Predictive Maintenance for Wind Turbine Bearings: An MLOps Approach with the DIAFS Machine Learning Model

Nataliya Shakhovska[1,2*†], Jaime Campos[3†]

[1] *Department of Systems of Artificial Intelligence, Lviv Polytechnic National University,Lviv,79905, Ukraine*
[2] *College of Engeneering, Design and Phisical Sciences, Brunel University London, Kingston Lane, Uxbridge, Middlesex UB8 3PH United Kingdom*
[3] *Department of Informatics, Linnaeus University, Sweden*

Abstract
This article underscores the importance of integrating machine learning analytics to enhance preventive maintenance methodologies, particularly emphasizing condition-based maintenance (CBM) within the wind energy domain. Through empirical evidence derived from wind turbine data, the paper outlines the efficacy and applicability of Machine Learning in Operations (MLOps) for predicting the residual operational life of wind turbine bearings. While the study's principal domain is renewable energy, especially wind power, it employs a specific wind turbine dataset for exhaustive model testing, leading to the proposition of an innovative ensemble model tailored for high-speed wind turbine bearing prognosis. The introduced model, "The Data Interpretation Algorithm for Forecasting Time Series" (DIAFS), crafted for assessing wind turbine bearing conditions, is predicated on an adaptive polynomial model approximation. It emerges as an indispensable asset for maintenance professionals implementing CBM methodologies.

Keywords
Condition-based maintenance, condition monitoring, machine learning, MLOps, Preventive maintenance, wind turbine bearing

1. Introduction

THE E.U.'s 2050 energy roadmap mandates member states to advance infrastructure for long-term energy system decarbonization. Estimates suggest a global population increase of 2 billion by 2050, requiring 47% more energy for a total of 10 billion people. Given the current energy system's inadequacies and climate goals, there's an urgent need for sustainable energy practices (energy.ec.europa.eu). In this context, aspects of the wind farms, especially the components, such as the wind turbines, are critical to investigate. Wind turbines (W.T.) are complex

---

SMARTINDUSTRY-2024: International Conference on Smart Automation & Robotics for Future Industry, April 18 - 20, 2024, Lviv, Ukraine
* Corresponding author.
† These authors contributed equally.
nataliya.b.shakhovska@lpnu.ua (N. Shakhovska)
0 0000-0002-6875-8534 (N. Shakhovska); 0000-0001-7048-8089 (J. Campos)

equipment with a design lifetime of approximately 20 years. However, its life span and availability can be augmented by the implementation of condition monitoring and preventive maintenance techniques related to operation and maintenance activities. As a result, this means that with effective operation and maintenance, these W.T.s are operational to be able to produce energy. In addition, Operating and maintenance (O&M) costs are part of a large amount of a wind farm's Levelized Cost of Electricity (LCOE). Thus, the reduction of O&M costs provides possibilities to control the LCOE in an effective way. Therefore, the optimization of operation and maintenance are critical factors in controlling the LCOE.

Consequently, one of the few measures of reducing the cost of production is to cut the cost of operation, including maintenance. It is the second-highest or even the highest element in operating expenses in some industries. Swanson [1] mentions that companies have started to undertake more efforts to improve quality and productivity as well as reduce costs to achieve world-class performance. This has led to examining the activities of the maintenance function. The author mentions that effective maintenance is crucial for many operations since it extends the life of the equipment, improves availability, and conserves/retains the equipment in appropriate conditions. Thus, if a company uses maintenance properly, it increases its production and revenue by increasing its availability [2]. This is so because the availability change allows the company to vary its production level and output. This, in turn, influences sales revenue and production costs. At the same time, the maintenance costs are affected since unplanned and preventive maintenance varies. In addition, the positive impact of the digital transformation efforts and their respective ICTs on maintenance and, thereby, productivity has been realized by academia and industry alike [3].

Additionally, it's well known that condition-based maintenance (CBM) preventive maintenance strategy is preferred wherever it is technically feasible and financially viable. The heart of CBM is condition monitoring (CM), which in principle, involves data acquisition, processing, analysis, interpretation, and extracting useful information. The information helps to identify whether asset health has deviated from the normal. If so, then fault diagnosis and prognosis usually follow. Finally, a decision is taken regarding when and what maintenance tasks are to be performed.

The CBM is now the most widely employed strategy in the Wind Farm industry [4–7]. Condition monitoring and maintenance can be performed with, for instance, vibration analysis, acoustic emission, sensory signals and signal processing methods, statistical methods, trend analysis, time-domain analysis, fast-Fourier transform (FFT), wavelet transforms, fault tree analysis (FTA) [8-9].

