to optimize the performance, maintenance, and monitoring of wind turbines. Nevertheless, a comprehensive analysis reveals certain gaps in these studies, including the need for standardized data collection methodologies, the establishment of universally applicable modeling frameworks, and the implementation of robust cyber security protocols for digital twin systems. Moreover, while these articles exemplify the potential for cost reduction and heightened efficiency through digital twins, it remains imperative to substantiate their long-term viability and effectiveness across diverse environmental conditions and operational contexts with empirical evidence.This important survey highlights the continued development of digital twin applications in the renewable energy sector and highlights the need for further interdisciplinary research and industry collaboration to realize their full potential in advancing sustainable energy solutions.

2.9.3. Justification for the Necessity of Further Research in the Field

Further research in the field of digital twins for wind turbines and wind farms is imperative due to its potential to revolutionize the efficiency, reliability, and sustainability of renewable energy production. Digital twins offer a comprehensive virtual representation of physical assets, enabling real-time monitoring, predictive maintenance, and performance optimization. In the context of wind energy, this technology could significantly enhance turbine operations by facilitating precise simulations, identifying potential faults before they occur, and fine-tuning operations for optimal energy output. Moreover, the dynamic nature of wind patterns and turbine behavior necessitates continued research to refine these digital replicas, ensuring their accuracy and adaptability to varying environmental conditions. Advancements in digital twins for wind turbines and farms promise not only increased energy generation but also cost reductions, improved safety, and a more sustainable energy future, making further research a crucial pursuit for the advancement of this technology.

3. Methodology

3.1. Data Acquisition and Data Check

For the construction of the preventive maintenance and fault detection models of a park consisting of 10 wind turbines, the data utilized and obtained are sourced from real-world measurements. Specifically, these data are derived from a Vestas V52 wind turbine located at Dundalk Institute of Technology in Ireland. The dataset is comprehensive and dynamic, capturing various operational parameters of the wind turbine. The dataset is comprehensive and dynamic, capturing various operational parameters of the wind turbine. This dataset includes the following factors:

Parameter	**Description**
Timestamps	*Timestampsin10-minuteintervals*
WindSpeed	*Average10-minutewindspeed(m/s)**
StdDevWindSpeed	*Windspeedstandarddeviation(m/s) in10-minuteperiod**
WindDirAbs	*Average10-minuteabsolutewinddirection(deg)**
WindDirRel	*RelativedirectionofnacellewithrespecttoWindDirAbs(deg)*
Power	*Average10-minutepoweroutput(kW)*
MaxPower	*Maximum10-minutepoweroutput(kW)*
MinPower	*Minimum10-minutepoweroutput(kW)*
StdDevPower	*Poweroutputstandarddeviation(kW)in10-minuteperiod*
AvgRPow	*Average10-minutereactivepoweroutput(kVAR)*
GenRPM	*Electricalgeneratorrevsperminute(RPM)*
RotorRPM	*Windturbinerotorrevsperminute(RPM)*
EnvirTemp	*Environmentaltemperatureoutsideofnacelle(°C)*
NacelTemp	*Temperatureinsidenacellespace(°C)*
GearOilTemp	*Gearboxoiltemperature(°C)*
GearBearTemp	*Gearboxbearingtemperature(°C)*
GenTemp	*Generatortemperature(notactive-999)*
GenPh1Temp	*Generatorphase1windingtemperature(°C)*
GenPh2Temp	*Generatorphase3windingtemperature(°C)*
GenPh3Temp	*Generatorphase3windingtemperature(°C)*
GenBearTemp	*Generatorbearingtemperature(°C)*

Each factor, critical to the turbine's performance and health, is recorded and updated every 10 minutes, providing a rich and detailed time series. This extensive data collection spans a significant period, beginning on 30 January 2006 and continuing until 12 March 2020. The use of real, time-specific data from a functioning wind turbine ensures that the models developed are grounded in actual operational conditions, enhancing their relevance and applicability in predicting maintenance needs and detecting faults in similar wind turbine setups (Byrne & MacArtain, 2022, 2023).

A basic error check is the next crucial step in refining the dataset, conducted to ensure its integrity and usability for modeling purposes. This error check encompasses several key aspects. Firstly, the dataset is meticulously scanned for any missing values across all factors for each of the 10 wind turbines, as missing data can significantly impact the accuracy and performance of predictive models, making this a vital step. Secondly, the data types for each factor are verified, ensuring they are correctly formatted and consistent throughout the dataset. This is essential for seamless processing and analysis. Thirdly, a thorough examination for any unusual or out-of-range values within each factor for each turbine is conducted. Such anomalies could indicate errors in data generation or potential edge cases in turbine behavior, which need to be addressed. In addition to these checks, a separate algorithm has been developed to create an additional file that maps the correlations between the different factors. This correlation analysis is a powerful tool in understanding the interdependencies and interactions among the variables like wind speed, temperature, output power, rotor speed, and vibration levels. By understanding these correlations, the quality of the error check can be enhanced, ensuring that the dataset is not only free from basic errors but also reflects realistic operational relationships. This comprehensive approach to data verification lays a solid foundation for building accurate and reliable predictive maintenance and fault detection models for wind turbines.

3.2. DataProcessing and Analysis

Following the meticulous data check that was conducted, the subsequent phase involves an extensive data processing and analysis procedure, which is critical for the construction of three models aimed at enhancing the operational efficiency of a wind farm. This phase commences with data cleaning, a crucial step in which any inconsistencies, outliers, or irrelevant data points identified during the data check are addressed and eliminated. This process ensures that the dataset is streamlined, containing only efficient and relevant values, thus establishing a robust foundation for model development. The cleaned dataset then undergoes a detailed analysis. This analysis is centered on dissecting and understanding the intricate patterns and trends within the data, with a particular focus on different factors each time depending on what is needed to be found. This level of analysis is key in identifying subtle nuances that could signal the onset of faults or indicate maintenance requirements. Equipped with this refined and deeply understood dataset, the research then moves towards the core objective: constructing three sophisticated models – a digital twin, a predictive maintenance model and a fault detection model. For the digital twin model, the data used were obtained as follows: having the original **Vestas V52 Wind Turbine, 10-minute SCADA Data, 2006-2020 - Dundalk Institute of Technology, Ireland.zip** dataset, we kept the necessary columns **WindSpeed:** Average 10-minute wind speed (m/s), **WindDirAbs:** Average 10-minute absolute wind direction (deg), **Power:** Average 10-minute power output (kW) and we named this new csv file **VestasV52_10_min_raw_A_DT.csv**. A code was then created which took over the cleaning of this data to be used in the alternatives of machine learning of the digital twin model (Decision Trees, Random Forest, Artificial Neural Networks (ANN), and Convolutional Neural Networks (CNN)).So this algorithm created a new csv file with the cleaned data which is

path_to_cleaned_dataset. csv and it was used for the digital twin construction as an input for the four algorithms. The same procedure was done for the Predictive Maintenance and Fault Detection models, except that this time the remaining columns of the original dataset were used, then again taking the file with the cleaned data for the algorithms. These models are specially designed to preemptively identify maintenance needs and detect operational anomalies in the wind turbines. To accomplish this, advanced algorithms that harness the processed data are employed. These algorithms are customized to effectively interpret the data, enabling them to predict potential issues and identify faults with high precision. This data-driven approach is poised to significantly enhance the reliability and efficiency of wind farm operations, representing a notable advancement in the field of renewable energy management.

3.3. Digital Twinning through AI-based forecasting methods

In this thesis a multi-faceted methodology centered around the utilization of a digital twin model was employed. This sophisticated virtual replica of the onshore wind farm served as the foundation for developing and testing various analytical approaches. The initial step involved the creation of a model that could accurately forecast the power output of the wind farm. This model incorporated crucial factors such as wind speed and direction. Building upon this foundation, four different machine learning techniques were examined: Decision Trees, Random Forest, Artificial Neural Networks (ANN), and Convolutional Neural Networks (CNN). To determine the most effective approach for the specific application, a thorough evaluation was conducted on each of these models, taking into account their R-squared (R^2), mean squared error (MSE), and root mean squared error (RMSE) metrics. This comparative analysis was crucial in determining the best fit model for predicting power output under varying environmental conditions. As mentioned in the previous section, the cleaned data originated from the following code(Figure 1. Algorithm for Data Cleaning)were used as inputs to create these four algorithms (Decision Trees, Random Forest, Artificial Neural Networks (ANN), and Convolutional Neural Networks (CNN)).

```python
# -*- coding: utf-8 -*-
"""
Created on Thu Jan 11 17:19:14 2024

@author: User
"""

import pandas as pd
import numpy as np
from scipy import stats

# Load dataset

file_path = r"C:\Users\nikos\OneDrive\Desktop\VestasV52_10_min_raw_A_DT.csv"
df = pd.read_csv(file_path, delimiter=';')

df['Timestamps'] = pd.to_datetime(df['Timestamps'], format='%d/%m/%Y %H:%M')


# Function to replace outliers with the mean of previous and next values
def replace_outliers(series):
    for i in range(1, len(series) - 1):
        if np.abs(stats.zscore([series[i]])) > 3:  # Outlier threshold
            series[i] = np.mean([series[i - 1], series[i + 1]])
    return series

# Apply the function to numerical columns
for col in ['WindSpeed', 'WindDirAbs', 'Power']:
    df[col] = replace_outliers(df[col])

# Remove rows with negative values
df = df[(df['WindSpeed'] >= 0) & (df['WindDirAbs'] >= 0) & (df['Power'] >= 0)]

# Reset index after dropping rows
df = df.reset_index(drop=True)

# Save cleaned dataset
df.to_csv('path_to_cleaned_dataset.csv', index=False)

# Optional: Display the first few rows to verify changes
print(df.head())
```

Figure 1. Algorithm for Data Cleaning

3.3.1. The Decision Trees Algorithm

```
import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.tree import DecisionTreeRegressor
from sklearn.metrics import mean_squared_error, r2_score
import math
import matplotlib.pyplot as plt

# Load dataset
file_path = r"C:\Users\User\Downloads\path_to_cleaned_dataset.csv"  # Replace with your file path
df = pd.read_csv(file_path, delimiter=';')

# Preprocess data
df['Timestamps'] = pd.to_datetime(df['Timestamps'], format='%d/%m/%Y %H:%M', errors='coerce')
df['Hour'] = df['Timestamps'].dt.hour  # Extract hour as a feature

# Splitting the dataset into features (X) and target (y)
X = df[['Timestamps', 'WindSpeed', 'WindDirAbs', 'Hour']]  # Include Timestamps
y = df['Power']

# Splitting the dataset into training (75%) and testing (25%) sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.25, random_state=42)

# Save timestamps for test set
timestamps_test = X_test['Timestamps']

# Drop Timestamps from X_train and X_test for model training
X_train.drop(['Timestamps'], axis=1, inplace=True)
X_test.drop(['Timestamps'], axis=1, inplace=True)

# Decision Tree Model
model = DecisionTreeRegressor(random_state=42)

# Training the model
model.fit(X_train, y_train)

# Predicting the Test set results
y_pred = model.predict(X_test)

# Create a DataFrame for test set with timestamps
test_results = pd.DataFrame({'Timestamps': timestamps_test, 'Actual': y_test, 'Predicted': y_pred}).sort_values(by='Timestamps')

# Filter for the last 12 hours
max_date = test_results['Timestamps'].max()
threshold_date = max_date - pd.Timedelta(hours=12)
```

Figure 2. Decision Trees Algorithm 1/2

```python

# Predicting the Test set results
y_pred = model.predict(X_test)

# Create a DataFrame for test set with timestamps
test_results = pd.DataFrame({'Timestamps': timestamps_test, 'Actual': y_test, 'Predicted': y_pred})

# Filter for the last 12 hours
max_date = test_results['Timestamps'].max()
threshold_date = max_date - pd.Timedelta(hours=12)
filtered_test_results = test_results[test_results['Timestamps'] >= threshold_date]

# Plotting
plt.figure(figsize=(15, 6))
width = 0.35  # Width of the bars
ind = np.arange(len(filtered_test_results))  # X locations for the groups

plt.bar(ind - width/2, filtered_test_results['Actual'], width, label='Actual')
plt.bar(ind + width/2, filtered_test_results['Predicted'], width, label='Predicted')

plt.xlabel('Timestamp')
plt.ylabel('Power')
plt.title('Actual vs Predicted Values for Last 12 Hours')
plt.xticks(ind, filtered_test_results['Timestamps'].dt.strftime('%Y-%m-%d %H:%M'), rotation=45)
plt.legend()
plt.tight_layout()
plt.show()

```

Figure 3. Decision Trees Algorithm 2/2

This algorithm(Figure 2. Decision Trees Algorithm 1/2,Figure 3. Decision Trees Algorithm 2/2) is a multi-step procedure designed to process, analyze, and visualize data using various Python libraries such as Pandas, NumPy, scikit-learn, and Matplotlib. Initially, essential libraries are imported for data manipulation (Pandas, NumPy), machine learning (scikit-learn), mathematical functions (math), and data visualization (matplotlib.pyplot). The process begins by loading a dataset from a CSV file into a DataFrame, where values are separated by semicolons. The data is then preprocessed; for example, timestamps are converted to datetime objects, and a new 'Hour' column is extracted for later use. The dataset is subsequently divided into features and a target variable, with features including 'Timestamps', 'WindSpeed', 'WindDirAbs', and 'Hour', and the target variable being 'Power'. A train-test split follows, allocating 75% of the data for training and 25% for testing, ensuring reproducibility with a fixed random state. Before model training, 'Timestamps' are saved from the test set and dropped from the training and test features. A DecisionTreeRegressor model is then created, trained with the training data, and used to predict the target variable for the test set. A new DataFrame is created to compare actual and predicted values, which is then filtered to focus on the last 12 hours of data. Finally, the results of the last 12 hours are visualized using a bar chart, meticulously crafted with Matplotlib, showcasing settings for figure size, bar width, labels, title, and legend. This comprehensive code effectively handles the complete process from data loading to model training, prediction, and result visualization for the dataset's most recent 12 hours. The next diagram (Diagram 1. Decision Trees Flowchart)is the flow chart of the algorithm showing each step of the algorithm.

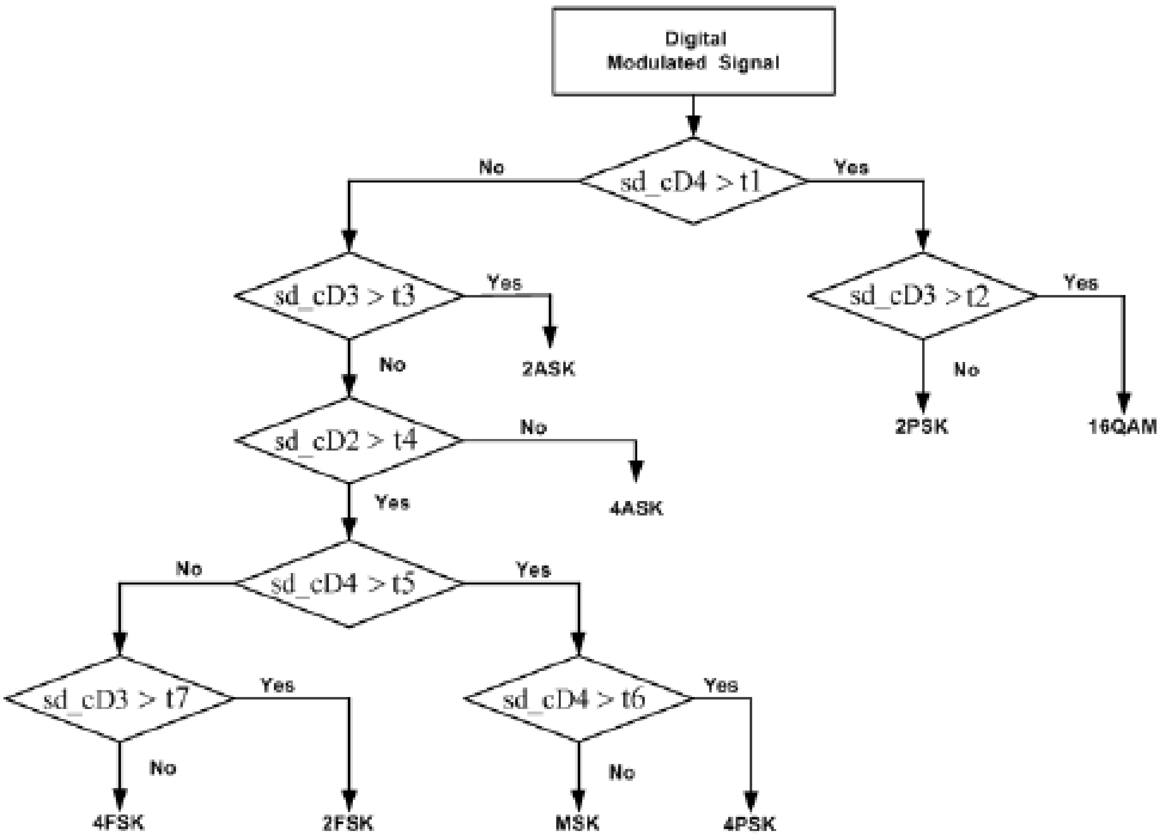

Diagram 1. Decision Trees Flowchart(*Flowchart-of-Decision-Tree-Classifier.Png (494×356)*, n.d.)

3.3.2. The Random Forest Algorithm

```python
import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestRegressor
from sklearn.metrics import mean_squared_error, r2_score
import math

# Load dataset
file_path = r"C:\Users\User\Downloads\path_to_cleaned_dataset.csv"  # Replace with your file path
df = pd.read_csv(file_path, delimiter=';')

# Preprocess data
df['Timestamps'] = pd.to_datetime(df['Timestamps'], format='%d/%m/%Y %H:%M', errors='coerce')
df['Hour'] = df['Timestamps'].dt.hour  # Extract hour as a feature
df.drop(['Timestamps'], axis=1, inplace=True)  # Drop original Timestamps column

# Splitting the dataset into features (X) and target (y)
X = df[['WindSpeed', 'WindDirAbs', 'Hour']]  # Including engineered hour feature
y = df['Power']

# Splitting the dataset into training (75%) and testing (25%) sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.25, random_state=42)

# Random Forest Model
model = RandomForestRegressor(n_estimators=100, random_state=42)  # 100 trees in the forest

# Training the model
model.fit(X_train, y_train)

# Predicting the Test set results
y_pred = model.predict(X_test)

# Model Evaluation
mse = mean_squared_error(y_test, y_pred)
rmse = math.sqrt(mse)
r2 = r2_score(y_test, y_pred)

print(f'Mean Squared Error: {mse}')
print(f'Root Mean Squared Error: {rmse}')
print(f'R^2 Score: {r2}')
```

Figure 4. Random Forest Algorithm 1/2

```python
# Save the Random Forest model
import joblib

model_filename = 'random_forest_model.pkl'
joblib.dump(model, model_filename)
print(f'Random Forest Model saved as {model_filename}')


from sklearn.metrics import mean_squared_error, r2_score
import matplotlib.pyplot as plt

# ... (your previous code)

# Calculate metrics for both training and testing data
y_train_pred = model.predict(X_train)
y_test_pred = model.predict(X_test)

mse_train = mean_squared_error(y_train, y_train_pred)
mse_test = mean_squared_error(y_test, y_test_pred)

print(f'Training MSE: {mse_train}')
print(f'Testing MSE: {mse_test}')

# Plot learning curves
plt.figure(figsize=(10, 6))
plt.plot(y_train, label='Actual Train')
plt.plot(y_train_pred, label='Predicted Train')
plt.plot(y_test, label='Actual Test')
plt.plot(y_test_pred, label='Predicted Test')
plt.xlabel('Sample')
plt.ylabel('Power')
plt.title('Actual vs. Predicted Values')
plt.legend()
plt.show()
```

Figure 5. Random Forest Algorithm 2/2

The Random Forest script(Figure 4. Random Forest Algorithm 1/2,Figure 5. Random Forest Algorithm 2/2)is meticulously organized into a sequence of distinct steps, each serving a unique purpose in the data analysis process. Initially, it commences by importing essential Python libraries: pandas for data manipulation, numpy for numerical operations, train_test_split from sklearn.model_selection for dividing data into training and testing sets, RandomForestRegressor from sklearn.ensemble to construct the regression model, and mean_squared_error and r2_score from sklearn.metrics for model evaluation, along with math for mathematical functions. Subsequently, the dataset is loaded into a DataFrame, df, using pandas.read_csv(), with a semicolon delimiter indicating separated data values. In the preprocessing stage, the 'Timestamps' column is converted into a datetime format, handling parsing errors with the errors='coerce' parameter. An 'Hour' feature is extracted from 'Timestamps' and added to the DataFrame, followed by dropping the original 'Timestamps' column as it becomes redundant. The dataset is then split into features (X) and the target variable (y), using 'WindSpeed', 'WindDirAbs', and 'Hour' as predictors, and 'Power' as the target. A train_test_split() function further divides the data into training and testing sets, maintaining a 75-25% split and ensuring reproducibility with the random_state parameter. A RandomForestRegressor model is instantiated with 100 decision trees and trained on the training data. The model's predictions for the test set are evaluated using two metrics: Mean Squared Error (MSE) and R^2 score. MSE is calculated by comparing predicted values to actual ones, and its square root, the Root Mean Squared Error (RMSE), measures the model's accuracy in the response variable's unit. The R^2 score assesses how well the model replicates observed outcomes based on the variation explained by the model. Each step in this script, from data loading and preprocessing to model training and evaluation, serves a specific function. The choice of a Random Forest Regressor is significant for its ability to manage non-linear relationships and feature interactions with minimal hyperparameter tuning, rendering it a robust option for various regression tasks. The next diagram(Diagram 2. Random Forest Flowchart)is the flow chart of the algorithm showing each step of the algorithm.

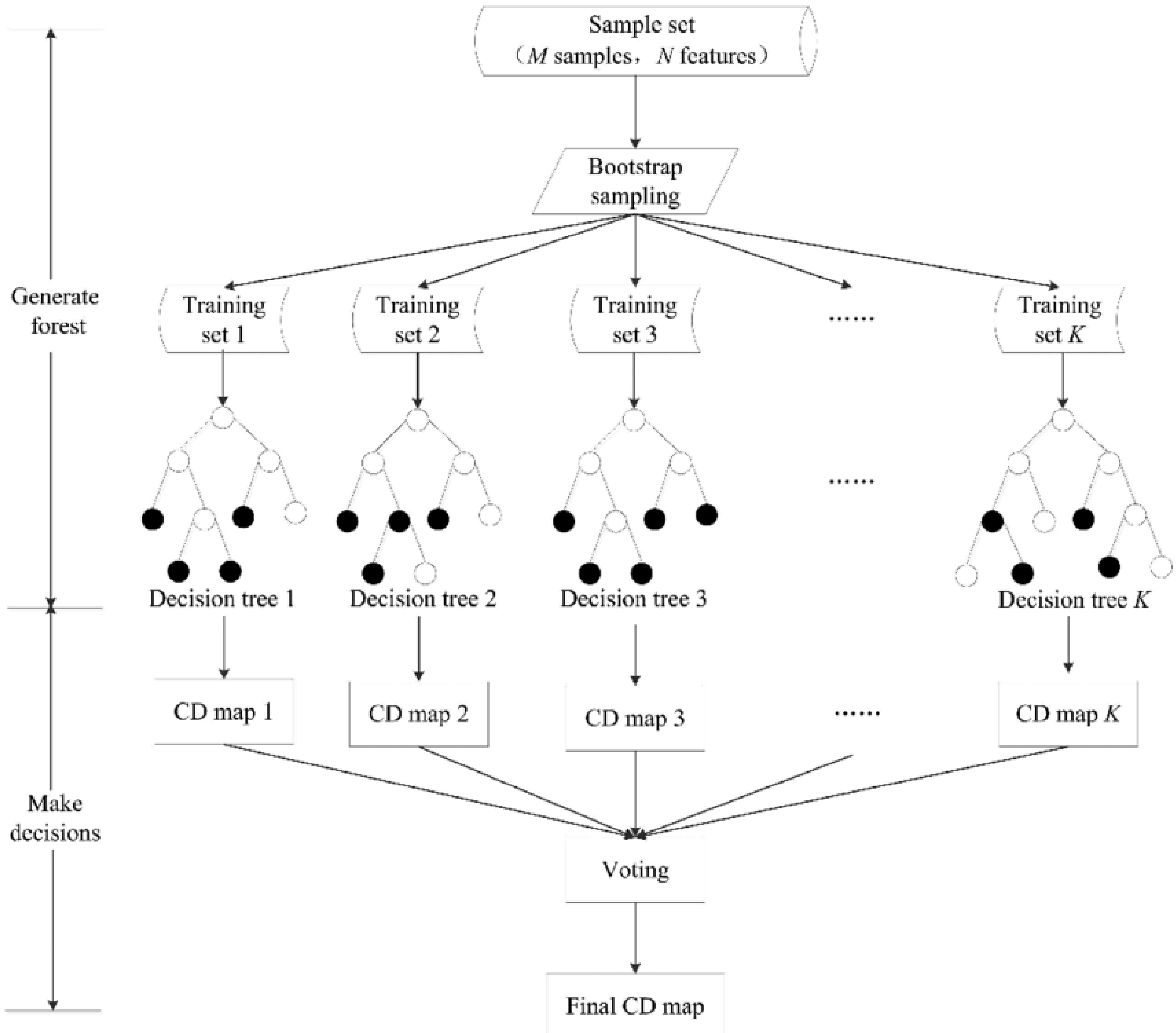

Diagram 2. Random Forest Flowchart (*The-Flow-Chart-of-Random-Forest-Classifier.Png (850×752)*, n.d.)

3.3.3. The Artificial Neural Networks (ANN) Algorithm

```python
import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.metrics import mean_squared_error, r2_score
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Dense
from tensorflow.keras.optimizers import Adam
import math

# Load dataset
file_path = r"C:\Users\User\Downloads\path_to_cleaned_dataset.csv"  # Replace with your file path
df = pd.read_csv(file_path, delimiter=';')

# Preprocess data
df['Timestamps'] = pd.to_datetime(df['Timestamps'], format='%d/%m/%Y %H:%M', errors='coerce')
df['Hour'] = df['Timestamps'].dt.hour  # Extract hour as a feature
df.drop(['Timestamps'], axis=1, inplace=True)  # Drop original Timestamps column

# Splitting the dataset into features (X) and target (y)
X = df[['WindSpeed', 'WindDirAbs', 'Hour']]  # Including engineered hour feature
y = df['Power']

# Splitting the dataset into training (75%) and testing (25%) sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.25, random_state=42)

# Standardizing the features (important for neural networks)
scaler = StandardScaler()
X_train = scaler.fit_transform(X_train)
X_test = scaler.transform(X_test)

# ANN Model
model = Sequential()

# Input layer and first hidden layer
# units: number of neurons, usually a number between the size of the input layer and the size of the output layer
# activation: relu is a common choice, good for hidden layers
model.add(Dense(units=64, activation='relu', input_dim=X_train.shape[1]))

# Second hidden layer
# Adding more layers increases the model's capacity to learn complex patterns
model.add(Dense(units=32, activation='relu'))

# Output layer
# units: 1, since we are predicting a single value (Power)
```

Figure 6. ANN Algorithm 1/2

```
model.add(Dense(units=1))
# Compiling the model
# optimizer: Adam is an efficient variant of Stochastic Gradient Descent
# loss: mean_squared_error, common choice for regression problems
model.compile(optimizer=Adam(learning_rate=0.001), loss='mean_squared_error')
# Training the model
# epochs: number of times the entire training set will be passed through the network
# batch_size: number of samples per gradient update, smaller batch size means more updates in one epoch
model.fit(X_train, y_train, epochs=100, batch_size=32)

# Predicting the Test set results
y_pred = model.predict(X_test)

# Model Evaluation
mse = mean_squared_error(y_test, y_pred)
rmse = math.sqrt(mse)
r2 = r2_score(y_test, y_pred)

print(f'Mean Squared Error: {mse}')
print(f'Root Mean Squared Error: {rmse}')
print(f'R^2 Score: {r2}')

import matplotlib.pyplot as plt

# Assuming y_test and y_pred are the actual and predicted values, respectively
last_12_actual = y_test[-12:]
last_12_predicted = y_pred[-12:]
# Create a list of indices for the last 12 values (e.g., 1, 2, ..., 12)
indices = np.arange(1, 13)
# Create a bar chart
plt.figure(figsize=(10, 6))
plt.bar(indices, last_12_actual, width=0.4, label='Actual', align='center', color='blue')
plt.bar(indices + 0.4, last_12_predicted, width=0.4, label='Predicted', align='center', color='orange')

# Add labels and title
plt.xlabel('Time Step')
plt.ylabel('Value')
plt.title('Actual vs. Predicted Values for the Last 12 Time Steps')
plt.xticks(indices + 0.2, indices)
plt.legend()

# Show the plot
plt.tight_layout()
plt.show()
```

Figure 7. ANN Algorithm 2/2

This script (Figure 6. ANN Algorithm 1/2,Figure 7. ANN Algorithm 2/2) outlines a thorough process for constructing and assessing an Artificial Neural Network (ANN) model using Python, encompassing various phases from data acquisition to model appraisal. The initial step involves importing essential libraries for data handling and processing, including pandas and numpy for data manipulation, train_test_split and StandardScaler from sklearn for data preparation, tensorflow.keras components for ANN construction, and math for mathematical operations. The dataset is then loaded into a pandas DataFramedf using read_csv(), with a semicolon as the delimiter. Data preprocessing involves converting 'Timestamps' to datetime objects, with parsing errors resulting in NaN values, extracting an 'Hour' feature from 'Timestamps' for model use, and removing the original 'Timestamps' column from the DataFrame. In preparing the data for modeling, features (X) such as 'WindSpeed', 'WindDirAbs', and 'Hour', and the target variable (y), 'Power', are defined. The dataset is split into training and testing sets, maintaining a 75-25% split, with train_test_split() ensuring reproducibility through the random_state parameter. Feature standardization is achieved using StandardScaler, a critical step for effective neural network performance. Building the ANN model involves initializing a Sequential model, indicating a linear stack of layers, and adding dense layers: the first as the input layer with 64 neurons using the ReLU activation function, a second hidden layer with 32 neurons, also with ReLU activation, and a single-neuron output layer for predicting the continuous variable 'Power'. The model is compiled with the Adam optimizer and mean squared error loss function, both standard for regression problems. The model training occurs over 100 epochs with a batch size of 32, allowing the model to learn and adjust its weights to minimize the loss function. Model prediction and evaluation are conducted on the test set, using the predict() method. The model's accuracy is gauged using Mean Squared Error (MSE) and the R^2 score, with MSE measuring the average squared difference between estimated and actual values, and RMSE providing error in the same units as the target variable. R^2 score indicates the proportion of variance in the dependent variable that is predictable from the independent variables.Overall, this script exemplifies a structured approach todeveloping and evaluating an ANN model for regression tasks, showcasing meticulous data preprocessing, feature scaling, and a multi-layered neural network, thereby reflecting a comprehensive understanding of the intricacies involved in crafting effective machine learning models. The next diagram (Diagram 3. ANN Flowchart) is the flow chart of the algorithm showing each step of the algorithm.

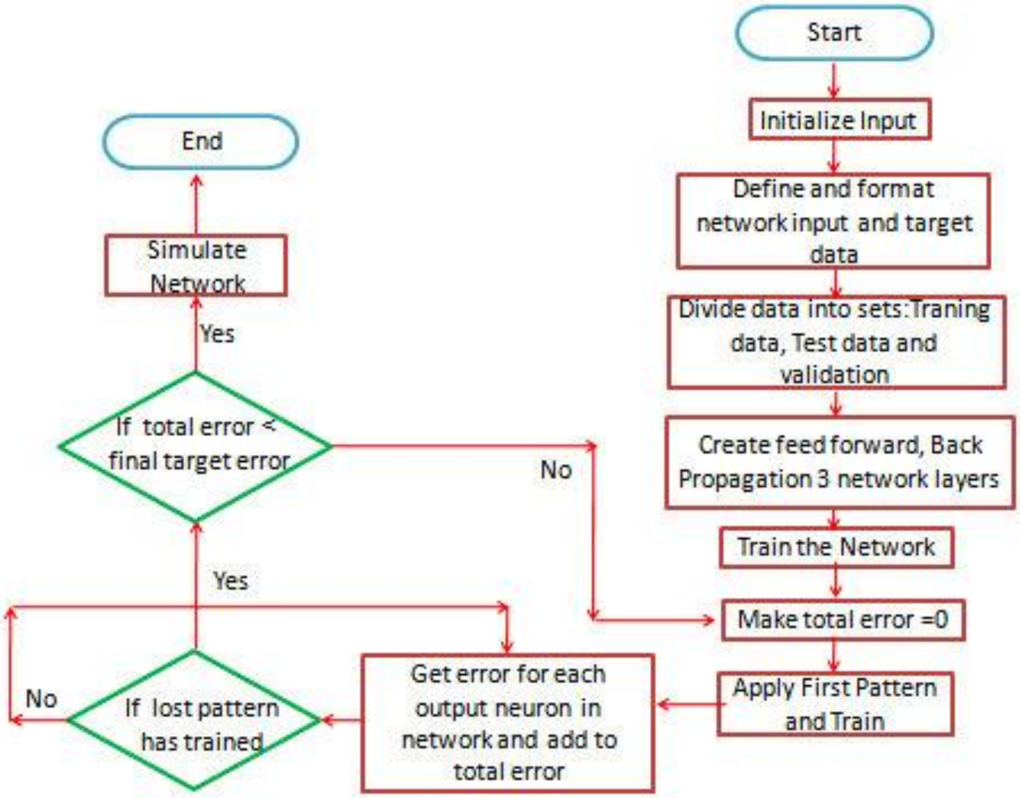

Diagram 3. ANN Flowchart (*Flowchart-for-Artificial-Neural-Network-ANN.Png (464×357)*, n.d.)

3.3.4. The Convolutional Neural Networks (CNN) Algorithm

```python
import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Conv1D, MaxPooling1D, Flatten, Dense
from tensorflow.keras.optimizers import Adam
from tensorflow.keras.callbacks import ModelCheckpoint
from sklearn.metrics import mean_squared_error, r2_score
import math

# Load dataset
file_path = r"C:\Users\User\Downloads\path_to_cleaned_dataset.csv"  # Replace with your file path
df = pd.read_csv(file_path, delimiter=';')

# Preprocess data
df['Timestamps'] = pd.to_datetime(df['Timestamps'], format='%d/%m/%Y %H:%M', errors='coerce')
df['Hour'] = df['Timestamps'].dt.hour  # Extract hour as a feature
df.drop(['Timestamps'], axis=1, inplace=True)  # Drop original Timestamps column

# Splitting the dataset into features (X) and target (y)
X = df[['WindSpeed', 'WindDirAbs', 'Hour']]  # Including engineered hour feature
y = df['Power']

# Splitting the dataset into training (75%) and testing (25%) sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.25, random_state=42)

# Standardizing the features
scaler = StandardScaler()
X_train = scaler.fit_transform(X_train)
X_test = scaler.transform(X_test)

# Reshaping for CNN input
X_train = X_train.reshape(X_train.shape[0], X_train.shape[1], 1)
X_test = X_test.reshape(X_test.shape[0], X_test.shape[1], 1)

# CNN Model
model = Sequential()

# Convolutional layer
model.add(Conv1D(filters=64, kernel_size=2, activation='relu', input_shape=(X_train.shape[1], 1)))

# MaxPooling layer
model.add(MaxPooling1D(pool_size=2))
```

Figure 8. CNN Algorithm 1/3

```python

# Flattening for the fully connected layers
model.add(Flatten())

# Fully connected layer
model.add(Dense(units=50, activation='relu'))

# Output layer
model.add(Dense(units=1))

# Compiling the model
model.compile(optimizer=Adam(learning_rate=0.001), loss='mean_squared_error')

# Model Checkpoint
checkpoint = ModelCheckpoint('best_model.h5', monitor='val_loss', mode='min', save_best_only=True)

# Training the model
model.fit(X_train, y_train, epochs=3, batch_size=32, callbacks=[checkpoint], validation_split=0.2)

# Load and Evaluate the Best Model
model.load_weights('best_model.h5')
y_pred = model.predict(X_test)

# Model Evaluation
mse = mean_squared_error(y_test, y_pred)
rmse = math.sqrt(mse)
r2 = r2_score(y_test, y_pred)

print(f'Mean Squared Error: {mse}')
print(f'Root Mean Squared Error: {rmse}')
print(f'R^2 Score: {r2}')

import matplotlib.pyplot as plt

# Assuming y_test and y_pred are the actual and predicted values, respectively
last_12_actual = y_test[-12:]
last_12_predicted = y_pred[-12:]

# Create a list of indices for the last 12 values (e.g., 1, 2, ..., 12)
indices = np.arange(1, 13)

# Create a bar chart
plt.figure(figsize=(10, 6))
plt.bar(indices, last_12_actual, width=0.4, label='Actual', align='center', color='blue')
plt.bar(indices + 0.4, last_12_predicted, width=0.4, label='Predicted', align='center', color='orange')
```

Figure 9. CNN Algorithm 2/3

```python

# Add labels and title
plt.xlabel('Time Step')
plt.ylabel('Value')
plt.title('Actual vs. Predicted Values for the Last 12 Time Steps')
plt.xticks(indices + 0.2, indices)
plt.legend()

# Show the plot
plt.tight_layout()
plt.show()
```

Figure 10. CNN Algorithm 3/3

This script (Figure 8. CNN Algorithm 1/3,Figure 9. CNN Algorithm 2/3,Figure 10. CNN Algorithm 3/3) provides a detailed walkthrough of constructing, training, and evaluating a Convolutional Neural Network (CNN) for regression analysis on a time series dataset using Python. It starts with importing necessary libraries for data manipulation (pandas, numpy), model preparation and evaluation (scikit-learn), CNN construction (TensorFlow, Keras), and plotting (matplotlib). The dataset, sourced from a CSV file, is loaded into a DataFramedf with a semicolon as the delimiter. Data preprocessing includes converting 'Timestamps' into datetime objects, handling parsing errors with NaN values, extracting an 'Hour' feature from 'Timestamps', and dropping the original column, thus tailoring the dataset for time series analysis. The dataset is then split into features (X) - 'WindSpeed', 'WindDirAbs', and 'Hour', and the target variable (y) - 'Power', followed by a standard 75%-25% division into training and testing sets. Feature standardization is achieved using StandardScaler, an essential step for neural networks. The data is reshaped to meet CNN input requirements, anticipating a three-dimensional format. Building the CNN model involves initializing a Sequential model, indicative of a linear layer stack. A Conv1D layer with 64 filters and a kernel size of 2 is used for sequential data feature extraction, followed by a MaxPooling1D layer to reduce data dimensionality. The data is then flattened, and a fully connected Dense layer with 50 neurons is added, concluding with a single-neuron output layer for the regression task. The model is compiled using the Adam optimizer and mean squared error loss function, typical for regression problems. Model checkpointing is set up to retain the model with the lowest validation loss, ensuring the best performance is captured. The model undergoes training for 3 epochs with a batch size of 32, including callbacks for checkpointing and a 20% validation split to monitor performance on unseen data. Post-training, the best-performing model is loaded for test set predictions, evaluated using Mean Squared Error (MSE) and R^2 score. MSE measures the average squared prediction errors, with RMSE offering a more interpretable version, and R^2 score indicates the proportion of variance in the dependent variable predictable from the independent variables. The script concludes by visualizing actual versus predicted values for the dataset's last 12 time steps using a matplotlib bar chart, providing a clear visual comparison between the model's predictions and true values. This script effectively demonstrates a CNN's application for time series regression, typically used for spatial data. Its use for time series data, which often contain local patterns or dependencies, is intriguing. The approach, including standard scaling, model checkpointing, and performancemetrics, is comprehensive. The final visualization presents the model's performance in an accessible, tangible manner, making it more understandable for audiences. The next diagram (Diagram 4. CNN Flowchart)is the flow chart of the algorithm showing each step of the algorithm.

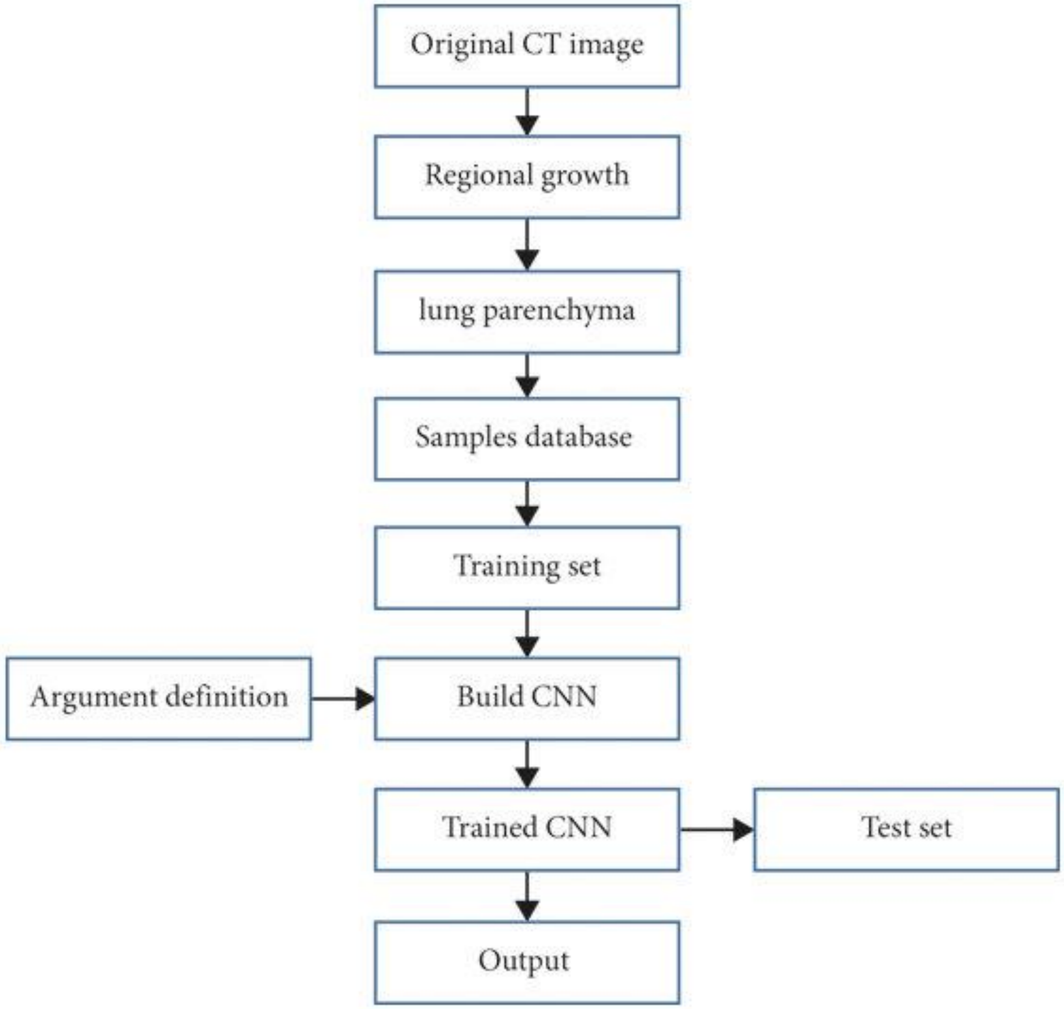

Diagram 4. CNN Flowchart (*Flowchart-for-the-CNN-Algorithm.Jpg (600×572)*, n.d.)

Beyond power output prediction, the digital twin model was pivotal in developing Predictive Maintenance and Fault Detection models, crucial for maximizing operational efficiency and dependability of the wind farm. The Predictive Maintenance model aims to foresee system malfunctions for strategic maintenance planning, reducing downtime and extending equipment lifespan. Conversely, the Fault Detection model rapidly identifies and diagnoses operational anomalies, enabling prompt corrective actions.

3.4. Fault Detection model through ML

The fault detection model plays a crucial role in ensuring the optimal functioning and durability of wind turbines within a wind farm. This model operates by analyzing data and utilizing algorithmic prediction, with the aim of promptly and accurately identifying potential faults in the turbines before they become major issues. The procedure commences with the continuous monitoring of various operational parameters such as wind speed, temperature, output power, rotor speed, and vibration levels. Real-time collection of these data points enables their input into the fault detection algorithm. The algorithm, which relies on sophisticated machine learning techniques, is trained to identify patterns and anomalies that may indicate potential faults. By comparing current operational data with historical trends and known fault signatures, it is capable of detecting any deviations that may suggest a problem. These deviations could manifest as unusual vibrations, signaling wear in bearings, changes in rotor speed, indicating possible blade issues, or inconsistencies in power output, hinting at electrical system malfunctions.Once a potential fault is identified, the system promptly notifies maintenance personnel, providing them with comprehensive information about the nature and location of the issue. This proactive approach to fault detection yields significant benefits. It facilitates timely maintenance interventions, preventing minor issues from escalating into major failures that may necessitate costly repairs, downtime, or even pose safety hazards. Furthermore, by minimizing unscheduled maintenance and extending the operational lifespan of the turbines, the fault detection model greatly enhances the overall efficiency and cost-effectiveness of the wind farm. Consequently, such models are regarded as indispensable tools in the field of renewable energy management, particularly within the rapidly evolving domain of wind energy. The fault detection model constructed for this study is presented below. For the implementation of this algorithm the data used follow the procedure mentioned in the digital twins' model, from the original data the corresponding columns were kept for this algorithm and then the cleaned data with a new file **path_to_cleaned_dataset_B. csv** created were the input data for the code. In addition each step for this algorithm is explained in detail below.

```python
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# Load the dataset
file_path = '/mnt/data/path_to_cleaned_dataset_B.csv'
df = pd.read_csv(file_path, delimiter=';')

# Parse the timestamps and extract features
df['Timestamps'] = pd.to_datetime(df['Timestamps'], format='%d/%m/%Y %H:%M')
df['Hour'] = df['Timestamps'].dt.hour
df['DayOfWeek'] = df['Timestamps'].dt.dayofweek
df['Month'] = df['Timestamps'].dt.month

# Define thresholds for each category
thresholds = {
    'Power': {'Extremely Higher': 3040.5, 'Higher': 1905.6},
    'Pitch': {'Extremely Higher': 10.7, 'Higher': 6.8},
    'GenRPM': {'Extremely Higher': 3275.0, 'Higher': 2445.5},
    'RotorRPM': {'Extremely Higher': 53.1, 'Higher': 39.6},
    'NacelTemp': {'Extremely Higher': 48.0, 'Higher': 36.0},
    'GearOilTemp': {'Extremely Higher': 85.0, 'Higher': 73.0},
    'GearBearTemp': {'Extremely Higher': 99.0, 'Higher': 84.0},
    'GenPh1Temp': {'Extremely Higher': 259.0, 'Higher': 179.5},
    'GenPh2Temp': {'Extremely Higher': 270.0, 'Higher': 189.0},
    'GenPh3Temp': {'Extremely Higher': 247.0, 'Higher': 172.375},
    'GenBearTemp': {'Extremely Higher': 123.0, 'Higher': 93.0}
}

# Function to classify fault status based on thresholds
def classify_fault_status(row):
    for column, threshold_values in thresholds.items():
        if row[column] > threshold_values['Extremely Higher']:
            return "Fault Detected"
        elif row[column] > threshold_values['Higher']:
            return "Imminent Fault"
    return "No Fault"

# Apply the function to each row of the dataframe
df['FaultStatus'] = df.apply(classify_fault_status, axis=1)

# Save the results to a new CSV file
results_file_path = '/mnt/data/modified_cleaned_dataset_B2-results.csv'
df.to_csv(results_file_path, index=False)
```

Figure 11. Fault Detection Algorithm 1/2

```python

# Distribution of Fault Categories
fault_counts = df['FaultStatus'].value_counts()
plt.figure(figsize=(8, 6))
sns.barplot(x=fault_counts.index, y=fault_counts.values)
plt.title('Distribution of Fault Categories')
plt.xlabel('Fault Category')
plt.ylabel('Count')
plt.savefig('/mnt/data/fault_distribution.png')
plt.show()

# Frequency of Faults Over Time (e.g., by Hour)
plt.figure(figsize=(12, 6))
sns.countplot(x='Hour', hue='FaultStatus', data=df)
plt.title('Frequency of Faults by Hour of the Day')
plt.xlabel('Hour')
plt.ylabel('Count')
plt.legend(title='Fault Status')
plt.savefig('/mnt/data/faults_by_hour.png')
plt.show()

# Correlation Heatmap
# Convert categorical data to numerical for correlation
df_numerical = df.copy()
df_numerical['FaultStatus'] = df['FaultStatus'].map({'No Fault': 0, 'Imminent Fault': 1, 'Fault Detected': 2})

# Calculating correlations
correlation_matrix = df_numerical.corr()

# Plotting the heatmap
plt.figure(figsize=(12, 10))
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm')
plt.title('Correlation Heatmap')
plt.savefig('/mnt/data/correlation_heatmap.png')
plt.show()

results_file_path
```

Figure 12. Fault Detection Algorithm 2/2

The algorithm (Figure 11. Fault Detection Algorithm 1/2,Figure 12. Fault Detection Algorithm 2/2) presents a thorough data analysis pipeline using Python, aimed at processing, analyzing, and visualizing a dataset comprehensively. It encompasses steps like data loading, parsing, feature extraction, fault status classification, data saving, and visualization. The process begins with importing libraries, where pandas is used for data manipulation, and matplotlib.pyplot and seaborn for data visualization. The dataset is then loaded into a DataFramedf from a specified path using pandas.read_csv(), with a semicolon as the delimiter. The next phase involves parsing 'Timestamps' into a datetime object and extracting new features such as 'Hour', 'DayOfWeek', and 'Month', which are pivotal for time-based analysis. The script then defines thresholds in a dictionary for various parameters like 'Power', 'Pitch', etc., to classify the fault status. A custom function, classify_fault_status, is created to categorize each row into 'Fault Detected', 'Imminent Fault', or 'No Fault' based on these thresholds, and this classification is added to the DataFrame as a new 'FaultStatus' column. Subsequently, the modified DataFrame is saved as a new CSV file at a specified path. To visualize the distribution of different fault categories, a bar plot is created using seaborn.barplot, and the plot is saved as an image file. Further, the frequency of faults over time is analyzed with another plot, showing fault occurrences by hour using seaborn.countplot, aiding in understanding the variation of fault occurrences throughout the day. Moreover, the DataFrame is converted to a numerical format for correlation analysis. A correlation matrix is computed and visualized using a heatmap, providing insights into the relationships between various variables. All these visualizations, including those for the distribution of fault categories, frequency of faults over time, and the correlation heatmap, are saved as image files. The next diagram (Diagram 5. Fault Detection Flowchart) is the flow chart of the algorithm showing each step of the algorithm.