When it comes to renewable energy sources, such as wind farms, the efficiency improvement in energy production goes hand in hand with the possibility of providing predictive maintenance to the different energy production equipment. For instance, to provide service to various equipment, the level of energy production is stable and can utilize its full potential without fluctuations due to equipment breakdowns. Although, a disadvantage of using wind farms is that they do not always produce energy. It requires storing energy when production diminishes because of changed weather conditions. Therefore, battery storage systems offer the possibility of having a more stable output of power to match the required demand. Obtaining energy while minimizing pollution/contamination and costs is worrying many due to climate change and global warming. In this aspect, renewable energy is an excellent alternative. Continuously, the fluctuations in energy production as a result of different weather conditions have led to the option of moving the wind turbines offshore since the conditions are better suited, i.e., higher and stable wind speeds are experienced [10]. The importance of renewable energy in the future is a fact; for instance, Balischewski et al. [11] mention that renewable energy production in the case of wind power accounted for approximately 12% of Germany's total power production in 2015. In addition, in 2019, wind energy generated enough electricity to cover 15% of Europe's power demand (windeurope.org). Thus, the wind-energy sector has grown significantly among renewable energy sources in the last two decades. The most common fault in wind turbines is linked to gearbox failures, though the bearing errors are overrepresented [12].

Furthermore, it is well known that difficulties arise when implementing the entire cycle of condition-based maintenance (CBM). It is connected with specific issues regarding CBM characteristics, such as the amount of data produced, data integrations, and the lack of experts. Consequently, these issues have been tried to be solved over time, namely with the support of various ICTs, such as expert systems, decision support systems, artificial intelligence, and distributed intelligence [13].

Consequently, the main contributions of this paper are connected with the use of machine learning (ML) algorithms supported with MLOps, e.g., machine learning in operations. Its contributions are the following:

•A Machine Learning (ML) pipeline that uses training data from wind turbines. This design identifies patterns and creates a model that can handle new data without breaking.

• A new ML model for predictive maintenance for wind turbine bearings. This model monitors the condition of wind turbine bearings using an adaptive polynomial model to approximate time series data. The data is split into two: level and trend. For forecasting, it considers past data trends. We also employed the Holt-Winters Exponential Smoothing method for more precise forecasting. The high R-Squared values of proposed model suggest it effectively predicts new data.

• In addition, two ML stacking models were developed. One combines regression and k-nearest neighbors (knn) with a regression tree, while the other combines regression with support vector regression (svr) and a regression tree. These combinations reduce predictor correlation and enhance model generalization. Notably, the second model has a low Root Mean Square Error, indicating minimized errors in predictions.

Overall, the approach described in this article is comprehensive and provides an effective solution for wind turbine-bearing data analysis and forecasting. The research methodology is built as follows:

1. Dataset preprocessing.
2. Models' development.
3. Results evaluation.
4. System architecture development.
5. System development and testing.

The contribution mentioned above is presented further in section 4 of the paper. Accordingly, section 2 highlights some aspects of the domain of interest. Then, later in section 3, the machine learning analytical capabilities are explained, and the term MLOps, i.e., machine learning in operations, is presented; hence, the section discusses what would be needed to achieve analytical capabilities supported by the MLOps in the field. Further on, in section 4, a case example using Wind turbine failure data is provided, namely for a bearing fault considering the essential aspects of implementing MLOps in the domain. Finally, in section 5, the discussions and conclusions are given.

2. Preventive Maintenance of Wind Turbines

Despite the widespread adoption of wind turbines, the effective prediction of bearing failures remains a significant challenge, leading to unplanned downtimes and reduced turbine availability. Therefore, operation and maintenance, and specially condition-based maintenance, become key aspects to consider in this context.

Thus, one of the most common faults in wind turbines is connected with gearbox failures, but where the bearing errors are overrepresented [12]. It is known that the main bearing failures are vital factors to consider when it comes to increasing wind farms' reliability and availability [14]. Furthermore, it is one of the critical reasons that condition monitoring on the bearings is crucial. Hence, the failures of rolling element bearings (REBs) considerably influence the entire machinery; therefore, it is of primary importance to perform effective condition monitoring of REBs not to incur lost production time and economic losses. Moreover, it is crucial to be able to decide when to perform maintenance actions to the equipment based on condition monitoring and its connected CBM approach. In this respect, diminishing energy costs, especially those connected with unplanned stoppages, faults, support delays, etc., is crucial.