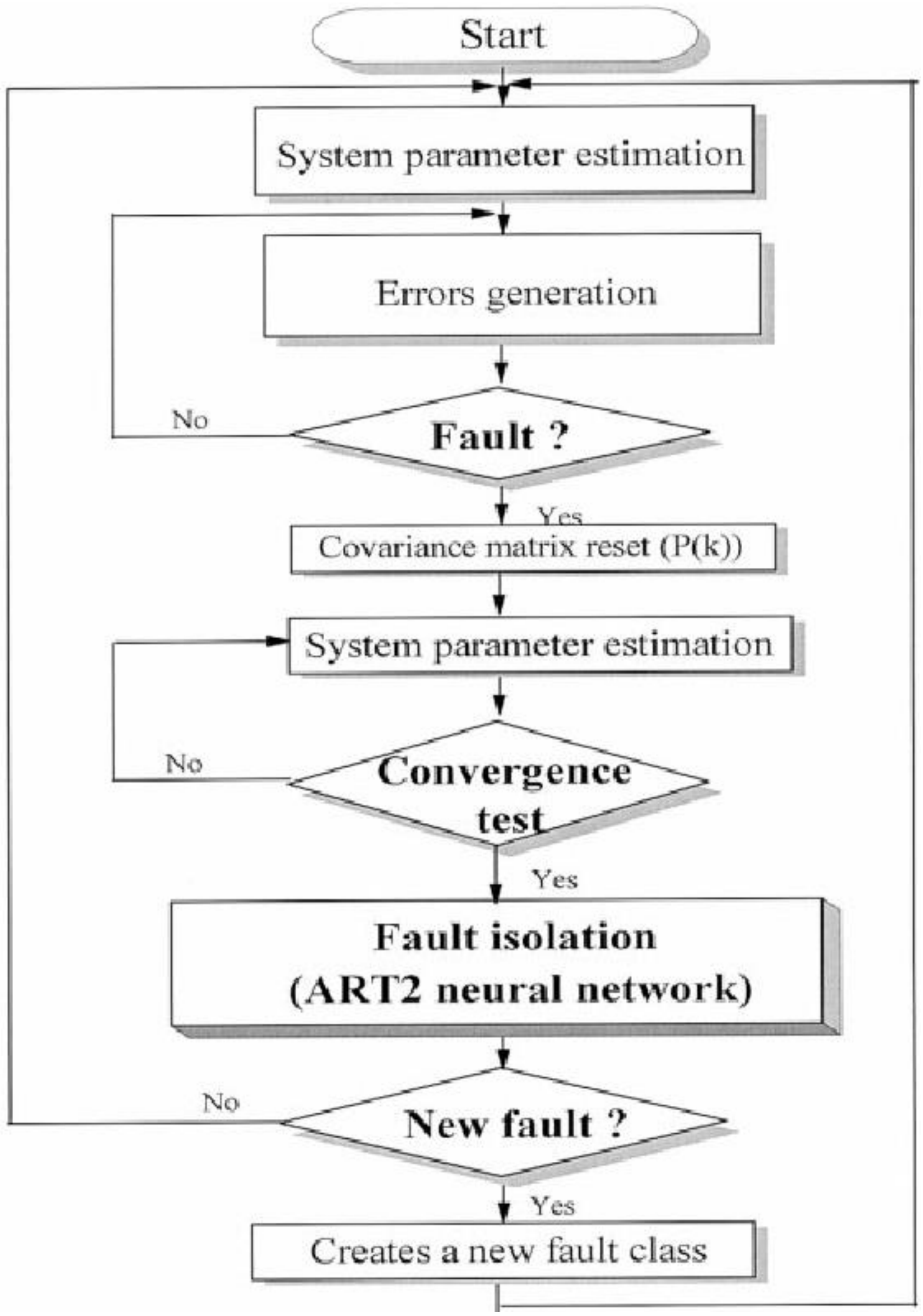

Diagram 5. Fault Detection Flowchart (*Flowchart-of-the-Fault-Diagnosis-Algorithm.Png (560×796)*, n.d.)

For the definition of the thresholds the procedure goes as follows. To establish a comprehensive understanding of outlier categorization in a dataset, specific numerical thresholds are set for five distinct categories: "Extremely Lower", "Lower", "Average Distance", "Higher", and "Extremely Higher". These categories are determined using the Interquartile Range (IQR), a standard statistical method. The categorization is defined as follows:

- **Extremely Lower:** Less than Q1 - 3 * IQR
- **Lower:** Between Q1 - 3 IQR and Q1 - 1.5 IQR
- **Average Distance:** Between Q1 - 1.5 IQR and Q3 + 1.5 IQR
- **Higher:** Between Q3 + 1.5 IQR and Q3 + 3 IQR
- **Extremely Higher:** Greater than Q3 + 3 * IQR These thresholds are calculated for each numerical column in the dataset.

Column	Extremely Lower	Lower	Average Distance	Higher	Extremely Higher
Power	Less than -2255.7	Between -2255.7 and -1120.8	Between -1120.8 and 1905.6	Between 1905.6 and 3040.5	Greater than 3040.5
Pitch	Less than -7.5	Between -7.5 and -3.6	Between -3.6 and 6.8	Between 6.8 and 10.7	Greater than 10.7
GenRPM	Less than -596.0	Between -596.0 and 233.5	Between 233.5 and 2445.5	Between 2445.5 and 3275.0	Greater than 3275.0
RotorRPM	Less than -9.9	Between -9.9 and 3.6	Between 3.6 and 39.6	Between 39.6 and 53.1	Greater than 53.1
NacelTemp	Less than -8.0	Between -8.0 and 4.0	Between 4.0 and 36.0	Between 36.0 and 48.0	Greater than 48.0
GearOilTemp	Less than 29.0	Between 29.0 and 41.0	Between 41.0 and 73.0	Between 73.0 and 85.0	Greater than 85.0

GearBearTemp	Less than 29.0	Between 29.0 and 44.0	Between 44.0 and 84.0	Between 84.0 and 99.0	Greater than 99.0
GenPh1Temp	Less than -112.0	Between -112.0 and -32.5	Between -32.5 and 179.5	Between 179.5 and 259.0	Greater than 259.0
GenPh2Temp	Less than -108.0	Between -108.0 and -27.0	Between -27.0 and 189.0	Between 189.0 and 270.0	Greater than 270.0
GenPh3Temp	Less than -101.25	Between -101.25 and -26.625	Between -26.625 and 172.375	Between 172.375 and 247.0	Greater than 247.0
GenBearTemp	Less than -17.0	Between -17.0 and 13.0	Between 13.0 and 93.0	Between 93.0 and 123.0	Greater than 123.0

These calculated thresholds provide a quantitative framework for categorizing outliers in each column of the dataset. By defining these boundaries, it becomes possible to systematically analyze and understand the extent of deviation in the dataset's numerical values, thus aiding in more informed data analysis and decision-making processes. The thresholds for categorizing outliers are calculated using the Interquartile Range (IQR) method. Here's a step-by-step explanation of how these thresholds are determined for each column:

Step 1: Determine the Quartiles

- **First Quartile (Q1):** This is the median of the lower half of the dataset (excluding the median if there is an odd number of data points). It marks the 25th percentile of the data.
- **Third Quartile (Q3):** This is the median of the upper half of the dataset. It marks the 75th percentile of the data.

Step 2: Calculate the Interquartile Range (IQR)

IQR: It is the difference between the Third (Q3) and the First Quartile (Q1).

$$IQR = Q3 - Q1$$

Step 3: Calculate Thresholds for Outlier Categories

- **Extremely Lower Threshold:** Values less than this are judged "Extremely Lower" outliers.

$$Extremely\ Lower\ =\ Q1\ -\ 3\ \times\ IQR$$

- **Lower Threshold:** Values between this threshold and the "Extremely Lower" threshold are considered "Lower" outliers.

$$Lower\ = Q1\ -\ 1.5\ \times\ IQ$$

- **Average Distance Lower and Higher Threshold:** These thresholds define the typical range of the data without outliers.

$$Average\ Distance\ Lower = Q1 - 1.5\ \times\ IQR$$

$$Average\ Distance\ Higher\ =\ Q3\ +\ 1.5\ \times\ IQR$$

- **Higher Threshold:** Values between this threshold and the "Average Distance Higher" threshold are considered "Higher" outliers.

$$Higher = Q3 +\ 1.5\ \times\ IQR$$

- **Extremely Higher Threshold:** Values greater than this are considered "Extremely Higher" outliers.

$$Extremely\ Higher\ =\ Q3\ +\ 3\ \times\ IQR$$

3.5. Predictive Maintenance Models

A predictive maintenance model in a wind farm plays a pivotal role in optimizing the maintenance schedule and enhancing the overall efficiency of the turbines. This model operates by analyzing a wealth of data collected from the wind turbines, including key parameters like wind speed, temperature, output power, rotor speed, and vibration levels. This data is continuously monitored and analyzed to predict when maintenance should be performed before a fault occurs. The procedure involves using advanced machine learning algorithms that learn from historical data to identify patterns and signs of wear or impending failure in turbine components. These algorithms can predict potential issues by detecting subtle changes in the data that may indicate deterioration or an anomaly in the turbine's performance. For instance, a gradual increase in vibration levels might suggest bearing wear, while changes in power output could indicate issues with the generator or blades. By implementing predictive maintenance, wind farm operators can schedule maintenance activities more effectively, targeting specific turbines at the optimal time before faults develop. This proactive approach significantly reduces the likelihood of unexpected turbine failures and downtime, which can be costly and disruptive. Moreover, predictive maintenance extends the lifespan of turbine components by ensuring they are serviced or replaced at the right time, leading to cost savings in the long run. Overall, the predictive maintenance model is a valuable asset for wind farms. It enhances operational reliability, reduces maintenance costs, and ensures a more consistent energy output. In an industry where maximizing efficiency and minimizing downtime is crucial, predictive maintenance models offer a smart, data-driven solution to maintain turbines at peak performance. The predictive model in this study is presented in the code below (Figure 13. Predictive Maintenance Algorithm).

```python
import pandas as pd
from sklearn.cluster import KMeans
import numpy as np

# Re-loading the dataset
file_path = '/mnt/data/modified_cleaned_dataset_B2-results.csv'
df = pd.read_csv(file_path)

# Continuing with the previously engineered features
df['FaultDetectedCount'] = np.random.randint(0, 10, size=len(df))
df['ImminentFaultCount'] = np.random.randint(0, 10, size=len(df))

# K-Means Clustering with increased clusters
kmeans = KMeans(n_clusters=5)  # Increasing clusters to 5 for more granularity
df['Cluster'] = kmeans.fit_predict(df[['FaultDetectedCount', 'ImminentFaultCount']])

# Analyzing Clusters for Maintenance Indication
# Assuming higher counts in either 'FaultDetectedCount' or 'ImminentFaultCount' indicate maintenance need
maintenance_threshold = max(df['FaultDetectedCount'].mean(), df['ImminentFaultCount'].mean())
df['MaintenanceNeeded'] = ((df['FaultDetectedCount'] >= maintenance_threshold) |
                           (df['ImminentFaultCount'] >= maintenance_threshold))

# Count of data points per cluster and maintenance need
cluster_counts = df['Cluster'].value_counts()
maintenance_counts = df['MaintenanceNeeded'].value_counts()

cluster_counts, maintenance_counts, df.head()
```

Figure 13. Predictive Maintenance Algorithm

The next diagram (Diagram 6. Predictive Maintenance Flowchart) is the flow chart of the algorithm showing each step of the algorithm.

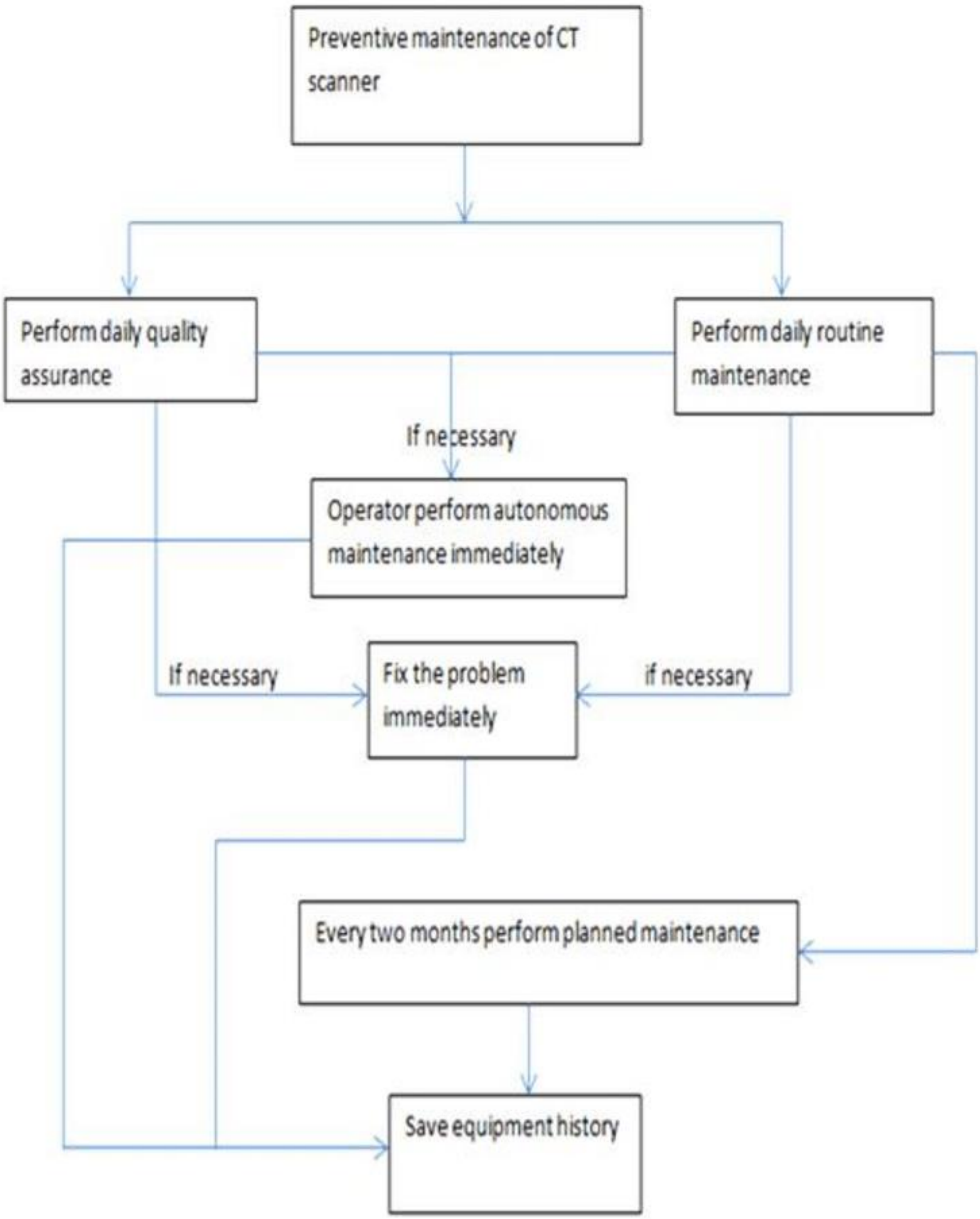

Diagram 6. Predictive Maintenance Flowchart (*Proposed-Preventive-Maintenance-Flow-Chart.Png (768×965)*, n.d.)

4. RESULTS

In this concluding chapter, the culmination of the thesis is achieved through an elaborate exposition of the findings, centering on the efficacy of the digital twin model. This sophisticated model incorporates four distinct machine learning algorithms: Decision Trees, Random Forest, Artificial Neural Networks (ANN), and Convolutional Neural Networks (CNN). The chapter meticulously presents the outcomes of each model, employing graphical representations, where applicable, to visually depict the data. To ascertain the most optimal approach for the specific application, a comprehensive evaluation of each model was undertaken. This evaluation was grounded in essential performance metrics: R-squared (R^2), mean squared error (MSE), and root mean squared error (RMSE). The chapter transcends these four models, delving into the results for Fault Detection and Predictive Maintenance models as well. In each instance, the findings are not only quantitatively scrutinized but also evaluated critically to comprehend their practical implications and relevance within the broader context of the field. This meticulous analysis provides a lucid perspective on the strengths and capabilities of each model, establishing a robust basis for both theoretical comprehension and practical implementation in the domain of digital twin technology. For the Digital Twin model the results for each Machine Learning models are as follows.

Decision Trees Model:

Mean Squared Error (MSE): 1090.2385682436002

Root Mean Squared Error (RMSE): 33.01876085263649

R-squared (R^2) Score: 0.9775002593283868

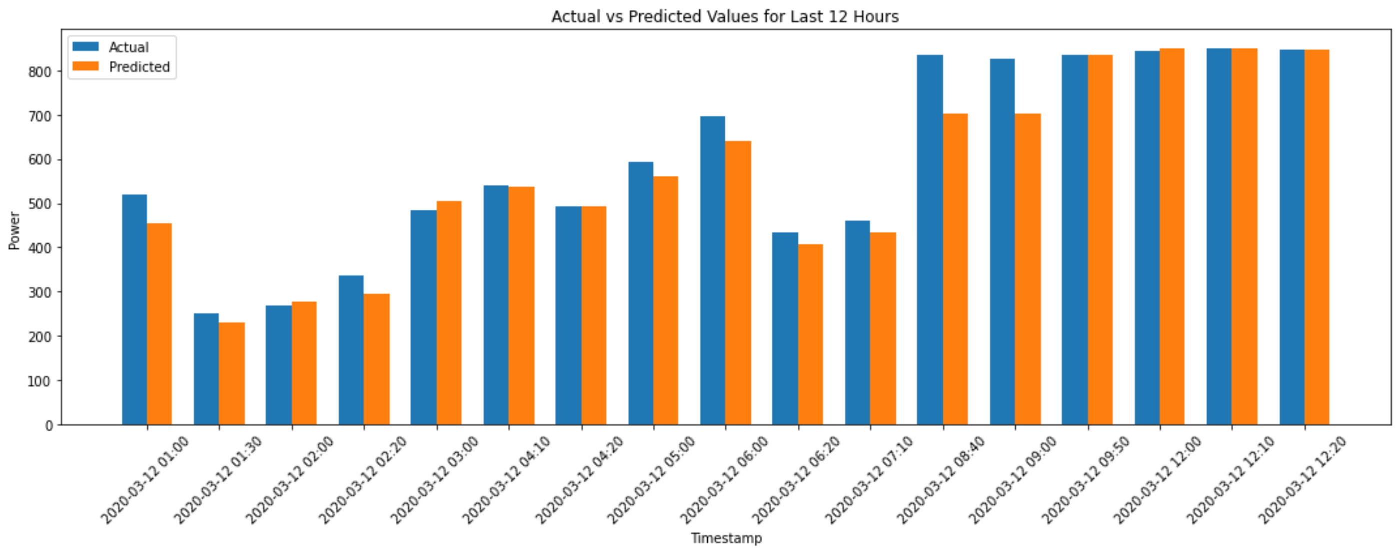

Chart 1. Decision Trees Results

This chart (Chart 1. Decision Trees Results)"Actual vs Predicted Values for Last 12 Hours" presents a side-by-side comparison of actual and predicted data over a 12-hour period. The X-axis displays time stamps in a 'yyyy-mm-dd hh:mm' format, marking consecutive hours. The Y-axis quantifies the power in kW. Each time stamp has two vertical bars — the blue bar represents the actual power values observed, and the orange bar indicates the predicted power values obtained from a model. The bars are arranged in pairs for each time stamp, allowing for a clear visual comparison between actual and forecasted power usage for each hour. The pattern of the bars suggests a close correlation between the predicted and actual values, with some discrepancies.

Random Forest Results:

Mean Squared Error: 652.1245282441845

Root Mean Squared Error: 25.536729004400396

R-squared (R^2) Score: 0.9865418146096866

Training MSE: 103.22065647075526

Testing MSE: 652.1245282441845

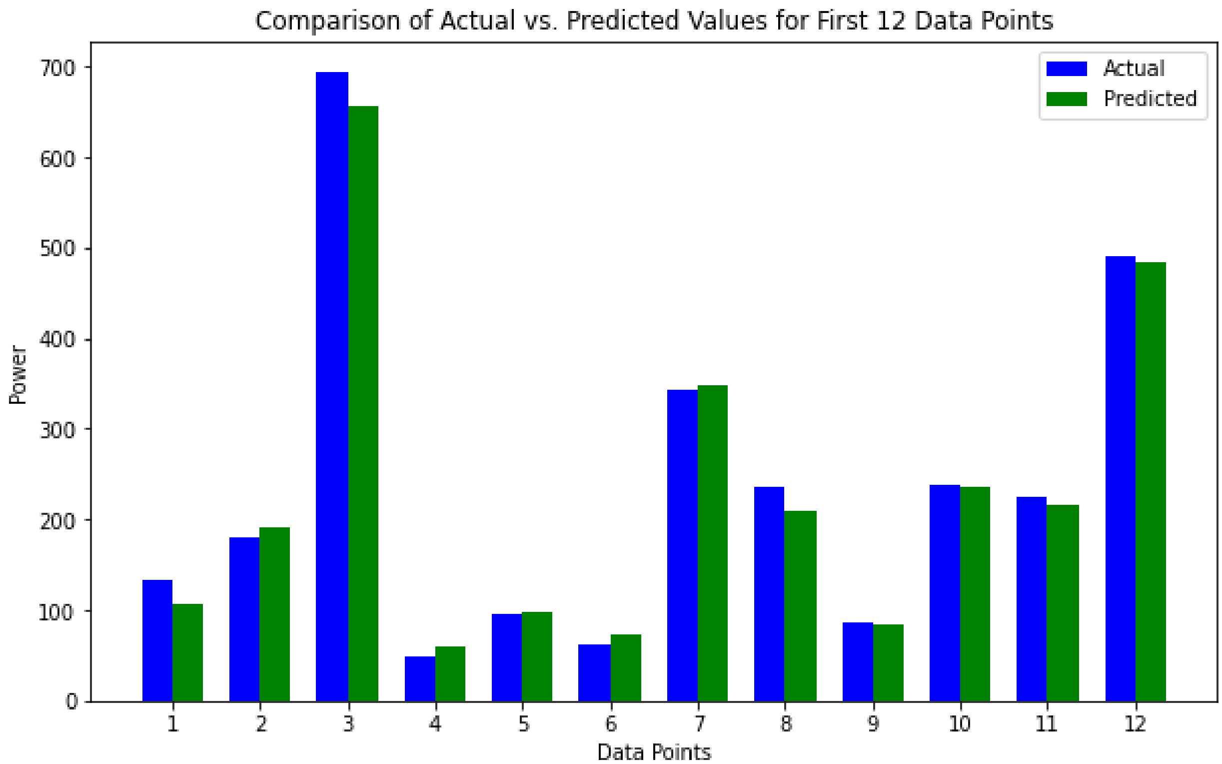

Chart 2. Actual Predictions of Random Forest

The provided (Chart 2. Actual Predictions of Random Forest) diagram is a visual representation comparing actual power output values to those predicted by a Random Forest regression model for a set of data points. This type of visualization is commonly used to assess the performance of regression models in machine learning. It contains two sets of bars for each data point: one for the actual value and one for the predicted value. These bars are color-coded—actual values in green and predicted values in blue—to distinguish between them easily. The model's predictions vary in accuracy across the dataset. At certain points, such as data points 2 and 12, the predictions significantly underestimate the actual values. Conversely, at data point 4, the model overestimates the actual value. The accuracy of the model's predictions can be assessed by comparing the height of the bars: the closer the bars are in height for a given data point, the more accurate the prediction. The vertical axis represents the power output and the horizontal axis lists the data points from 1 to 12. The title "Comparison of Actual vs. Predicted Values for First 12 Data Points" suggests that these are the initial results from a larger set of predictions made by the model, implying that the diagram is part of a larger evaluation of the model's predictive performance on a dataset related to power output, possibly from a renewable energy source like wind or solar power.

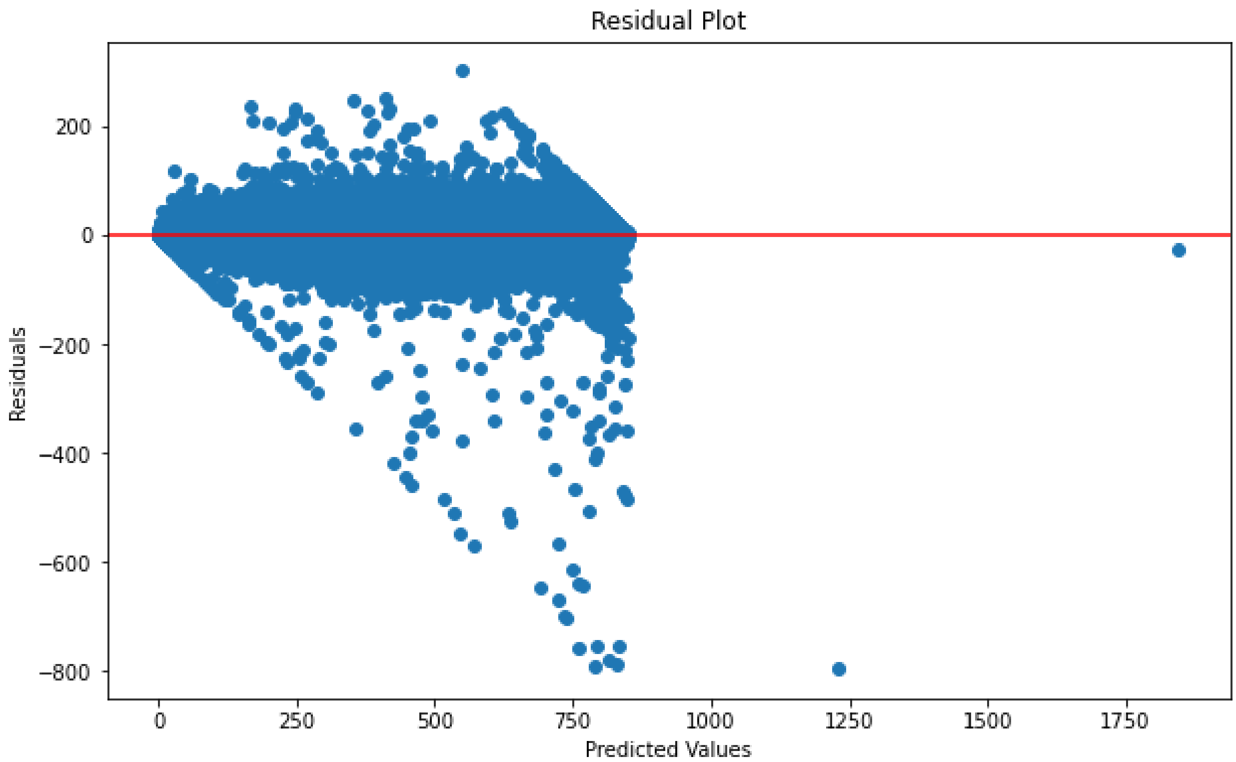

Chart 3. Residual Plot of Random Forest

This diagram (Chart 3. Residual Plot of Random Forest) is a residual plot, a type of scatter plot that is used to visualize the residuals (differences between observed and predicted values) of a predictive model. The horizontal axis represents the predicted values, while the vertical axis represents the residuals. In this residual plot, each point corresponds to a single prediction made by the model. The location of a point along the vertical axis shows the residual for that prediction: points that lie on the horizontal line at zero indicate perfect predictions, while points above or below the line indicate that the model has underpredicted or overpredicted, respectively. The distribution of points in this plot is heteroscedastic, as the spread of residuals appears to increase with the increase in predicted values. This suggests that the model's predictive errors vary across the range of predictions. A model is ideally homoscedastic, where the residuals are evenly distributed across all levels of predicted values, indicating consistent accuracy of the model regardless of the size of the prediction. The red line at the zero mark on the vertical axis serves as a reference to quickly gauge whether the residuals are above or below the expected value (a perfect prediction). The fact that there are many points far from this line suggests there are many predictions that are not very close to the actual values. This kind of plot is instrumental in regression analysis to diagnose issues with the model, such as whether the errors are systematically high or low across the range of predictions.

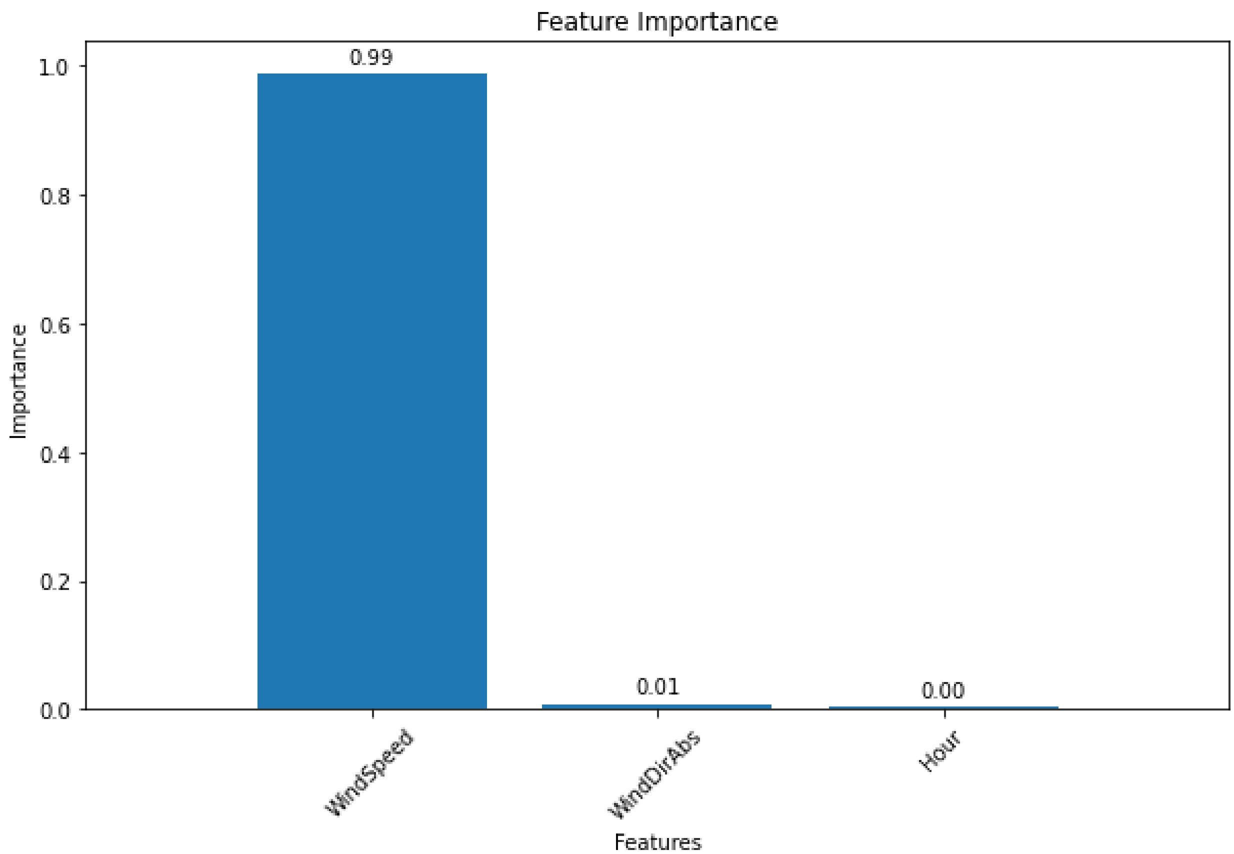

Chart 4. Feature Importance of Random Forest

This diagram (Chart 4. Feature Importance of Random Forest) is a bar chart representing the feature importance as determined by a Random Forest algorithm. The chart lists three features along the horizontal axis: 'WindSpeed', 'WindDirAbs', and 'Hour'. The vertical axis quantifies the importance of each feature in the model, likely on a scale from 0 to 1, where 1 would indicate the highest possible importance. The feature 'WindSpeed' has a bar extending almost to the top of the chart, labeled with a numerical value of 0.99, suggesting that it is by far the most important feature in predicting the target variable in the model. In stark contrast, 'WindDirAbs' has a very small bar, with a value of 0.01, indicating that it has some, but very little importance. The 'Hour' feature has a bar that is not visible above the axis, and its label indicates an importance of 0.00, which suggests that it has no predictive power according to the model used. This kind of visualization is crucial for understanding which variables the model is primarily relying on to make predictions. In this case, the model seems to be heavily dependent on 'WindSpeed', which might be expected in a model predicting something like energy output from a wind turbine. However, the time of day ('Hour') appears to have no bearing on the model's predictions, which could be an insight into the nature of the data or suggest that time of day was not a relevant predictor in this specific context.

ANN Results:

Mean Squared Error: 603.0259533432173

Root Mean Squared Error: 24.5565867608513

R-squared (R^2) Score: 0.9875550838470767

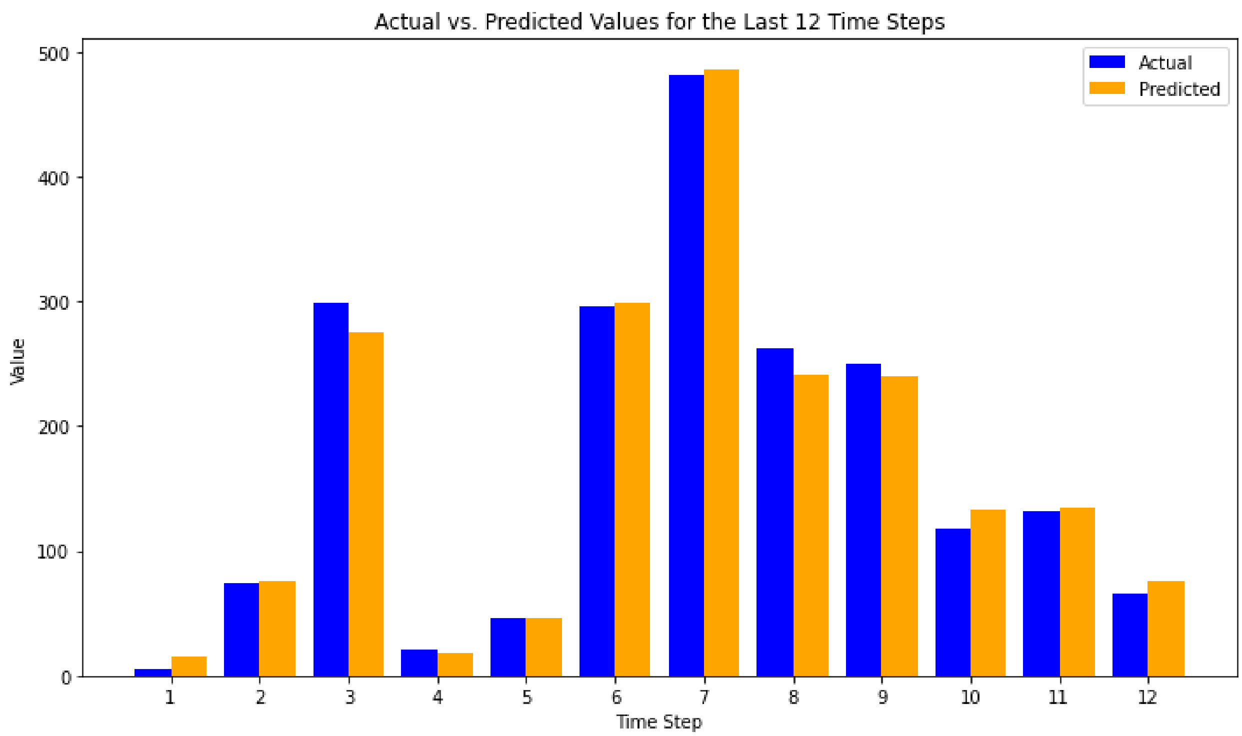

Chart 5. Actual Predictions of ANN

The chart (Chart 5. Actual Predictions of ANN) displayed is a bar graph comparing actual and predicted values over a series of 12 time steps, which are the result of an Artificial Neural Network (ANN) model as suggested by the file name. Each time step is represented by a pair of bars on the horizontal axis, labeled from 1 to 12. The vertical axis represents the value of the output variable, though the specific nature of the value (e.g., sales, temperatures, etc.) and its units are not provided. For each time step, there are two bars: one for the actual value (colored in blue) and one for the predicted value (colored in orange). These colors make it easy to distinguish between the actual and predicted values and assess the accuracy of the ANN model's predictions at each time step. The bar heights represent the magnitude of the values. By comparing the height of the blue and orange bars at each time step, one can evaluate how well the model's predictions align with the actual data. For example, at time steps 5 and 6, the model's predictions closely match the actual values, indicated by the nearly equal height of the bars. However, at other time steps, such as 2 and 4, there are noticeable differences between the predicted and actual values, suggesting discrepancies in the model's performance. Overall, this kind of visualization is commonly used to illustrate the performance of predictive models, highlighting areas where the model performs well or where it may need improvement.

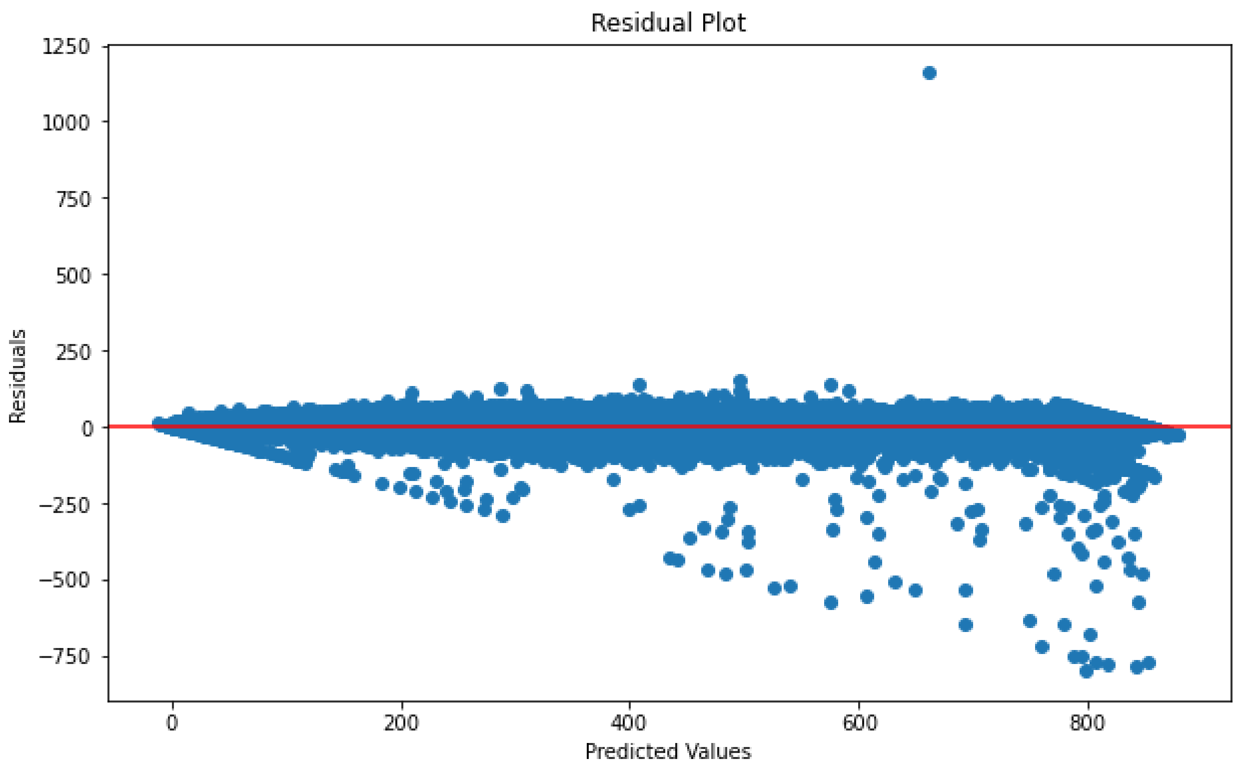

Chart 6.Residual Plot of ANN

This diagram (Chart 6.Residual Plot of ANN) is another residual plot, which is utilized to assess the performance of a predictive model, in this case, an Artificial Neural Network (ANN). The horizontal axis shows the predicted values while the vertical axis shows the residuals—the difference between the actual values and the predicted values. Ideally, if the predictions were perfect, all points would lie on the horizontal line where the residual is zero, indicated by the red line. In this plot, most residuals cluster around this line, suggesting that the predictions are generally close to the actual values. However, there are some notable outliers, particularly one that is significantly above the rest, which could indicate an instance where the model's prediction was far from the actual value. The presence of such outliers might suggest the model has limitations in capturing all the patterns in the data or that there are anomalies in the dataset.

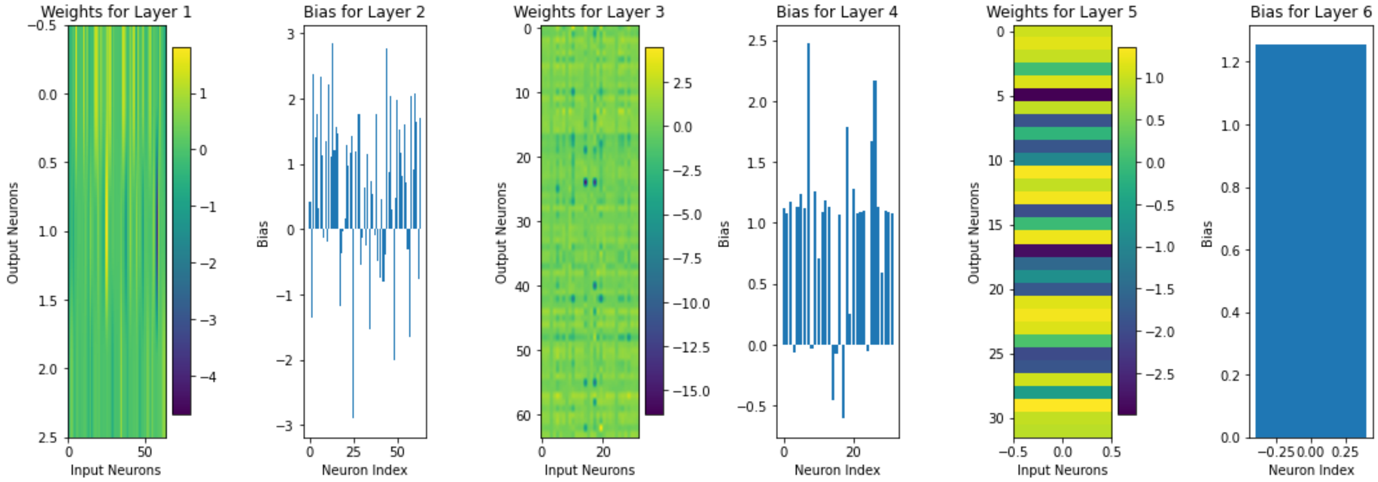

Chart 7.Weights for ANN

This diagram (Chart 7.Weights for ANN) is a series of plots illustrating the weights and biases of different layers within an ANN. These visualizations are crucial for understanding how the ANN model processes input data:

- The first plot shows the weights for Layer 1. Each column represents a neuron in the input layer, and each row represents a neuron in the subsequent layer. The color intensity indicates the strength of the weight (positive or negative).
- The second plot presents the biases for Layer 2. Each line represents a neuron in Layer 2, and the length of the line indicates the size of the bias.
- The third plot shows the weights for Layer 3, with a similar structure to the first plot, albeit for a different layer.
- The fourth plot shows the biases for Layer 4, again, each line representing the bias for a neuron in that layer.
- The fifth plot shows the weights for Layer 5, which, like the other weight plots, indicates how input neurons affect the output neurons.
- The final plot shows the bias for Layer 6, which is a singular value given that it seems to be the output layer or a layer with a single neuron.

These weights and biases are the parameters the ANN adjusts during training to minimize prediction error. They are the essence of the neural network's learning process. The complexity and patterns within these plots reflect the complexity of the relationships that the ANN has learned from the training data.