Consequently, all the former aspects are related to operation and maintenance practices, with considerable potential for cost reductions and innovations. Therefore, it is essential to consider the operation and maintenance of the wind industry, especially when it is believed to drastically cover a significant part of the E.U.'s total electricity consumption. Further, the domain's state of the art and challenges and its wind turbine bearing condition monitoring methods can be found in [15]. In addition, a review of the use of machine learning methods for wind farm turbines is also highlighted in [16]. The remaining useful life (RUL) is the length of time a machine is likely to operate before it requires repair or replacement. By taking RUL into account, engineers can schedule maintenance, optimize operating efficiency, and avoid unplanned downtime. For this reason, estimating RUL is a top priority in predictive maintenance programs. The paper calculates RUL based on methodology from [17]. In this case, the availability of wind farms, as mentioned above, is vital to produce energy. Their availability can be seen as technical availability, which involves the percentage of time that a wind turbine/wind farm is available to generate energy. It is stated as the percentage of the theoretical maximum of its availability [18, 19]. The equation below is from the IEC Standard 61400-26, highlighting the percentage of the theoretical maximum, as mentioned above. This formula is widely used in the wind industry and indicates the operational performance of wind turbines/wind farms.

$$\text{Availability} = 1 - (\text{Unavailable Time})/(\text{Available Time} + \text{Unavailable Time})$$

Thus, low availability can be poor wind farm reliability performance or a below-standard maintenance action. Therefore, an effective operation and maintenance approach (O&M) is crucial to be able to make the conditions of optimal availability of the wind turbine, which at the same time can produce the expected amount of energy.

Consequently, the efficiency improvement in energy production goes in hand with the possibility of providing predictive maintenance to the different energy production equipment. For instance, to provide service to the various equipment by doing so is the level of energy production stable and can utilize its full potential without fluctuations because of equipment breakdowns. It is, therefore, crucial to optimize the O&M. Thus, preventive maintenance based on the CBM approach provides an understanding of when it is suitable to maintain the equipment to avoid unplanned stoppages; in this case, the energy production and try to keep the wind turbines available so they can be operational and receive service when it is needed. It is, therefore, key to understand the various offshore wind turbine operation and maintenance approaches and how they might impact O&M [20]. Consequently, it is essential to find tools, such as machine learning algorithms supported with Mlops, to support the preventive maintenance of wind turbines..

3. Machine Learning Analytical Capabilities via Mlops

In the application of machine learning methodologies, it is imperative to source appropriate datasets and subsequently undergo rigorous data preprocessing. This involves the acquisition of relevant data, subsequent cleansing procedures, and preparation for analytical diagnosis. Periodic retraining of models is essential to ensure their robustness and accuracy. Furthermore, for optimal performance, it's paramount that models autonomously adapt their parameters based on data-driven insights, especially in response to dynamic changes in the input dataset [21,22]. This self-adjustment capability ensures the model's relevance and precision across varying data landscapes, which are all key aspects. ML is used where it can be applied to learn by comparing and correlating numerous similar patterns from various data sources to develop models to understand and foresee different things of interest [23-24]. ML learning learns from experience based on big data or specific datasets. It can detect different patterns in the labeled and not labeled data, i.e., supervised and unsupervised learning, and from where the results are not known [25]. The role of the ML in industrial asset management is to provide further information about when to provide services to the production equipment to keep production running and avoid unplanned stoppages. The increase in the use of A.I. and ML depends on, for instance, emergent technologies, such as new sensor technologies, the Internet of Things (IoT), and big data, which goes hand in hand while intending to reach industry 4.0 and all that it involves in the digitization process. The characteristics of these technologies are that they increase the amount of data that is created, gathered, and possible to analyze. Several different machine-learning algorithms have been tested and suggested [23-25]. However, in the domain of industrial asset management, i.e., condition monitoring and maintenance and its related CBM approach, the specific tools and techniques have been successfully used for several years. Hence, they are slow and rather conservative and belong to a vertical market, e.g., to acquire/adopt new emerging technologies. However, the domain still leans on its well-accepted signal processing technologies, among others, for the CBM approach. Therefore, it is essential to understand if those suggested ML algorithms have a place in a running CBM strategy, i.e., if they are suitable for operation, namely the so-called MLOps (Machine learning in operations).