CNN Results:

Mean Squared Error: 5932.643032918852

Root Mean Squared Error: 77.02365242520541

R-squared (R^2) Score: 0.877565393826627

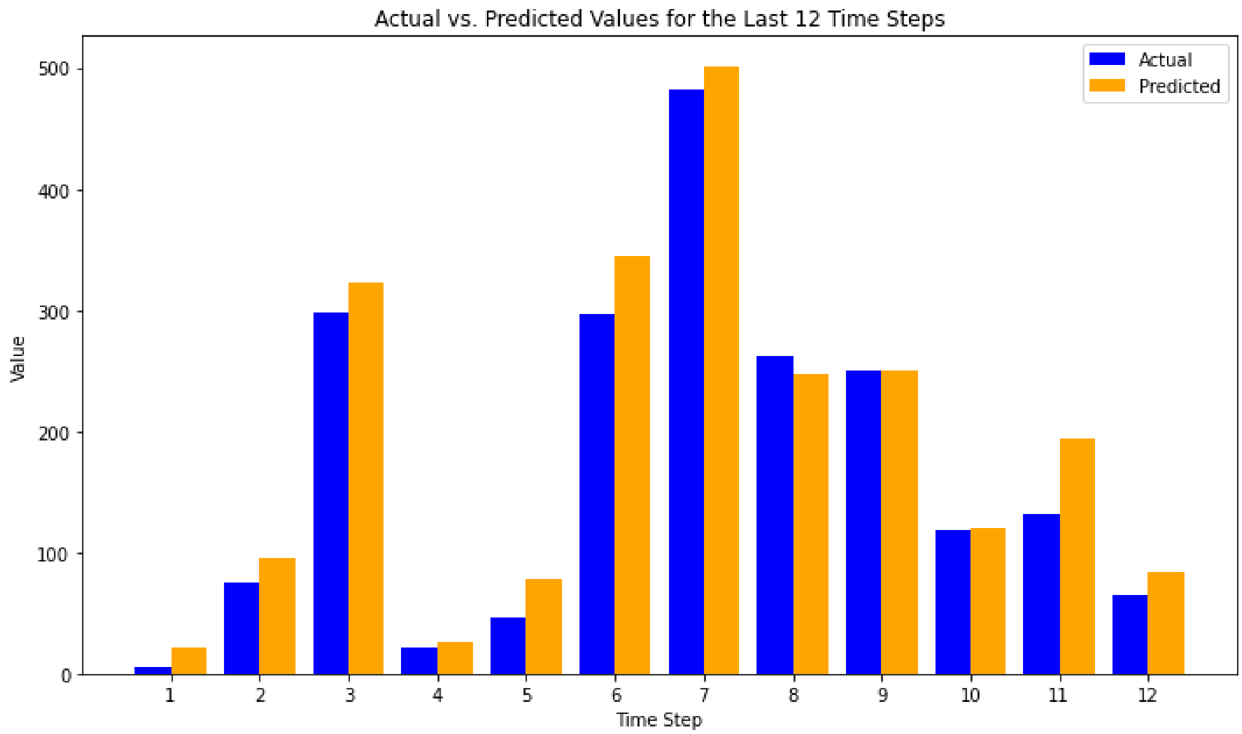

Chart 8.Actual Predictions of CNN

This chart (Chart 8.Actual Predictions of CNN) depicting the actual versus predicted values for the last 12 time steps from a Convolutional Neural Network (CNN) model. The horizontal axis is labeled with the time steps from 1 to 12, and the vertical axis represents the value of the output variable, which is not specified. Each time step has two bars adjacent to each other: the blue bar represents the actual value, and the orange bar represents the predicted value from the CNN model. The chart allows for a direct visual comparison of the predicted values against the actual ones at each time step, highlighting the model's prediction accuracy. For instance, at time steps 5 and 6, the predicted values are almost identical to the actual values, indicating a high level of accuracy. In contrast, at other time steps, such as 2, 3, and 4, the model's predictions deviate from the actual values, which could indicate periods or conditions where the model is less accurate.

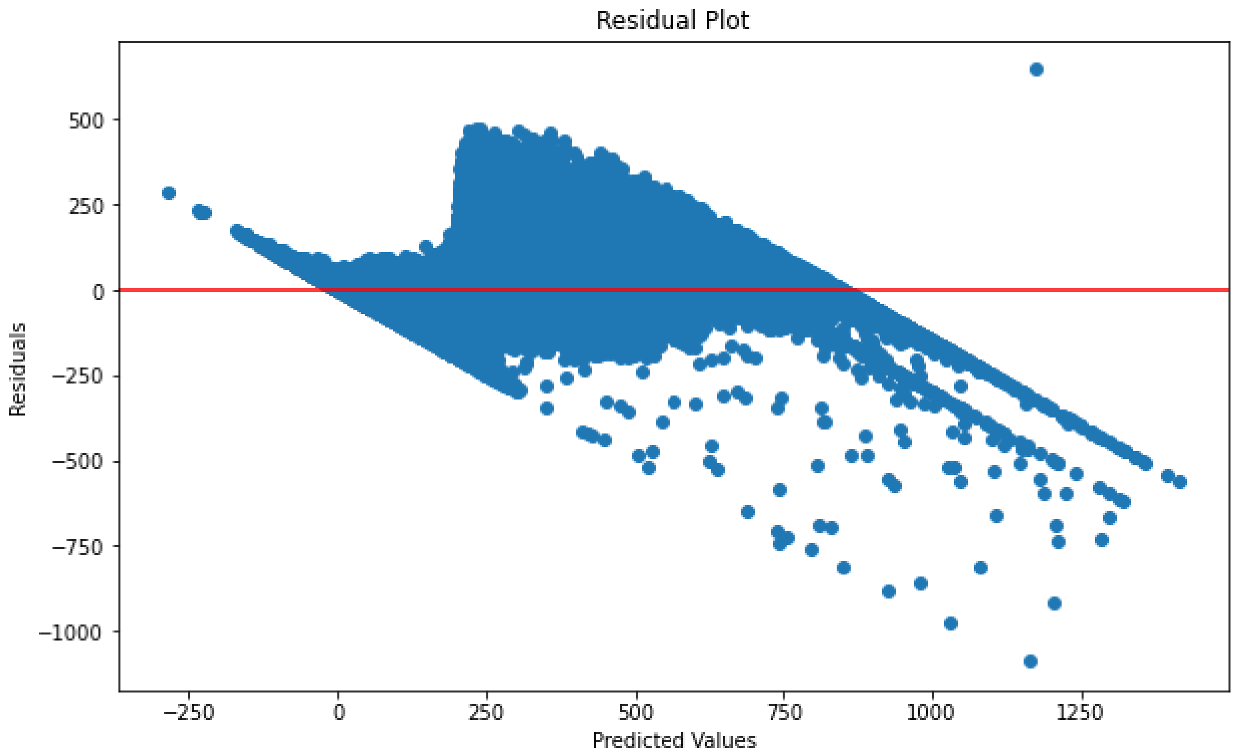

Chart 9.Residual Plot of CNN

This diagram (Chart 9.Residual Plot of CNN) shows a residual plot for a CNN model. In this scatter plot, the horizontal axis displays the predicted values, and the vertical axis shows the residuals or the differences between the predicted and actual values. A red line at zero on the vertical axis indicates where the residuals would be if the predictions were perfect. The plot shows a pattern where the residuals increase negatively with higher predicted values, suggesting that the model tends to overestimate the lower values and underestimate the higher values. This heteroscedastic pattern—where the spread of residuals changes across the range of predictions—might indicate that the model's performance is not consistent across all values it predicts, which could be due to various factors including model architecture or the nature of the data.

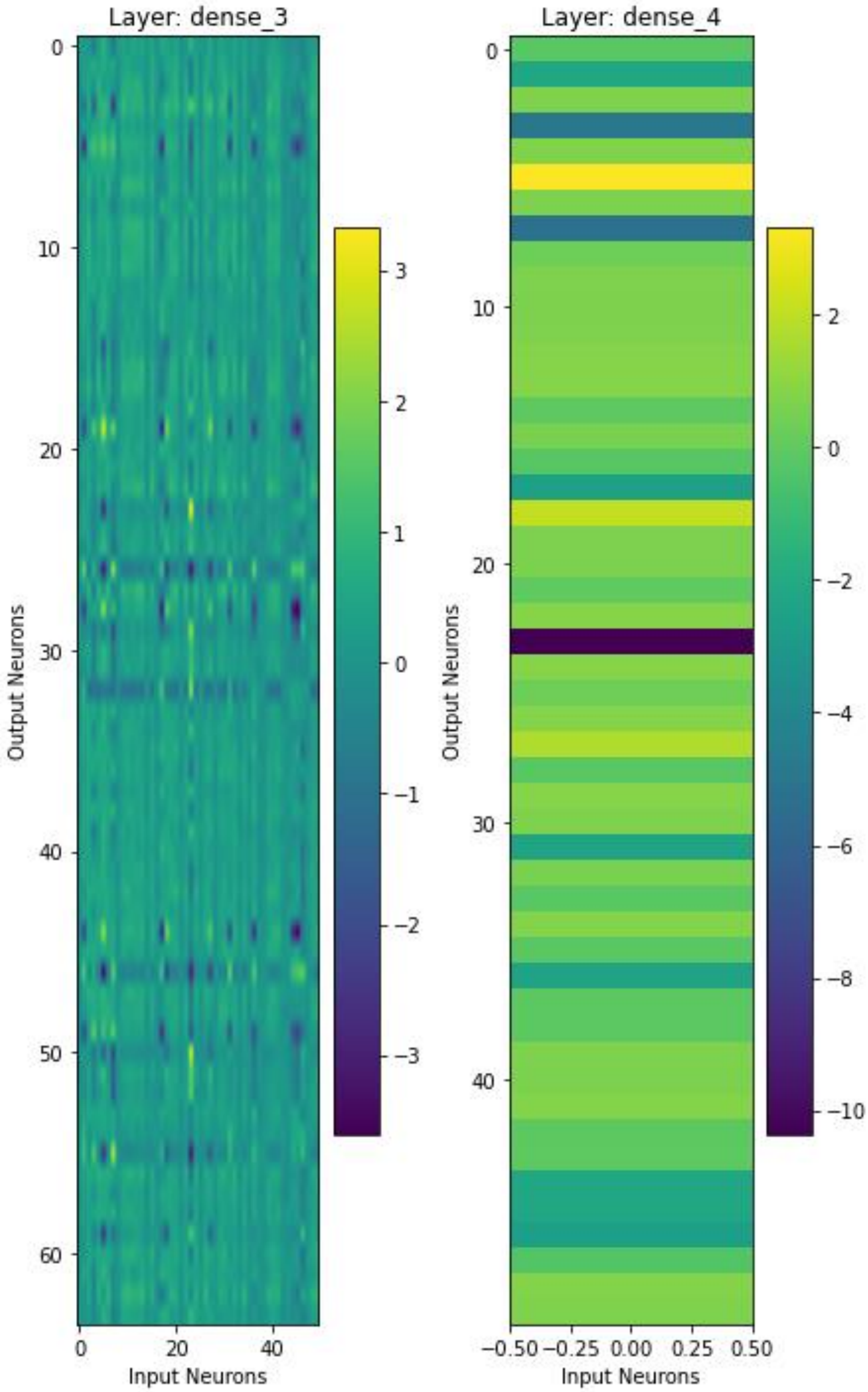

Chart 10.Weights for CNN

The above heatmap (Chart 10.Weights for CNN) representing the weights in two layers of a CNN, named 'dense_3' and 'dense_4'. Heatmaps like this are used to visualize the strength and structure of the connections between neurons in different layers of a neural network. In the heatmap, the x-axis corresponds to input neurons and the y-axis to output neurons. The color intensity indicates the strength of the weight, with cooler colors (e.g., purples and blues) typically representing smaller or negative weights, and warmer colors (e.g., yellows) representing larger or positive weights. Such visualizations can provide insights into which features are being emphasized by the network in these layers and can be useful for diagnosing the behavior of the network, understanding feature representations, and potentially informing adjustments to the network architecture for improved performance.

<u>Fault Detection Results:</u>

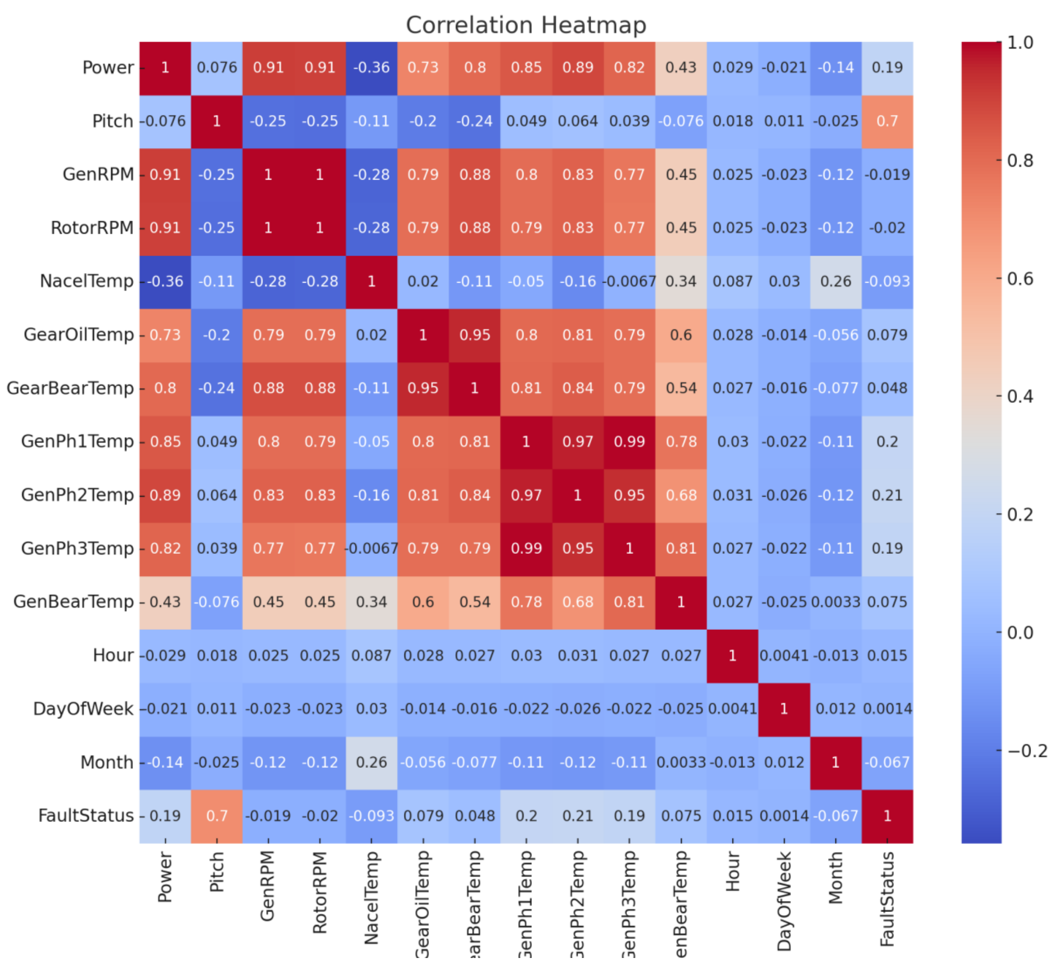

Chart 11. Fault Detection Correlations Heatmap

This heatmap (Chart 11. Fault Detection Correlations Heatmap) is a visual representation of the correlation matrix for various parameters in a fault detection dataset. The X and Y axes of the heatmap list the same parameters, such as Power, Pitch, Generator RPM (GenRPM), Rotor RPM, Nacelle Temperature (NacelleTemp), and several temperature readings from different generator parts (GearOilTemp, GearBearTemp, GenPhTemp1, etc.), as well as time-related factors (Hour, DayOfWeek, Month) and FaultStatus. Each cell within the heatmap shows the correlation coefficient between the parameters on the corresponding X and Y axes, with values ranging from -1 to 1. A value of 1 indicates a perfect positive correlation, -1 indicates a perfect negative correlation, and a value close to 0 implies no correlation. The colors provide a quick visual cue of the strength and direction of the correlation: red tones suggest a positive correlation, blue tones indicate a negative correlation, and lighter colors or white suggest little to no correlation. A closer look reveals strong positive correlations within generator phase temperatures (GenPhTemp1, GenPhTemp2, GenPhTemp3) as indicated by the deep red colors. Similarly, a strong negative correlation is observed between GearOilTemp and Power, as seen by the deep blue shade. Time factors like Hour, DayOfWeek, and Month show very light colors, implying they have negligible linear relationships with the other parameters. FaultStatus shows varying degrees of correlation with other parameters, notably a moderate negative correlation with Power, which could be of particular interest in predicting or detecting faults. This heatmap serves as a critical tool for understanding the relationships between variables in the dataset, which can significantly inform the feature selection process for predictive modeling in fault detection.

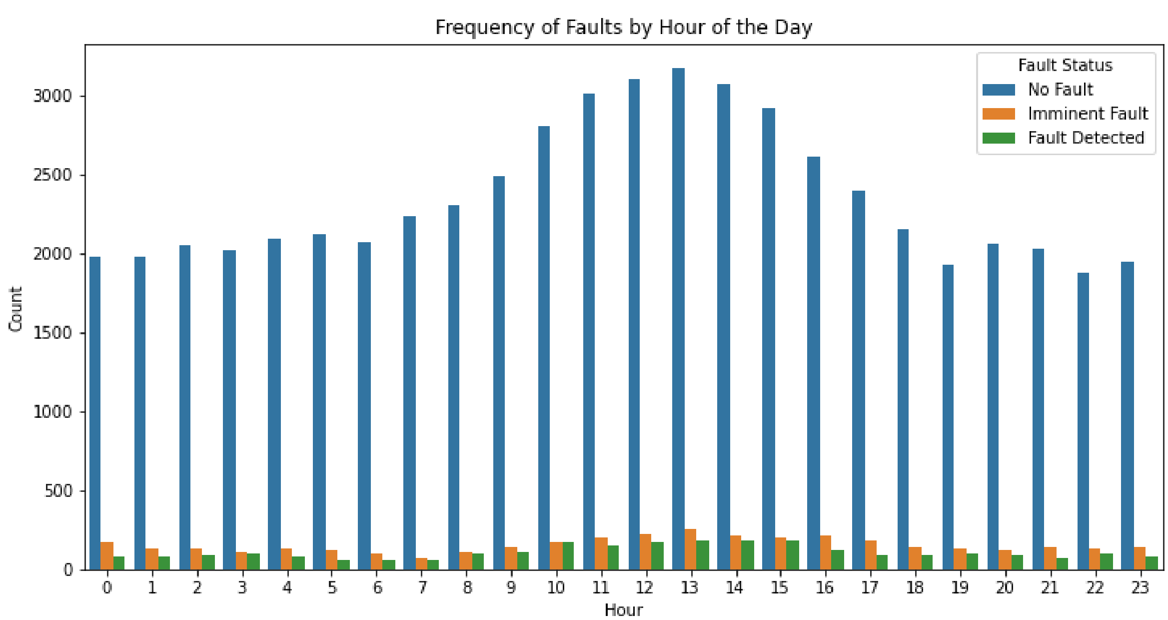

Chart 12. Fault Detection Frequency of Faults

The diagram (Chart 12. Fault Detection Frequency of Faults)"Frequency of Faults by Hour of the Day" presents a histogram that classifies occurrences of faults within a system throughout a 24-hour period. The X-axis represents the hours of the day, ranging from 0 to 23, corresponding to a 24-hour time format. The Y-axis indicates the number of faults detected. The histogram is divided into three categories as indicated by the legend: 'No Fault', 'Imminent Fault', and 'Fault Detected'. Each hour displays the total counts of each fault status type, however, the diagram primarily showcases the occurrences of 'No Fault', implying a dataset with a prevalent number of normal operational readings. There seems to be a recurring pattern, with the counts of 'No Fault' reaching their peak at specific hours, which may suggest periods of heightened system activity or monitoring frequency. The 'Imminent Fault' and 'Fault Detected' statuses exhibit significantly lower counts, with 'Imminent Fault' displaying a very sparse distribution across the hours, and 'Fault Detected' occurrences being slightly more frequent but still considerably less than instances of 'No Fault'. The data indicates that the monitored system experiences faults infrequently compared to normal operation, with certain hours displaying a slight increase in fault detection. This pattern could be advantageous for establishing predictive maintenance schedules or for conducting further examination into the reasons why particular hours may have higher incidences of faults, which could be attributed to factors such as increased load or operational stress during specific times of the day.

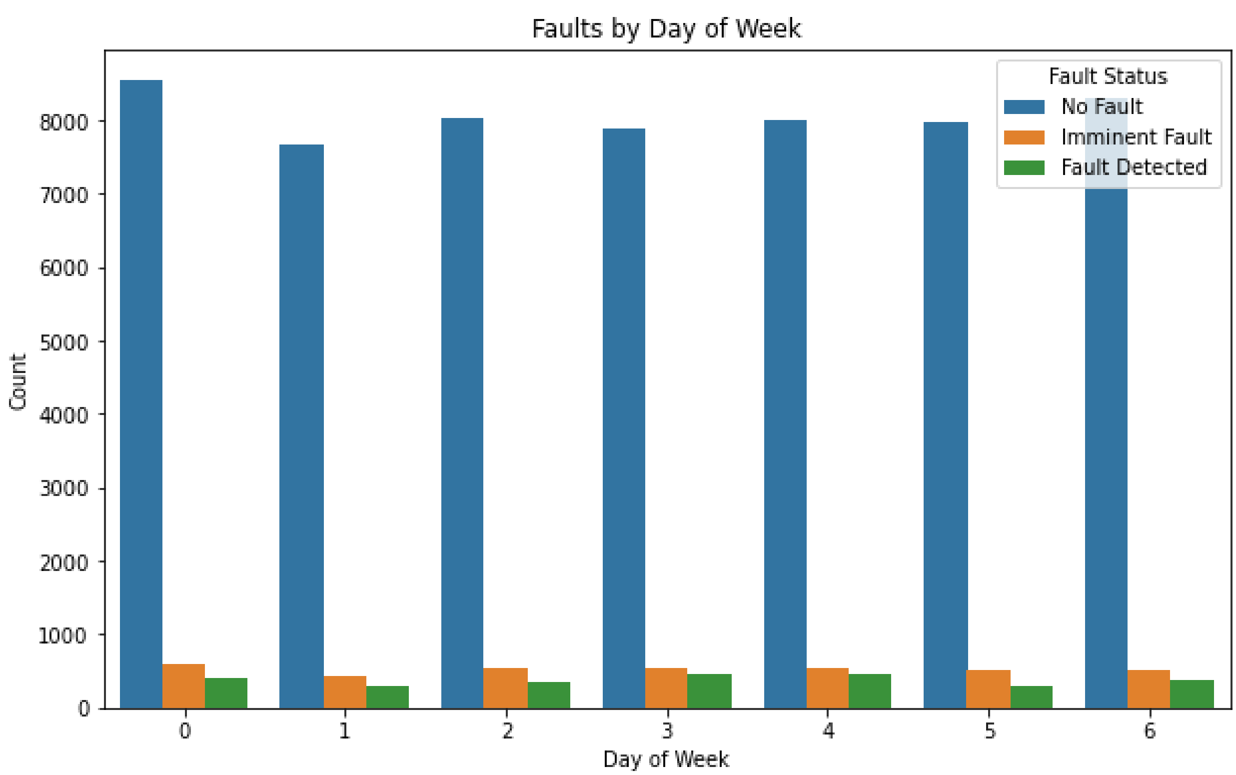

Chart 13.Day of Week Faults

This diagram (Chart 13.Day of Week Faults) is a stacked bar chart that represents the count of fault statuses for each day of the week. The horizontal axis is labeled "Day of Week" with numerical values ranging from 0 to 6, which presumably correspond to the days of the week, starting from Sunday (0) to Saturday (6). The vertical axis is labeled "Count," indicating the number of occurrences. There are three categories represented in each bar, differentiated by color:

- "No Fault" (shown in blue) represents the count of cases where no fault was detected.
- "Imminent Fault" (shown in green) represents the count of cases where a fault was about to occur.
- "Fault Detected" (shown in orange) represents the count of cases where a fault was identified.

The chart shows that for each day, the majority of cases have no faults. The counts of "No Fault" are significantly higher than the counts for "Imminent Fault" and "Fault Detected," which are quite small in comparison. The small sections of green and orange at the bottom of each bar indicate that faults, whether imminent or detected, occur much less frequently than non-fault situations. The similarity in heights of the bars across different days might suggest that there is no significant variation in fault occurrence by day of the week.