Nevertheless, there are efforts to use machine learning methods for wind farms turbine, especially for blade fault detection or generator temperature monitoring [16]. The use of machine learning for wind turbine bearing fault detection in [26, 27] is reported. However, the implementation and use of the MLOps approach are not considered. Thus, these new ICTs, such as the IoT, big data, and machine learning, involve substantial modifications in several aspects for their successful implementation and use, i.e., how the maintenance department creates, uses, and manages their different Information Systems (I.S) & digital capabilities. The current work considers the use of the bearings faults as a case to highlight the use of ML in the domain of interest. MLOps come into the picture to provide organizations with the capabilities embedded in those technologies, in this case, machine learning into MLOps for industrial asset management. The term MLOps was coined by Sculley et al. [28] in the article titled "Hidden technical debt in machine learning systems." In this case, the needs of MLOps emerge, i.e., machine learning in operations, e.g., a term that is derived from the DevOps for machine learning. It highlights critical issues in machine learning and the gap from the pilot development into the actual deployment, which needs to consider many more aspects. This is highlighted in Figure 1 below. As seen in Figure 1, there is a need to cover several elements to implement ML successfully into operations in the domain.

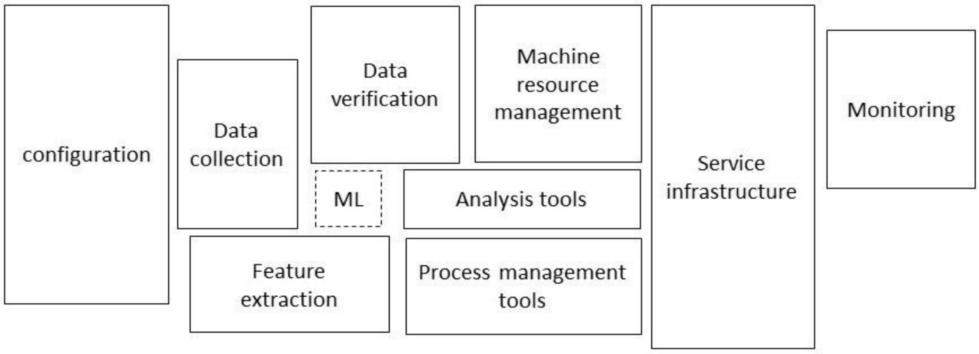

Figure 1: Highlights the small fraction of real-world machine learning systems composed of ML codes, as seen by the dotted box in the middle (modified from [28]).

At the same time, the prerequisite surrounding infrastructure is vast and multifaceted. The different business needs to organize continuous cooperation and interaction between all participants in the processes of working with machine learning models, from business to engineers and Big Data developers, including Data Scientists and ML specialists. MLOps is important for industries with needs of streaming data processing, such as wind turbines remaining useful life forecasting. Thus, MLOps provides organizational and business capabilities, i.e., automation, engagement, insight/decision-making, and innovation [25]. Hence, one of the main challenges in connection with Artificial Intelligence and machine learning, in this case, is that many of the systems are in the experimental phases, and few of them are deployed in production because of their complexities. Furthermore, deployment entails multiple factors, such as data and system integration with existing technologies, architectures, and legacy infrastructure. In addition, modification of business processes and the organizational culture, adequate employee skills, data engineering, organizational change management, etc. [29]. Consequently, as a result, the total production deployment is a lengthier process than pilot projects and has higher costs. It is, therefore, crucial to have a clear strategy, vision, and purpose for taking advantage of the inherent organizational intellectual skills and material resources [29]. Thus, digital transformation and its solutions must be combined with people and a smart approach to successfully digitalizing the area of interest. Hence, the focus should be on optimizing maintenance throughout the asset life and based on the operations' needs with suitable digital solutions.

IV. Creating analytics capabilities for wind farms supported by the MLOps approach

This section outlines the process of implementing MLOps for wind turbines equipped with sensors on their bearings. Initially, the MLOps framework facilitates the integration of resources from diverse origins. Additionally, the selection of an appropriate model may vary based on the source of data. A schematic representation of the MLOps pipeline is provided in Figure 2.