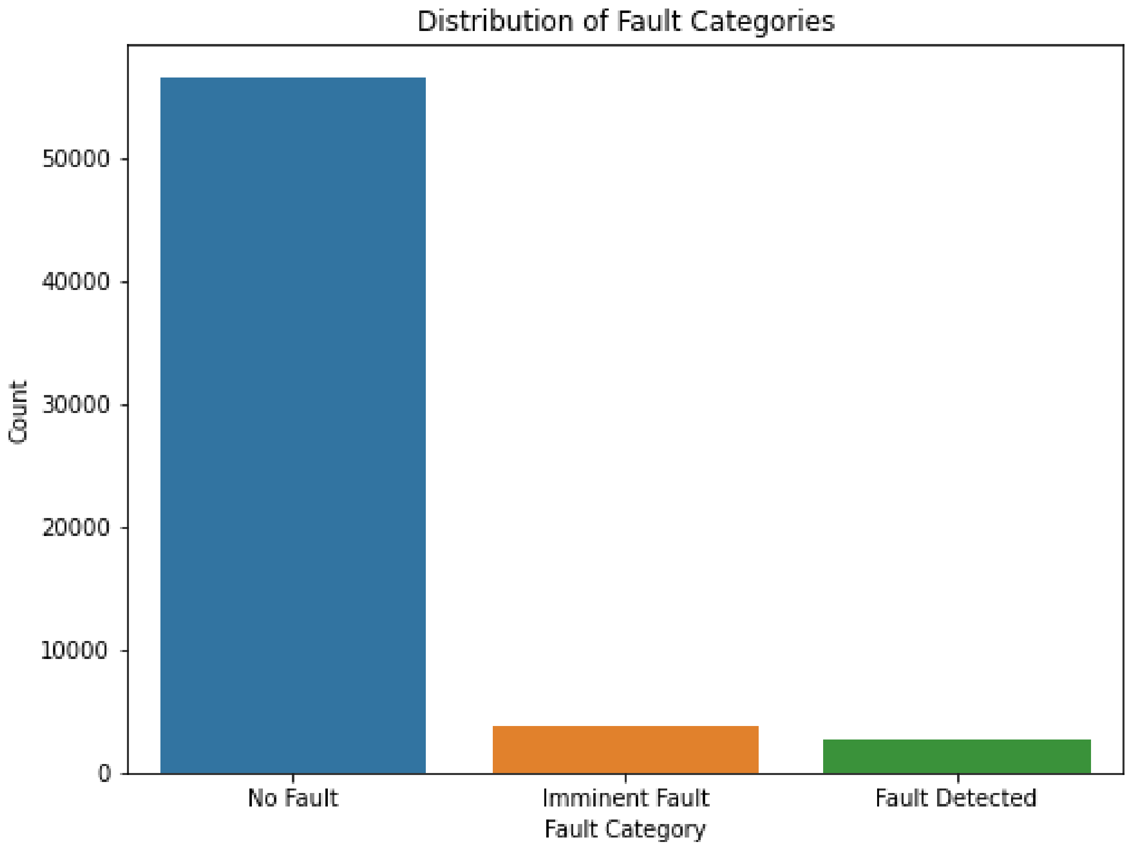

Chart 14.Fault Distribution

This diagram (Chart 14.Fault Distribution), titled "Distribution of Fault Categories," is a bar chart showing the count of instances across three fault categories: "No Fault," "Imminent Fault," and "Fault Detected." The height of the bars indicates the frequency of each category. This visualization is critical for understanding the overall distribution of fault statuses within the dataset, highlighting that "No Fault" instances occur most frequently, with "Imminent Fault" and "Fault Detected" occurring less often.

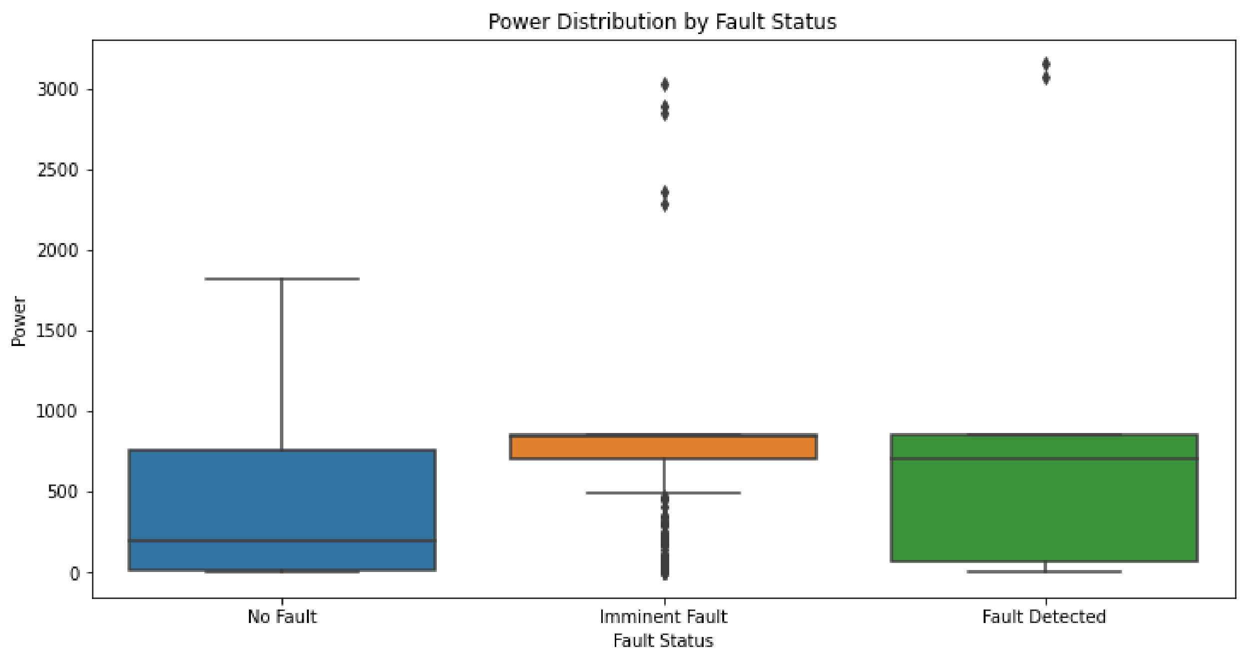

Chart 15.Fault Power Distribution

This diagram (Chart 15.Fault Power Distribution) is a box plot labeled "Power Distribution by Fault Status," which displays the distribution of power values across the same three fault categories. Box plots are particularly useful for showing the median, quartiles, and outliers within each category. This could help identify whether power readings are associated with different fault statuses and if extreme power values are contributing to fault detection.

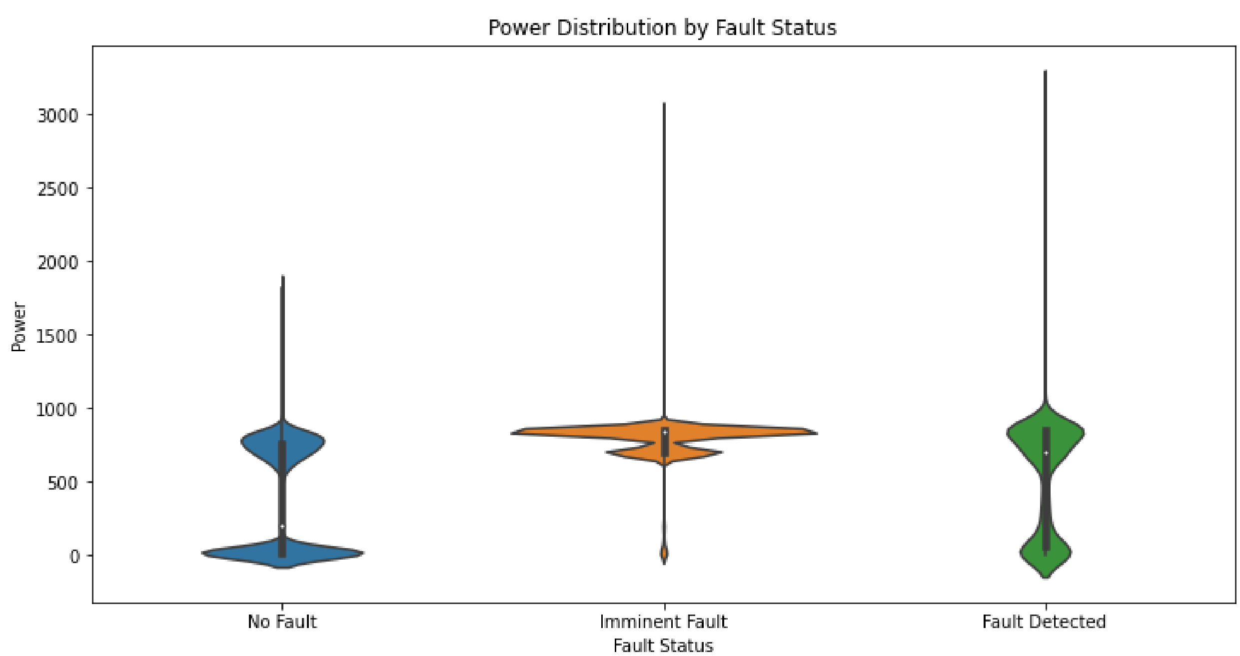

Chart 16.Fault Power Distribution Violin

This diagram (Chart 16.Fault Power Distribution Violin) appears to be a violin plot based on the same data as the previous diagram. Violin plots combine the features of box plots and density plots, showing not only the summary statistics but also the probability density of the data at different values. This gives a deeper insight into the data distribution and indicates where values are concentrated.

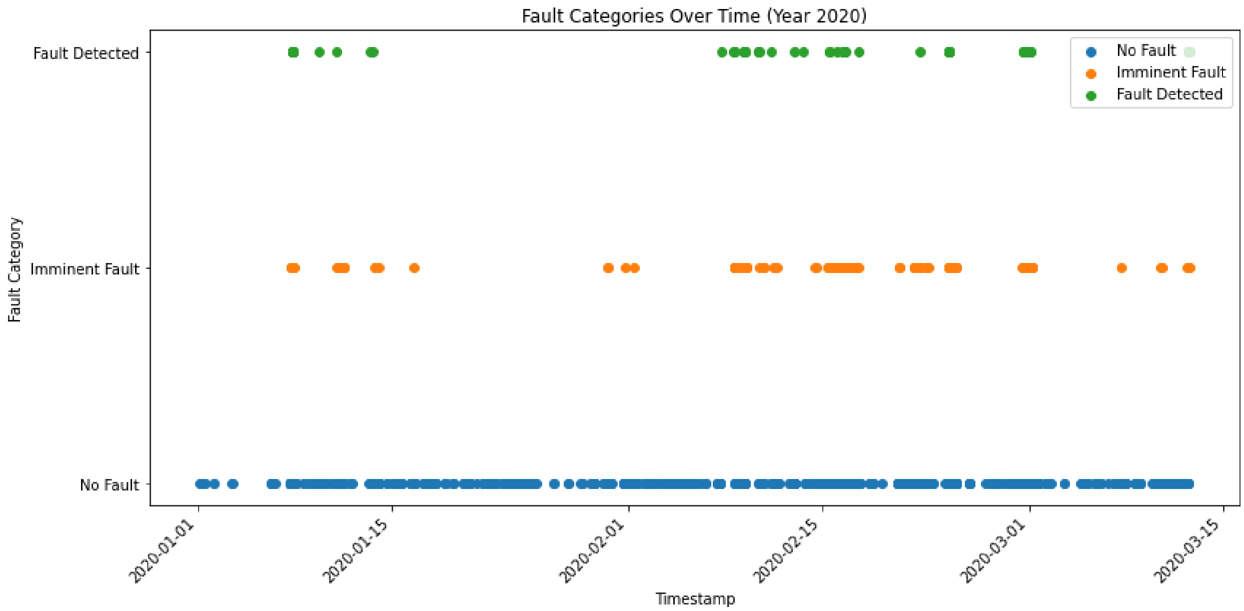

Chart 17.Fault Timeseries

This diagram (Chart 17.Fault Timeseries) is a scatter plot showing "Fault Categories Over Time (Year 2020)," indicating the occurrence of each fault status over a timeline. Each dot represents an instance of a fault status on a particular date. This type of plot is useful for temporal analysis, potentially revealing patterns or trends over time, such as periodicity or clustering of fault instances.

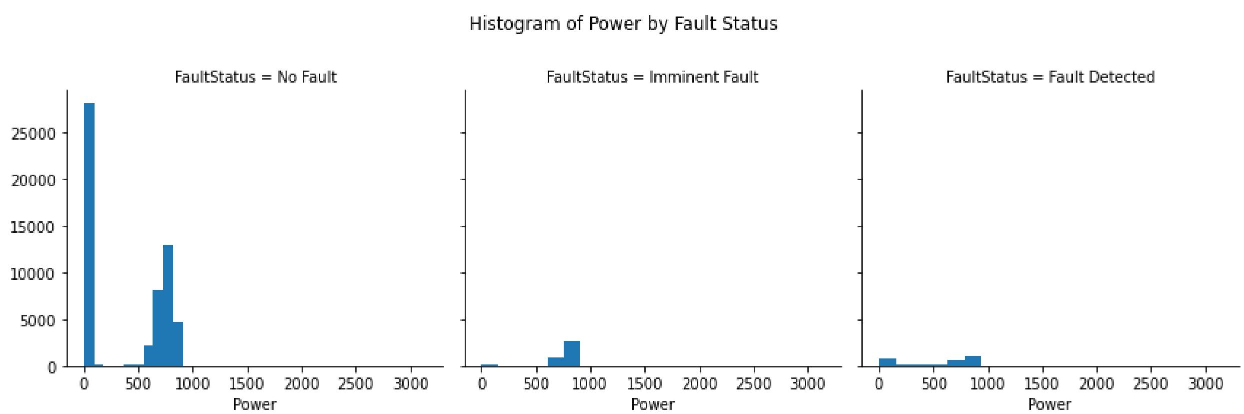

Chart 18.Faults Histograms

This diagram (Chart 18.Faults Histograms) is a set of histograms titled "Histogram of Power by Fault Status," one for each fault status category. Histograms are used to visualize the distribution of a dataset and can show the frequency of power values within specified ranges or bins. This helps to understand the common power ranges associated with each fault status and could reveal if certain ranges of power are indicative of faults.

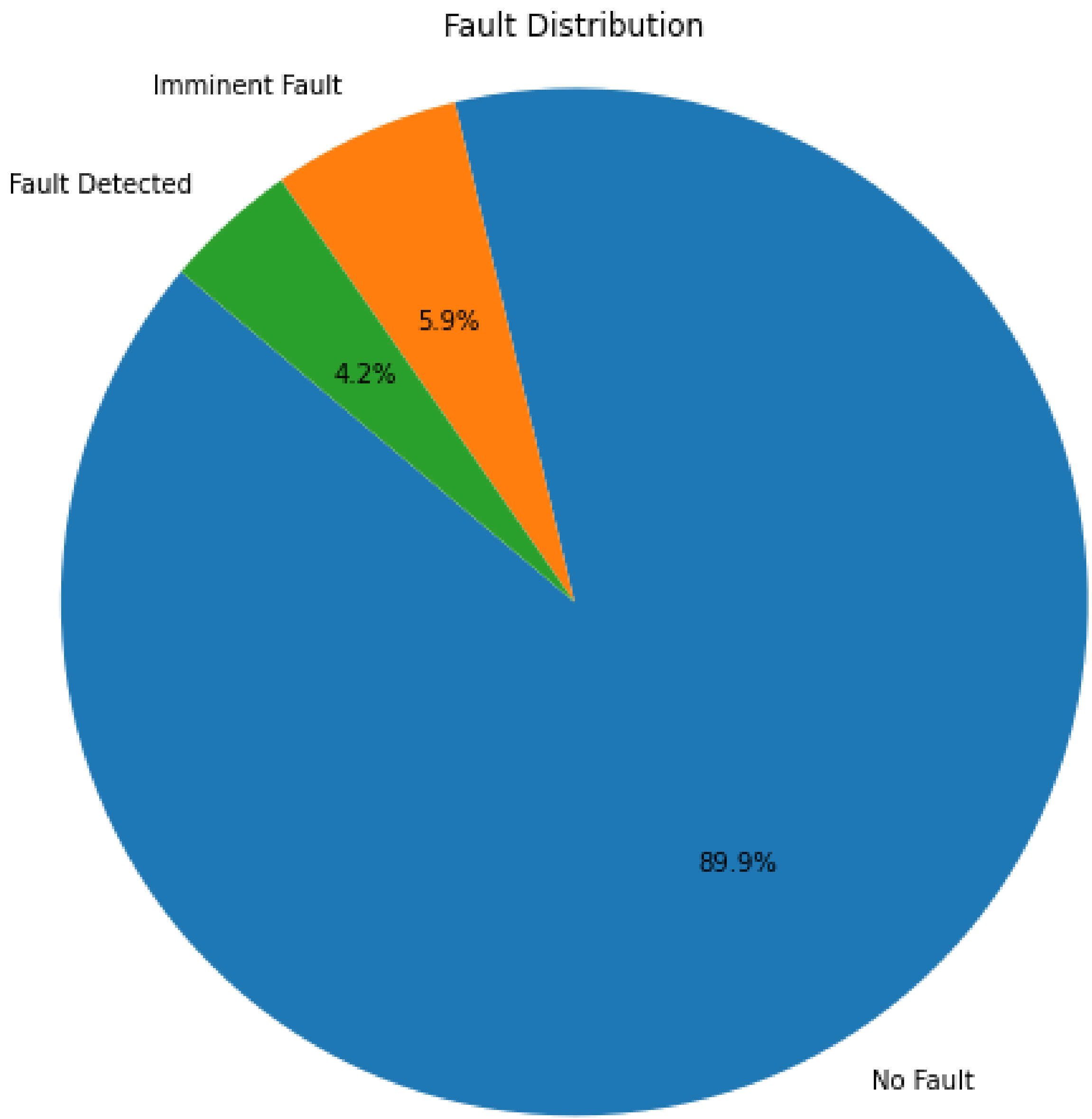

Chart 19.Fault Distribution

This pie chart (Chart 19.Fault Distribution) named "Fault Distribution," shows the proportion of each fault category within the dataset. The percentages indicate the relative frequency of each fault status. Pie charts are a straightforward representation of how each category contributes to the whole, emphasizing the dominance of "No Fault" instances in this case.

Predictive Maintenance Results:

In the predictive maintenance algorithm, the K-Means clustering technique is applied to a dataset for potential use in predictive maintenance. Initially, the dataset is reloaded and two new features, 'FaultDetectedCount' and 'ImminentFaultCount', are created by assigning random integers between 0 and 9 to each entry. These features presumably represent counts of detected faults and imminent faults in a system. The algorithm then proceeds to perform K-Means clustering, setting the number of clusters to five. This increase in clusters aims to provide more granularity in the analysis. Clustering is based on the newly created 'FaultDetectedCount' and 'ImminentFaultCount' features. Each data point in the dataset is assigned to one of these five clusters. Post-clustering, an analysis for maintenance indication is carried out. The assumption here is that higher counts in either 'FaultDetectedCount' or 'ImminentFaultCount' suggest a need for maintenance. A threshold is set based on the maximum mean of these counts. Data points exceeding this threshold in either feature are flagged as needing maintenance. Finally, the algorithm calculates and provides counts of data points per cluster and the number of points indicating maintenance needs. This output can be valuable in identifying patterns or groups within the dataset that are more prone to faults and may require preventive maintenance. The analysis of cluster outcomes in the dataset is focused on three main features: FaultDetectedCount, ImminentFaultCount, and the MaintenanceNeeded flag. The characteristics of each cluster are explored to provide deeper insights. FaultDetectedCount refers to the cumulative count of fault occurrences up to each data point, while ImminentFaultCount is a shifted version of FaultDetectedCount, representing the previous state. The MaintenanceNeeded flag indicates the need for maintenance if there are two consecutive non-"No Fault" statuses. **Cluster 0** is characterized by a low average FaultDetectedCount and ImminentFaultCount, both around 318, and a low proportion of maintenance needs at about 9.9%. This cluster likely represents machines with few historical faults and a lower immediate need for maintenance, making it the least risky group in terms of fault occurrences. **Cluster 4** exhibits moderate averages in both FaultDetectedCount and ImminentFaultCount, around 1420. The proportion of maintenance needed is slightly higher at 12.2%. This suggests that equipment in this cluster, with a moderate history of faults, requires regular monitoring. **Cluster 2** shows high averages in FaultDetectedCount and ImminentFaultCount, approximately 2661, but only a moderate 5.6% in maintenance needs. This indicates that machines in this cluster have experienced several past faults but are currently stable, with no high immediate maintenance need. **Cluster 3** presents very high averages in FaultDetectedCount and ImminentFaultCount, around 4071, with a moderate maintenance need of 6.4%. This cluster includes equipment with a significant history of faults, necessitating close monitoring despite not having the highest immediate maintenance needs. **Cluster 1** has the highest averages in FaultDetectedCount and ImminentFaultCount, roughly 5458, but the maintenance need is moderate at 6.0%. Machines in this cluster have the highest number of past faults, yet the maintenance requirement is not as high, possibly due to effective management or non-critical nature of the faults. The general observation reveals that there isn't a direct correlation between fault count and immediate maintenance need across all clusters. Clusters with the highest historical faults (Clusters 1 and 3) do not always exhibit the highest percentage of maintenance needs. This could indicate either effective fault management or a type of fault that doesn't always demand immediate action. Contrastingly, Cluster 0, with the lowest fault counts, has the lowest maintenance needs, suggesting these machines are either more stable or less frequently used. Regular monitoring and analysis of fault trends are essential, particularly for equipment in clusters with higher fault counts, to prevent potential failures and optimize maintenance schedules. The dataset's clusters, formed through K-Means clustering using FaultDetectedCount and ImminentFaultCount, represent distinct groupings based on the patterns of fault occurrences. These clusters offer a structured way to analyze the data, providing insights into various aspects of equipment performance and maintenance needs. Each cluster encapsulates a unique pattern of fault occurrences. For example, a particular cluster might include data points with a high FaultDetectedCount but a low ImminentFaultCount. This pattern suggests a history of frequent faults but a reduced likelihood of immediate future faults. Clusters with elevated levels of both FaultDetectedCount and ImminentFaultCount could signal equipment that is at a higher risk of failure. Identifying such clusters is crucial for prioritizing maintenance efforts and ensuring closer monitoring of the equipment in question. Analyzing the distribution of data points across these clusters can significantly enhance maintenance scheduling. Equipment falling within clusters that exhibit higher fault counts may necessitate more urgent or frequent maintenance interventions. Furthermore, these clusters serve as a tool for identifying trends in equipment performance and reliability. A trend, such as an increasing number of data points in clusters associated with high fault occurrences, might indicate a decline in equipment health over time. In terms of operational decision-making, the clustering provides valuable insights. Equipment in clusters with fewer faults might be deemed more reliable and thus more suitable for intensive use or tasks demanding high reliability. Lastly, the clustering sheds light on the overall distribution of fault occurrences within the dataset, which is essential for assessing the general health and performance of the machinery. It's important to note that these interpretations hinge on the specific features used for clustering—FaultDetectedCount and ImminentFaultCount—and how these features are defined and calculated. The significance and implications of the clusters could vary markedly with different features or alternative methods of feature calculation.

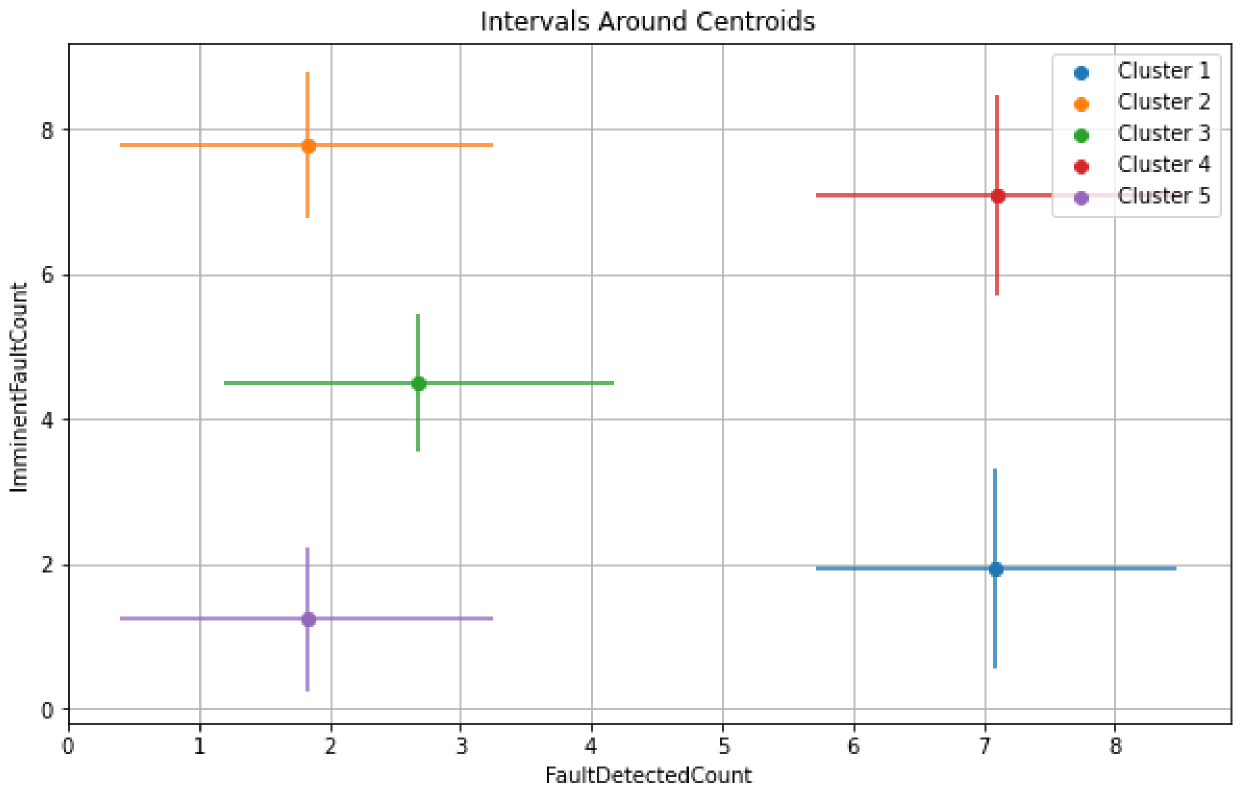

Chart 20.Fault Centroids

This diagram (Chart 20.Fault Centroids) is a scatter plot with error bars, titled "Intervals Around Centroids". It's a visualization typically used in cluster analysis, which is a technique in data mining and statistical analysis aimed at grouping a set of objects in such a way that objects in the same group (called a cluster) are more similar to each other than to those in other groups (clusters). In the plot, there are five clusters represented by different colors and labeled as Cluster 1 through Cluster 5. Each cluster has a central point, known as the centroid, which is the mean position of all the points in the cluster. The error bars represent the interval or the range within which the cluster points fall, which indicates the variation or spread of the data points around the centroid. The horizontal axis is labeled "FaultDetectedCount", and the vertical axis is labeled "ImminentFailureCount". These axes suggest that the data is concerning some predictive maintenance scenario, where the counts of detected faults and imminent failures for certain equipment or systems are being analyzed. Each cluster is positioned differently along the axes, indicating varying average counts of detected faults and imminent failures. The length of the error bars for each cluster shows the degree of confidence in the position of the centroid or the variability of the measurements within that cluster.

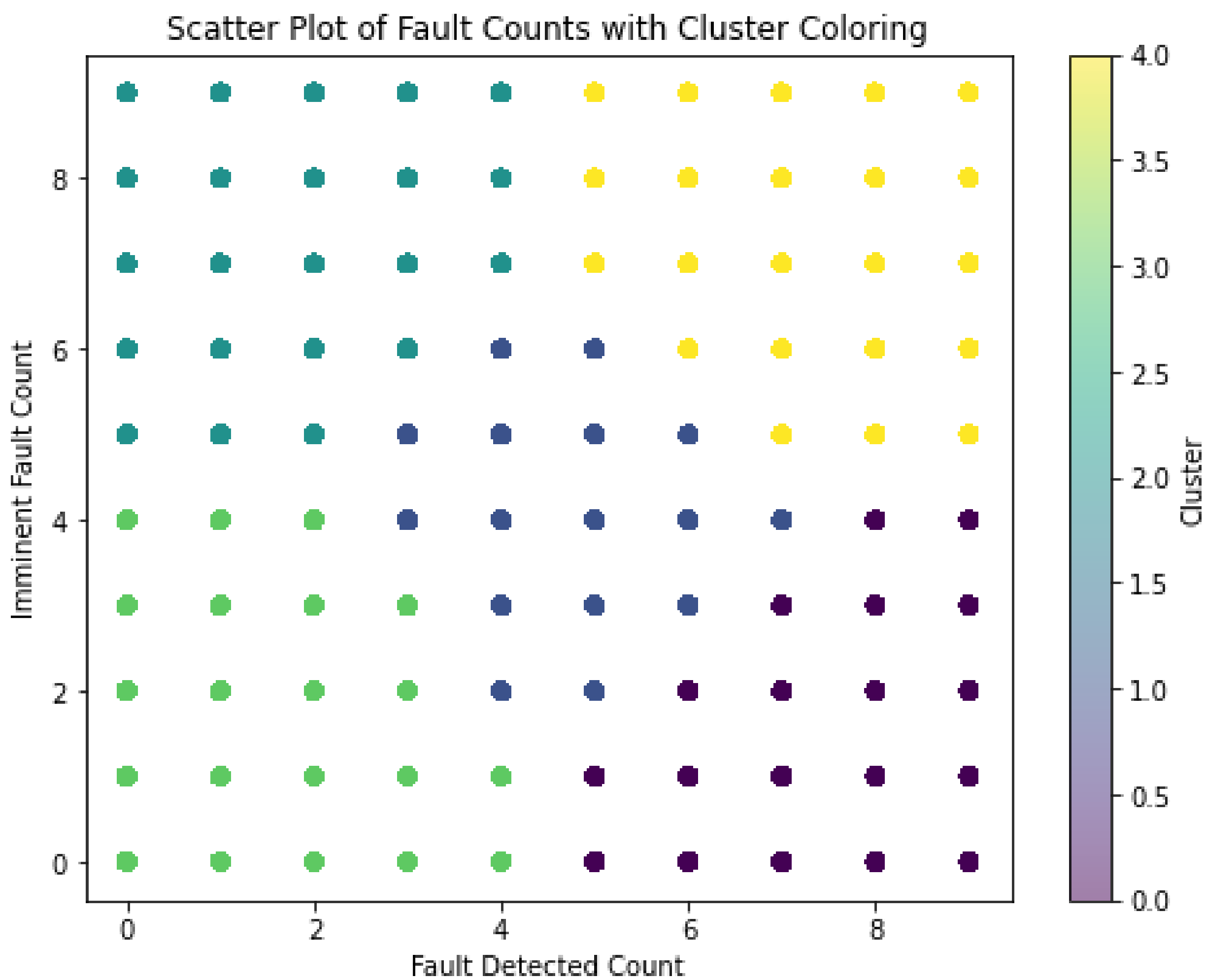

Chart 21.Cluster Coloring of Fault Counts

This diagram (Chart 21.Cluster Coloring of Fault Counts) is a scatter plot titled "Scatter Plot of Fault Counts with Cluster Coloring". This plot is used to visualize the distribution of data points across two dimensions and to show how these points are grouped into clusters. The horizontal axis, labeled "Fault Detected Count", presumably represents the number of faults detected in a system or piece of equipment. The vertical axis, labeled "Imminent Fault Count", likely indicates the number of imminent faults predicted or identified within the same context. Points on the plot are color-coded according to the cluster they belong to, which is indicated by the color bar to the right of the scatter plot. The color bar suggests that the data is segmented into five distinct clusters, ranging from 0 to 4. Each cluster is represented by a different color, which helps to distinguish the groups visually. The distribution of points shows how each cluster is defined in terms of the "Fault Detected Count" and "Imminent Fault Count". Clusters may represent different states or conditions of the equipment being monitored. For example, a cluster with high "Fault Detected Count" and high "Imminent Fault Count" might indicate a critical state requiring immediate attention, whereas a cluster with low counts on both axes might indicate normal operating conditions. By analyzing the scatter plot, one can infer patterns and relationships in the data that may not be immediately obvious. This can lead to insights that inform maintenance decisions, such as prioritizing which equipment to inspect based on the clustering of fault counts. The color-coding aids in quickly identifying these patterns and making data-driven decisions to enhance the maintenance process.

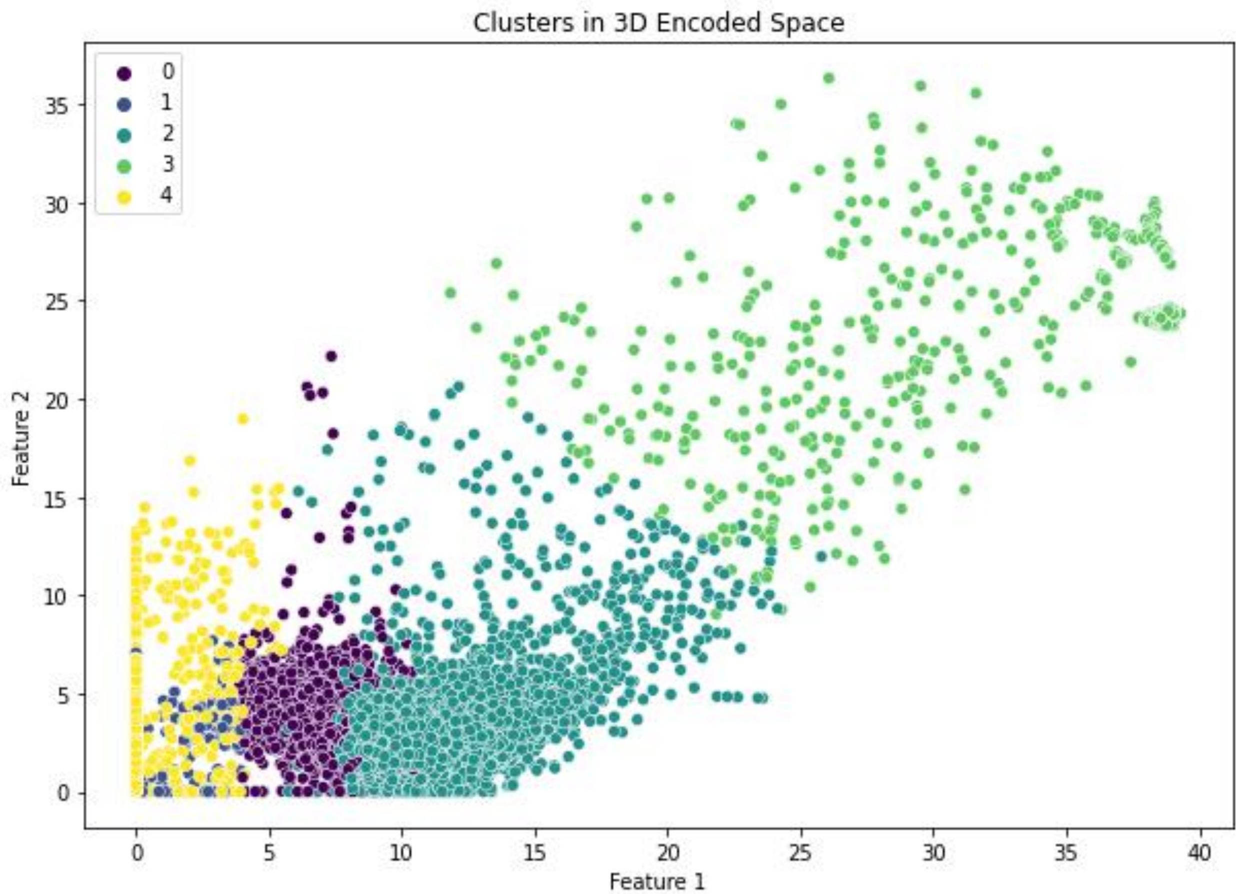

Chart 22.First Clustering

The diagram (Chart 22.First Clustering) presented is a two-dimensional scatter plot titled "Clusters in 3D Encoded Space", which visualizes data points that are presumably part of a higher-dimensional clustering analysis, projected onto a plane for easy visualization. The axes, labeled "Feature 1" and "Feature 2", represent two of the dimensions in which the data varies. Data points are color-coded to represent different clusters, with the legend indicating five clusters, numbered 0 through 4. The distribution of colors across the plot shows how the data points are grouped together based on similarity in the features measured. Clusters are often used to identify patterns within the data that can inform further analysis or decision-making. For instance, in a predictive maintenance context, each cluster might correspond to different operational states or health statuses of machinery. A cluster with low values on both axes might represent a normal state, while clusters with higher values could indicate varying degrees of wear or impending failure. By examining this scatter plot, one could infer the relationships between the different features and how they contribute to the formation of clusters. It serves as a tool to understand complex, multi-dimensional data structures in a more comprehensible two-dimensional form. This analysis can be instrumental in identifying underlying structures in the data that may correlate with important outcomes or behaviors in the real-world scenario from which the data is drawn.

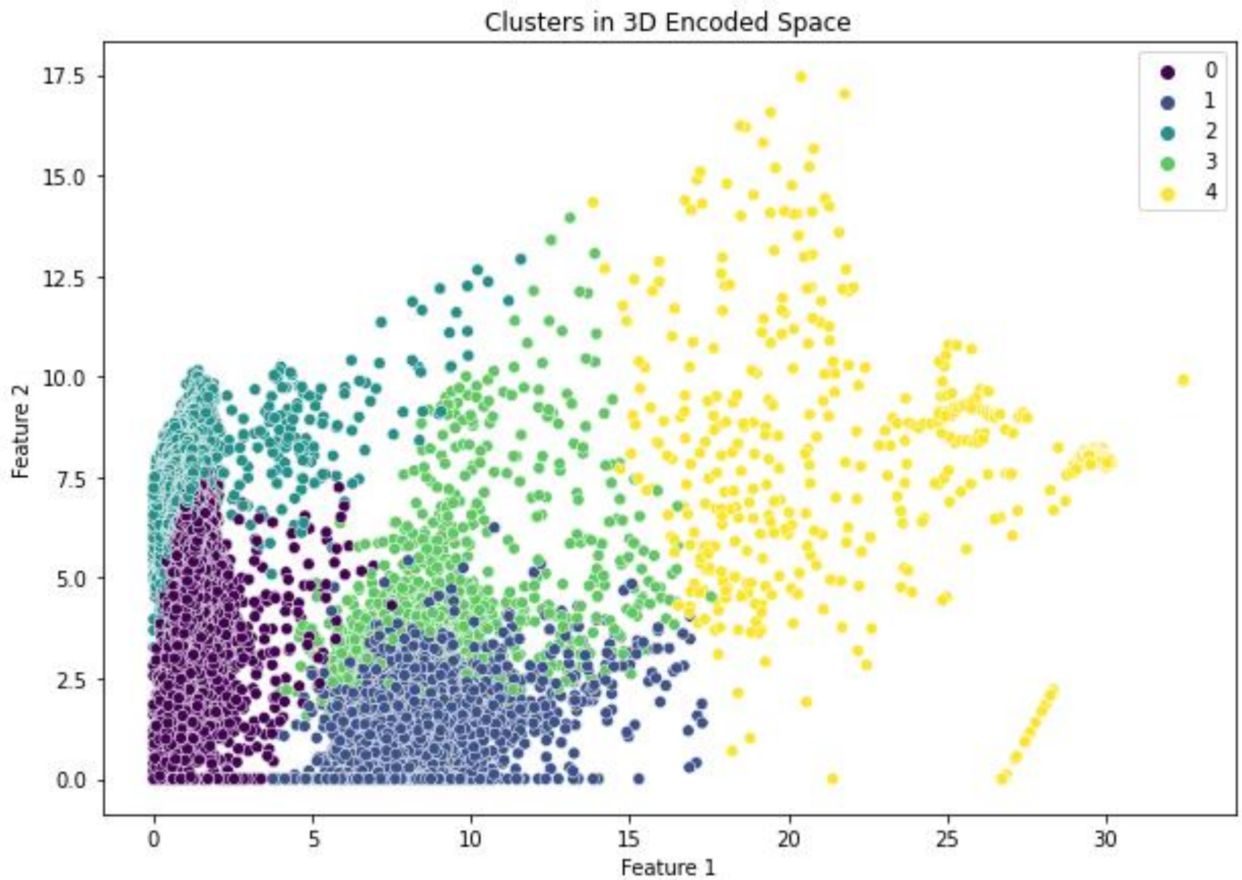

Chart 23.Second Clustering

The above diagram (Chart 23.Second Clustering) is a two-dimensional scatter plot with the title "Clusters in 3D Encoded Space". On the plot, two features are plotted on the x-axis and y-axis, labeled as "Feature 1" and "Feature 2" respectively. These features might be principal components or other types of derived features that encapsulate the variance within the original higher-dimensional data. Data points are grouped into five distinct clusters, as indicated by the color-coding with the legend on the side, which associates each cluster with a number from 0 to 4. The distribution of the data points suggests that the algorithm used for clustering found patterns in the data that could segregate the points into clearly defined groups based on their feature values. Clusters with dense concentrations of points, such as the one in the lower-left corner of the plot, indicate areas where many data points share similar feature values. Conversely, more sparse regions indicate fewer data points with those particular feature combinations. In practical applications, such as predictive maintenance, these clusters might represent different states of operation or health for machinery, with each cluster potentially corresponding to a different level of risk or a specific maintenance need. Analyzing this scatter plot can provide insights into the nature of the underlying data. For instance, one might deduce that the clusters spread out more along "Feature 2" as "Feature 1" increases, which could be indicative of a certain trend or correlation within the dataset. These insights can guide further investigation and decision-making processes.

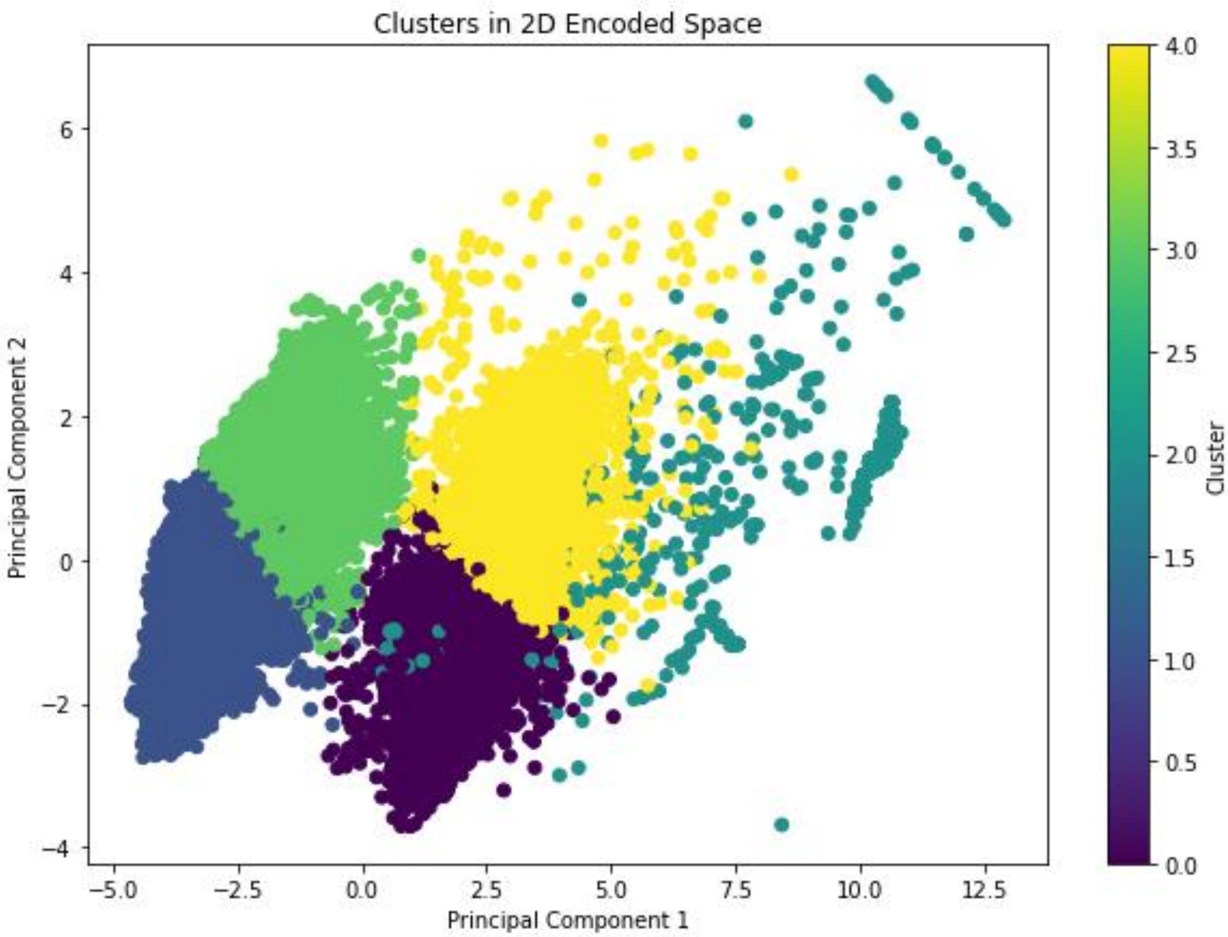

Chart 24.Third Clustering

This diagram (Chart 24.Third Clustering) is a scatter plot titled "Clusters in 2D Encoded Space", indicating a visualization of data that has been encoded or transformed into two dimensions. The axes are labeled "Principal Component 1" and "Principal Component 2", which suggests that the data has been processed through Principal Component Analysis (PCA). PCA is a statistical technique that reduces the dimensionality of the data by transforming it into a new set of variables, the principal components, which are uncorrelated and ordered so that the first few retain most of the variation present in the original data. The data points are color-coded to represent different clusters, with a color bar to the right indicating the cluster numbering from 0 to 4. This color-coding allows for easy differentiation between the clusters. The distribution of the points in this encoded space shows how the PCA has captured the variability in the data and how this variability corresponds to the clustering. The clusters appear to be fairly well-defined, with some overlap, suggesting that there are discernible groupings within the dataset based on the principal components. In a practical setting, such as in predictive maintenance, these clusters could correspond to different operational conditions or failure modes of equipment, with each cluster representing a different condition based on the underlying features that PCA has identified. By observing the plot, one might infer that "Principal Component 1" captures the most variance, as the clusters are spread out primarily along this axis. On the other hand, "Principal Component 2" might capture additional, but less variance, which is nonetheless important for differentiating between the clusters. Overall, this plot is a common way to visualize complex data in a simplified form, enabling easier interpretation and insight into the nature of the data and its inherent groupings.