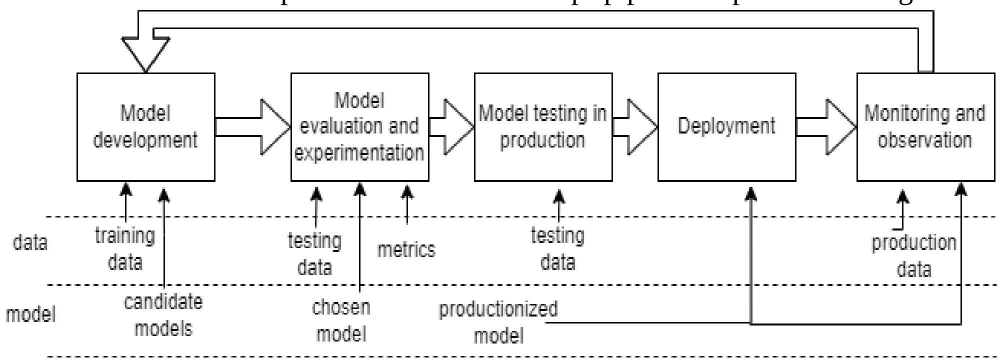

Figure. 2: The general scheme of the MLOps pipeline.

The reason for monitoring the ML model is to understand how it solves the business problem. Concerning wind farms, data quality is crucial. Often this data is presented as a time series. In addition, seasonality may be different for different data sources because wind farms are situated in different places with various conditions. Despite its "flexibility" in finding relationships in large datasets, the ML model has many vulnerabilities.

Therefore, effective monitoring of machine learning models is essential for several reasons. First, the quality and structure of input data play a pivotal role in the accuracy of the model. Second, as models evolve, their performance can degrade over time, necessitating consistent evaluations. Third, there are challenges related to interdependent models and unique pipeline configurations. Fourth, there may be instances of abnormal values or predictions that the model has not previously encountered, emphasizing the need for robust outlier detection mechanisms. Additionally, understanding the model's inner workings and decision-making process is vital, particularly in contexts where interpretability is crucial. In many scenarios, there's ambiguity regarding the true values in queries a priori, leading to uncertainty about the precise class or cluster to which a particular component belongs. Furthermore, the time required for model computations, potential unavailability of deployment endpoints, alterations in the application's business logic, susceptibility to cyberattacks, and potential data losses all underscore the importance of rigorous monitoring. Such considerations are fundamental in ensuring the model's accuracy and reliability throughout its lifecycle.

In the present study, we analyzed a dataset obtained from a 2MW wind turbine's high-speed shaft driven by a 20-tooth pinion gear [30]. Vibration signals, each lasting 6 seconds, were captured daily over a span of 50 consecutive days. Notably, on March 17, two measurements were taken and are considered as separate days for this analysis. Over these 50 days, an inner race fault emerged, leading to the bearing's failure. In its compact form, the dataset has a measurement time step of 5 days. The Remaining Useful Life (RUL) computed using the methodology from [17] is utilized as a time series. Various models have been employed for data analysis, and a novel ensemble has been explicitly introduced for prognostics and RUL forecasting of wind turbine high-speed bearings. The structure of this research is outlined as follows:

1. Time Series Analysis: Classic time series models are employed for a primary data investigation.
2. Data Interpretation Algorithm Development: A novel Data Interpretation Algorithm for Forecasting Time Series is crafted based on the modified adaptive mono-parameter Braun model. Data is segmented into levels and trends. The Holt-Winters Exponential Smoothing method, tailored for time series data with both trends and seasonal variations, is employed. This method facilitates forecasting by leveraging previously observed weighted changes.
3. Predictive Modeling: Classical predictive machine learning models are utilized alongside the development of a new ensemble schema. The innovative stacking model we developed suggests deforming meta-features based on pairwise multiplication results. These are then integrated with the training dataset in a meta-model.
4. Result Analysis and MLOps Architecture Development: A comparative analysis of the results is conducted, leading to the formulation of an MLOps architectural framework.

Initially, classical time series models are applied. The polynomial trend model essentially functions as a multiple regression equation, making the methods and procedures of regression analysis, as discussed in the initial segment of this publication, largely relevant for its delineation. The AutoRegressive Integrated Moving Average (ARIMA) constructed time series is illustrated in Fig. 3. It begins with a visualization of the vibration signals in the time domain. Forecasts are generated for the subsequent five values (depicted by the red line). This visualization offers forecasted data without any smoothing.