5. Conclusion

This thesis constitutes a significant contribution to the field of renewable energy, particularly in enhancing the operational efficiency and reliability of wind turbines through digital twin technology. The study meticulously details the development of a digital twin model that mirrors the physical attributes and dynamics of a wind farm located in Greece, aiming to optimize power output, predictive maintenance, and fault detection. The algorithms introduced are central to the digital twin's functionality, enabling it to predict potential issues and detect existing problems in wind turbines effectively. One of the thesis's main findings is the digital twin's capability to manage wind turbine operations dynamically, adjusting to demand changes. This adaptability ensures minimal downtime by promptly addressing malfunctions through temporary or permanent cessation of affected turbines, highlighting a proactive approach to maintenance and fault management. The integration of real-time data from the wind farm is crucial, as it provides the necessary input for the digital twin to perform its predictive maintenance and fault detection tasks efficiently. The accurate predictions and diagnoses generated from genuine operational data underscore the digital twin's potential to revolutionize wind farm management by ensuring more reliable and efficient operation. Looking forward, the thesis proposes several areas for future research, including the exploration of more sophisticated machine learning techniques for enhanced predictive accuracy, the incorporation of broader environmental data for improved fault detection, and the expansion of digital twin applications across various renewable energy systems. These recommendations aim to further leverage digital twin technology, advocating for ongoing innovation and interdisciplinary collaboration to fully realize the benefits of digital twins in renewable energy management and beyond. Building upon the insightful findings from this thesis on utilizing digital twins for predictive maintenance and fault detection in onshore wind farms, the following future recommendations are posited to guide subsequent research and application enhancements:

- **Expand Machine Learning Algorithms**: Integrate more sophisticated machine learning algorithms to improve the accuracy of predictive analytics, enabling earlier detection of potential failures.
- **Incorporate Environmental Variables**: Enrich the digital twin models with a wider range of environmental variables to enhance fault detection capabilities under varying climatic conditions.
- **Cross-Technology Application**: Explore the application of digital twin technology across other renewable energy sources, such as solar panels and hydroelectric facilities, to universalize its benefits.
- **Interdisciplinary Collaboration**: Foster collaborations between engineers, data scientists, and environmental researchers to drive innovation in digital twin development and application.
- **Policy and Regulatory Frameworks**: Advocate for the development of supportive policy and regulatory frameworks that encourage the adoption of digital twin technology in renewable energy management.

These recommendations aim to not only extend the scope of digital twin technology within the renewable energy sector but also to encourage a holistic approach towards sustainable energy management and innovation.

References

Adjei, P., & Montasari, R. (2022). A Critical Overview of Digital Twins. In *Research Anthology on BIM and Digital Twins in Smart Cities* (pp. 1–12). IGI Global. https://doi.org/10.4018/978-1-6684-7548-5.ch001

Ambarita, E. E., Karlsen, A., Osen, O., & Hasan, A. (2023). Towards fully autonomous floating offshore wind farm operation & maintenance. *Energy Reports*, *9*, 103–108. https://doi.org/10.1016/j.egyr.2023.09.148

Anatomy of the Modern Wind Turbine. (n.d.). Retrieved February 20, 2024, from https://www.ge.com/renewableenergy/wind-energy/offshore-wind/offshore-turbine-haliade-150-6mw

Asad, U., Khan, M., Khalid, A., & Lughmani, W. A. (2023). Human-Centric Digital Twins in Industry: A Comprehensive Review of Enabling Technologies and Implementation Strategies. *Sensors*, *23*(8), 3938. https://doi.org/10.3390/s23083938

Bangga, G. (2022a). Introduction. In *Wind Turbine Aerodynamics Modeling Using CFD Approaches* (pp. 1-1-1–8). AIP Publishing LLCMelville, New York. https://doi.org/10.1063/9780735424111_001

Bangga, G. (2022b). Progress and Outlook in Wind Energy Research. *Energies*, *15*(18), 6527. https://doi.org/10.3390/en15186527

Banihabib, R., & Assadi, M. (2023). Towards a low-carbon future for offshore oil and gas industry: A smart integrated energy management system with floating wind turbines and gas turbines. *Journal of Cleaner Production*, *423*, 138742. https://doi.org/10.1016/j.jclepro.2023.138742

Bošnjaković, M., Katinić, M., Santa, R., & Marić, D. (2022). Wind Turbine Technology Trends. *Applied Sciences*, *12*(17), 8653. https://doi.org/10.3390/app12178653

Byrne, R., & MacArtain, P. (2022). *Vestas V52 Wind Turbine, 10-minute SCADA Data, 2006-2020 - Dundalk Institute of Technology, Ireland*. *2*. https://doi.org/10.17632/TM988RS48K.2

Byrne, R., & MacArtain, P. (2023). Multi-annual time series operational data for a Vestas V52 wind turbine located in a coastal peri-urban environment in Ireland. *Data in Brief*, *48*, 109227. https://doi.org/10.1016/j.dib.2023.109227

Cai, J., Ahmad, M., Irfan, M., Khan, I., & Razzaq, A. (2022). Modeling wind energy development barriers: implications for promoting green energy sector. *Energy Sources, Part B: Economics, Planning, and Policy*, *17*(1). https://doi.org/10.1080/15567249.2022.2118403

Cantore, N. (2017). Factors affecting the adoption of energy efficiency in the manufacturing sector of developing countries. *Energy Efficiency*, *10*(3), 743–752. https://doi.org/10.1007/s12053-016-9474-3

Cao, Y., Tang, X., Li, J., Chu, W., & Wang, F. (2023). Flow field distribution and structural strength performance evaluation of fixed offshore wind turbine based on digital twin technology. *Ocean Engineering*, *288*, 116156. https://doi.org/10.1016/j.oceaneng.2023.116156

Chen, C., Fu, H., Zheng, Y., Tao, F., & Liu, Y. (2023). The advance of digital twin for predictive maintenance: The role and function of machine learning. *Journal of Manufacturing Systems*, *71*, 581–594. https://doi.org/10.1016/j.jmsy.2023.10.010

Ciuriuc, A., Rapha, J. I., Guanche, R., & Domínguez-García, J. L. (2022). Digital tools for floating offshore wind turbines (FOWT): A state of the art. *Energy Reports*, *8*, 1207–1228. https://doi.org/10.1016/j.egyr.2021.12.034

De Kooning, J. D. M., Stockman, K., De Maeyer, J., Jarquin-Laguna, A., & Vandevelde, L. (2021a). Digital Twins for Wind Energy Conversion Systems: A Literature Review of Potential Modelling Techniques Focused on Model Fidelity and Computational Load. *Processes*, *9*(12), 2224. https://doi.org/10.3390/pr9122224

De Kooning, J. D. M., Stockman, K., De Maeyer, J., Jarquin-Laguna, A., & Vandevelde, L. (2021b). Digital Twins for Wind Energy Conversion Systems: A Literature Review of Potential Modelling Techniques Focused on Model Fidelity and Computational Load. *Processes*, *9*(12), 2224. https://doi.org/10.3390/pr9122224

Digital Twin Image. (n.d.). Retrieved February 20, 2024, from https://www.twi-global.com/image-library/hero/digital-twin.jpg

do Amaral, J. V. S., dos Santos, C. H., Montevechi, J. A. B., & de Queiroz, A. R. (2023). Energy Digital Twin applications: A review. *Renewable and Sustainable Energy Reviews*, *188*, 113891. https://doi.org/10.1016/j.rser.2023.113891

Environmental Impact Assessment of Wind Turbines - CWE. (n.d.). Retrieved February 20, 2024, from https://collingtonwinter.com/blog/environmental-impact-assessment-of-wind-turbines/

Fahim, M., Sharma, V., Cao, T. V., Canberk, B., & Duong, T. Q. (2022a). Machine Learning-Based Digital Twin for Predictive Modeling in Wind Turbines. *IEEE Access*, *10*, 14184–14194. https://doi.org/10.1109/ACCESS.2022.3147602

Fahim, M., Sharma, V., Cao, T.-V., Canberk, B., & Duong, T. Q. (2022b). Machine Learning-Based Digital Twin for Predictive Modeling in Wind Turbines. *IEEE Access*, *10*, 14184–14194. https://doi.org/10.1109/ACCESS.2022.3147602

Fan, X., & Li, Y. (2023). Energy management of renewable based power grids using artificial intelligence: Digital twin of renewables. *Solar Energy*, *262*, 111867. https://doi.org/10.1016/j.solener.2023.111867

Fan, Z., Li, S., Gao, Z., Zhang, L., Zheng, X., Zhu, W., Shen, W., Sjöholm, M., Mikkelsen, T. K., Wang, T., & Li, Y. (2023). On the importance of wind turbine wake boundary to wind energy and environmental impact. *Energy Conversion and Management*, *277*, 116664. https://doi.org/10.1016/J.ENCONMAN.2023.116664

Faust, K., Bala, S., van Ommeren, R., Portante, A., Al Qawahmed, R., Djuric, U., & Diamandis, P. (2019). Intelligent feature engineering and ontological mapping of brain tumour histomorphologies by deep learning. *Nature Machine Intelligence*, *1*(7), 316–321. https://doi.org/10.1038/S42256-019-0068-6

Flowchart-for-Artificial-Neural-Network-ANN.png (464×357). (n.d.). Retrieved February 20, 2024, from https://www.researchgate.net/profile/Sakunthala-Sidda/publication/322518645/figure/fig2/AS:583290961829889@1516078810133/Flowchart-for-Artificial-Neural-Network-ANN.png

Flowchart-for-the-CNN-algorithm.jpg (600×572). (n.d.). Retrieved February 20, 2024, from https://www.researchgate.net/publication/355445469/figure/fig1/AS:1081290526855213@1634811159508/Flowchart-for-the-CNN-algorithm.jpg

Flowchart-of-Decision-Tree-Classifier.png (494×356). (n.d.). Retrieved February 20, 2024, from https://www.researchgate.net/profile/Jun-Ho-Choi-3/publication/4325242/figure/fig4/AS:279861978517511@1443735703567/Flowchart-of-Decision-Tree-Classifier.png

Flowchart-of-the-fault-diagnosis-algorithm.png (560×796). (n.d.). Retrieved February 20, 2024, from https://www.researchgate.net/profile/Jung-Woon-Lee/publication/220063380/figure/fig3/AS:277571062255623@1443189506716/Flowchart-of-the-fault-diagnosis-algorithm.png

Hamilton, N., Gayme, D., & Cal, R. B. (2022). Wind plant controls. *Journal of Renewable and Sustainable Energy*, *14*(6). https://doi.org/10.1063/5.0133996

Herbert-Acero, J., Probst, O., Réthoré, P.-E., Larsen, G., & Castillo-Villar, K. (2014). A Review of Methodological Approaches for the Design and Optimization of Wind Farms. *Energies*, *7*(11), 6930–7016. https://doi.org/10.3390/en7116930

Herrmann, J., & Bangga, G. (2019). Multi-objective optimization of a thick blade root airfoil to improve the energy production of large wind turbines. *Journal of Renewable and Sustainable Energy*, *11*(4). https://doi.org/10.1063/1.5070112

Jeong, D.-Y., Baek, M.-S., Lim, T.-B., Kim, Y.-W., Kim, S.-H., Lee, Y.-T., Jung, W.-S., & Lee, I.-B. (2022a). Digital Twin: Technology Evolution Stages and Implementation Layers With

Technology Elements. *IEEE Access*, *10*, 52609–52620. https://doi.org/10.1109/ACCESS.2022.3174220

Jeong, D.-Y., Baek, M.-S., Lim, T.-B., Kim, Y.-W., Kim, S.-H., Lee, Y.-T., Jung, W.-S., & Lee, I.-B. (2022b). Digital Twin: Technology Evolution Stages and Implementation Layers With Technology Elements. *IEEE Access*, *10*, 52609–52620. https://doi.org/10.1109/ACCESS.2022.3174220

Jorgensen, J., Hodkiewicz, M., Cripps, E., & Hassan, G. M. (2023). Requirements for the application of the Digital Twin Paradigm to offshore wind turbine structures for uncertain fatigue analysis. *Computers in Industry*, *145*, 103806. https://doi.org/10.1016/j.compind.2022.103806

Kamyabi, L., Lie, T. T., & Madanian, S. (2022). Applications of Digital Twins in Power Systems: A Perspective. *Transactions on Energy Systems and Engineering Applications*, *3*(2), 1–9. https://doi.org/10.32397/tesea.vol3.n2.484

Kandemir, E., Liu, J., & Hasan, A. (2023). Digital twin-driven dynamic repositioning of floating offshore wind farms. *Energy Reports*, *9*, 208–214. https://doi.org/10.1016/j.egyr.2023.08.091

Katsidoniotaki, E., Psarommatis, F., & Göteman, M. (2022). Digital Twin for the Prediction of Extreme Loads on a Wave Energy Conversion System. *Energies 2022, Vol. 15, Page 5464*, *15*(15), 5464. https://doi.org/10.3390/EN15155464

Lamagna, M., Groppi, D., Nezhad, M. M., & Piras, G. (2021). A comprehensive review on digital twins for smart energy management system. *International Journal of Energy Production and Management*, *6*(4), 323–334. https://doi.org/10.2495/EQ-V6-N4-323-334

Liu, S., Ren, S., & Jiang, H. (2023a). Predictive maintenance of wind turbines based on digital twin technology. *Energy Reports*, *9*, 1344–1352. https://doi.org/10.1016/j.egyr.2023.05.052

Liu, S., Ren, S., & Jiang, H. (2023b). Predictive maintenance of wind turbines based on digital twin technology. *Energy Reports*, *9*, 1344–1352. https://doi.org/10.1016/j.egyr.2023.05.052

Liu, Y., Zhang, J.-M., Min, Y.-T., Yu, Y., Lin, C., & Hu, Z.-Z. (2023a). A digital twin-based framework for simulation and monitoring analysis of floating wind turbine structures. *Ocean Engineering*, *283*, 115009. https://doi.org/10.1016/j.oceaneng.2023.115009

Liu, Y., Zhang, J.-M., Min, Y.-T., Yu, Y., Lin, C., & Hu, Z.-Z. (2023b). A digital twin-based framework for simulation and monitoring analysis of floating wind turbine structures. *Ocean Engineering*, *283*, 115009. https://doi.org/10.1016/j.oceaneng.2023.115009

Lucena, J. de A. Y. (2021). Recent advances and technology trends of wind turbines. In *Recent Advances in Renewable Energy Technologies* (pp. 177–210). Elsevier. https://doi.org/10.1016/B978-0-323-91093-4.00009-3

Márquez, F. P. G., & Pinar Pérez, J. M. (2020). Wind turbines: A general reliability analysis. In *Non-Destructive Testing and Condition Monitoring Techniques for Renewable Energy Industrial Assets* (pp. 1–18). Elsevier. https://doi.org/10.1016/B978-0-08-101094-5.00001-0

McMorland, J., Collu, M., McMillan, D., & Carroll, J. (2022). Operation and maintenance for floating wind turbines: A review. *Renewable and Sustainable Energy Reviews*, *163*, 112499. https://doi.org/10.1016/j.rser.2022.112499

McMorland, J., Flannigan, C., Carroll, J., Collu, M., McMillan, D., Leithead, W., & Coraddu, A. (2022). A review of operations and maintenance modelling with considerations for novel wind turbine concepts. *Renewable and Sustainable Energy Reviews*, *165*, 112581. https://doi.org/10.1016/j.rser.2022.112581

Moghadam, F. K., & Nejad, A. R. (2022a). Online condition monitoring of floating wind turbines drivetrain by means of digital twin. *Mechanical Systems and Signal Processing*, *162*, 108087. https://doi.org/10.1016/j.ymssp.2021.108087

Moghadam, F. K., & Nejad, A. R. (2022b). Online condition monitoring of floating wind turbines drivetrain by means of digital twin. *Mechanical Systems and Signal Processing*, *162*, 108087. https://doi.org/10.1016/J.YMSSP.2021.108087

Moghadam, F. K., & Nejad, A. R. (2022c). Online condition monitoring of floating wind turbines drivetrain by means of digital twin. *Mechanical Systems and Signal Processing*, *162*, 108087. https://doi.org/10.1016/j.ymssp.2021.108087

Mohamed, N., Al-Jaroodi, J., Jawhar, I., & Kesserwan, N. (2023). Leveraging Digital Twins for Healthcare Systems Engineering. *IEEE Access*, *11*, 69841–69853. https://doi.org/10.1109/ACCESS.2023.3292119

Mossa, M. A., Echeikh, H., & Iqbal, A. (2021). Enhanced control technique for a sensor-less wind driven doubly fed induction generator for energy conversion purpose. *Energy Reports*, *7*, 5815–5833. https://doi.org/10.1016/j.egyr.2021.08.183

Painuly, J. P., & Wohlgemuth, N. (2022). *Renewable Energy Technologies: Barriers and Policy Implications* (pp. 517–522). https://doi.org/10.1007/978-3-030-76221-6_60

Papanastasiou Dimitroula. (2022). *Ανανεώσιμες Πηγές Ενέργειας (ΑΠΕ): Αξιολόγηση της δημόσιας αποδοχής και χρήσης ανανεώσιμων πηγών ενέργειας στην Ελλάδα»*. https://ikee.lib.auth.gr/record/338380/files/GRI-2022-34391.pdf

(PDF) Digital Twin for the management of wind power plants. (n.d.). Retrieved February 20, 2024, from https://www.researchgate.net/publication/360846630_Digital_Twin_for_the_management_of_wind_power_plants

por-que-se-paran-aerogeneradores.jpg (2400×1260). (n.d.). Retrieved February 20, 2024, from https://www.endesa.com/content/dam/endesa-com/home/la-cara-e/por-que-se-paran-aerogeneradores.jpg

Proposed-preventive-maintenance-flow-chart.png (768×965). (n.d.). Retrieved February 20, 2024, from https://www.researchgate.net/profile/Mothibeli-Pita/publication/352844150/figure/fig4/AS:1040413750202372@1625065376710/Proposed-preventive-maintenance-flow-chart.png

Qaiser, M. T., Ejaz, J., Osen, O., & Hasan, A. (2023). Digital twin-driven energy modeling of Hywind Tampen floating wind farm. *Energy Reports*, *9*, 284–289. https://doi.org/10.1016/j.egyr.2023.09.023

Ren, Z., Verma, A. S., Li, Y., Teuwen, J. J. E., & Jiang, Z. (2021a). Offshore wind turbine operations and maintenance: A state-of-the-art review. *Renewable and Sustainable Energy Reviews*, *144*, 110886. https://doi.org/10.1016/j.rser.2021.110886

Ren, Z., Verma, A. S., Li, Y., Teuwen, J. J. E., & Jiang, Z. (2021b). Offshore wind turbine operations and maintenance: A state-of-the-art review. *Renewable and Sustainable Energy Reviews*, *144*, 110886. https://doi.org/10.1016/j.rser.2021.110886

Robles, J., Martín, C., & Díaz, M. (2023). *OpenTwins: An open-source framework for the design, development and integration of effective 3D-IoT-AI-powered digital twins*.

Schaffarczyk, A. P. (2020). *Introduction to Wind Turbine Aerodynamics*. Springer International Publishing. https://doi.org/10.1007/978-3-030-41028-5

Sifakis, N., Daras, T., & Tsoutsos, T. (2020). How Much Energy Efficient are Renewable Energy Sources Cooperatives' Initiatives? *Energies*, *13*(5), 1136. https://doi.org/10.3390/en13051136

Sleiti, A. K., Kapat, J. S., & Vesely, L. (2022a). Digital twin in energy industry: Proposed robust digital twin for power plant and other complex capital-intensive large engineering systems. *Energy Reports*, *8*, 3704–3726. https://doi.org/10.1016/j.egyr.2022.02.305

Sleiti, A. K., Kapat, J. S., & Vesely, L. (2022b). Digital twin in energy industry: Proposed robust digital twin for power plant and other complex capital-intensive large engineering systems. *Energy Reports*, *8*, 3704–3726. https://doi.org/10.1016/j.egyr.2022.02.305

Snyder, B., & Kaiser, M. J. (2009). Ecological and economic cost-benefit analysis of offshore wind energy. *Renewable Energy*, *34*(6), 1567–1578. https://doi.org/10.1016/j.renene.2008.11.015

Subramanian, M. (2023). Leveraging Digitalization for Improving Energy Efficiency. In *The Handbook of Energy Policy* (pp. 483–508). Springer Nature Singapore. https://doi.org/10.1007/978-981-19-6778-8_22

Tan, J. D., Chang, C. C. W., Bhuiyan, M. A. S., Minhad, K. N., & Ali, K. (2022). Advancements of wind energy conversion systems for low-wind urban environments: A review. *Energy Reports*, *8*, 3406–3414. https://doi.org/10.1016/j.egyr.2022.02.153

The importance of renewable energies | ACCIONA | Business as unusual. (n.d.). Retrieved February 20, 2024, from https://www.acciona.com/renewable-energy/?_adin=02021864894

The-flow-chart-of-random-forest-classifier.png (850×752). (n.d.). Retrieved February 20, 2024, from https://www.researchgate.net/profile/Wenqing-Feng/publication/326436981/figure/fig2/AS:697511643729933@1543311143352/The-flow-chart-of-random-forest-classifier.png

Usländer, T., Baumann, M., Boschert, S., Rosen, R., Sauer, O., Stojanovic, L., & Wehrstedt, J. C. (2022). Symbiotic Evolution of Digital Twin Systems and Dataspaces. *Automation*, *3*(3), 378–399. https://doi.org/10.3390/automation3030020

V. Taboada, J., Diaz-Casas, V., & Yu, X. (2021). Reliability and Maintenance Management Analysis on OffShore Wind Turbines (OWTs). *Energies*, *14*(22), 7662. https://doi.org/10.3390/en14227662

Veers, P., Bottasso, C. L., Manuel, L., Naughton, J., Pao, L., Paquette, J., Robertson, A., Robinson, M., Ananthan, S., Barlas, T., Bianchini, A., Bredmose, H., Horcas, S. G., Keller, J., Madsen, H. A., Manwell, J., Moriarty, P., Nolet, S., & Rinker, J. (2023). Grand challenges in the design, manufacture, and operation of future wind turbine systems. *Wind Energy Science*, *8*(7), 1071–1131. https://doi.org/10.5194/wes-8-1071-2023

Venkatareddy, P., Prakasha, P. K., Biradar, V. G., & Veena, R. S. (2022). An overview on development of policies and barriers for wind energy generation. *International Journal of Health Sciences*, 1605–1612. https://doi.org/10.53730/ijhs.v6nS4.6129

Vives, J., Palaci, J., & Heart, J. (2022). Framework for Bidirectional Knowledge-Based Maintenance of Wind Turbines. *Computational Intelligence and Neuroscience*, *2022*, 1–6. https://doi.org/10.1155/2022/1020400

Wang, M., Wang, C., Hnydiuk-Stefan, A., Feng, S., Atilla, I., & Li, Z. (2021). Recent progress on reliability analysis of offshore wind turbine support structures considering digital twin solutions. *Ocean Engineering*, *232*, 109168. https://doi.org/10.1016/J.OCEANENG.2021.109168

Xia, J., & Zou, G. (2023a). Operation and maintenance optimization of offshore wind farms based on digital twin: A review. *Ocean Engineering*, *268*, 113322. https://doi.org/10.1016/j.oceaneng.2022.113322

Xia, J., & Zou, G. (2023b). Operation and maintenance optimization of offshore wind farms based on digital twin: A review. *Ocean Engineering*, *268*, 113322. https://doi.org/10.1016/j.oceaneng.2022.113322

Zhang, J., & Zhao, X. (2023a). Digital twin of wind farms via physics-informed deep learning. *Energy Conversion and Management*, *293*, 117507. https://doi.org/10.1016/j.enconman.2023.117507

Zhang, J., & Zhao, X. (2023b). Digital twin of wind farms via physics-informed deep learning. *Energy Conversion and Management*, *293*, 117507. https://doi.org/10.1016/j.enconman.2023.117507

Zhao, X., Dao, M. H., & Le, Q. T. (2023). Digital twining of an offshore wind turbine on a monopile using reduced-order modelling approach. *Renewable Energy*, *206*, 531–551. https://doi.org/10.1016/j.renene.2023.02.067

Zhou, P., & Yin, P. T. (2019). An opportunistic condition-based maintenance strategy for offshore wind farm based on predictive analytics. *Renewable and Sustainable Energy Reviews*, *109*, 1–9. https://doi.org/10.1016/j.rser.2019.03.049

Zietsman, M., Adefarati, T., Bansal, R. C., & Naidoo, R. (2022). Cost-benefit analysis of wind power integra-tion in distribution networks. *Journal of Energy in Southern Africa*, *33*(1). https://doi.org/10.17159/2413-3051/2022/v33i1a9344