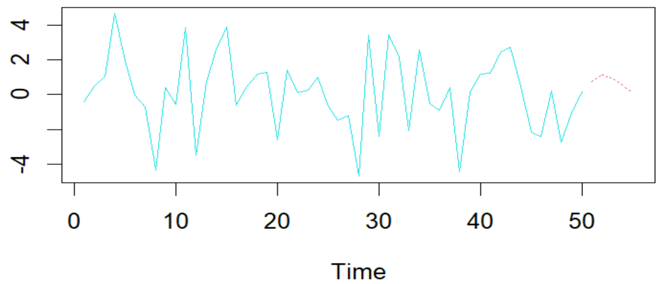

Figure. 3: The diagram of ARIMA

For time-series forecasts, Root Mean Squared Error (RMSE) measure is used. It is equal to 17.93.

Next, based on vibration level, RUL is calculated.

To decrease RMSE, smoothing is proposed. The adaptive mono-parameter Braun model is used for stationary time series based on simple exponential smoothing:

$$\hat{y}_{(t+1)}=S_t, S_t=\alpha y_t+(1-\alpha) S_{(t-1)}, t=1,2,3,\ldots, \quad (1)$$

where $\hat{y}_{(t+1)}$ is the prognostic value of time series level in time (t+1), S_t is the exponential mean, α is the adaptation coefficient, and y_t is the current time series value.

In this context, the model's value is derived from a weighted average of the current actual value and preceding model values. This weight is commonly referred to as the smoothing factor. It dictates the rate at which the most recent observable data point diminishes in influence. A lower weight ensures that prior model values exert a stronger influence, leading to a smoother data series. Taking the adaptation coefficient α and the warning period τ, it is necessary to approximate the series using an adaptive polynomial model.

The novel method for predictive maintenance of wind turbine bearing is developed in the paper.

This method consists of the following steps:

1. Zero-order time series analysis (p = 0);
2. First order time series analysis (p = 1);
3. Assess the accuracy and quality of forecasts;
4. Make a forecast.

The first two steps of proposed method are presented below.

Step 1.

The procedure for Step 1 was developed as a sequence of the following steps:

Let $\hat{y}_0= y_0$.

Append array $\hat{y}$ using the following formula:

$$\hat{y}_{(t)}= \alpha * y_t+(1-\alpha) * \hat{y}_{(t-1)} , \quad (2)$$

where $\hat{y}_{(t)}$ is an actual value, and $\hat{y}_{(t-1)}$ is the previous number from the prediction array.

Repeat step 2 for all values in the dataset.

Step 2.

1. Let x=1, $\hat{y}_0= y_0$, $l_0= y_0$, and .$b_0= y_1-y_0$, where y is our initial dataset.
2. Define new level value using the formula: $l_x= \alpha y_x+(1- \alpha)(l_{(x-1)}+b_{(x-1)})$.
3. Define new trend value: $b_x=\beta(l_x-l_{(x-1)})+(1-\beta)b_{(x-1)}$.
4. Define our prediction $\hat{y}_{(x+1)}=l_x+b_x$.
5. Define x=x+1 and repeat steps 2-5 until x<n.

The results of proposed method are given in Table 1.

The performance of the time series models on the presented testing data is given below:

1. Exponential Smoothing: MAPE is appr. 8.5 %
2. Proposed method: MAPE is appr. 2.2 %
3. Holt's Trend Method: MAPE is appr. 6.6 %
4. ARIMA: MAPE is appr. 3.1 %
5. TBATS: MAPE is appr. 3.2 %

The Exponential Smoothing model did well by achieving a lower MAPE of 8.5 percent. All the other models outperformed them by producing lower MAPE. However, the DIAFS model emerged as the winner based on its test data with MAPE performance, which was close to 2.2 %.

Next, machine learning predictive models are used for data analysis.

First of all, kurtosis was calculated. The kurtosis measure describes the tail of distribution – how similar are the outlying values of the distribution to the standard normal distribution? For example, the standard normal distribution has a kurtosis of 0.

Stochastic variables and numerous uncertainties can substantially complicate the problem. To circumvent these challenges, various ensembles are proposed. An ensemble method in both statistics and machine learning leverages multiple trained algorithms to achieve superior predictive performance than what could be attained by any single algorithm alone. In contrast to a statistical ensemble, a machine learning ensemble encompasses a distinct finite set of alternative models but typically allows for much more flexible structures. The central premise is to employ fundamentally diverse models to enhance the capability of processing unfamiliar data. The results derived from regression, KNN, ANN models, and the proposed ensemble of predictive models were subsequently compared.

Bagging is an ensemble technique where models are trained in parallel on different random subsets of the training data. The final decision is derived from the majority voting of the ensemble classifiers, selecting the class predicted by the majority.

Boosting involves training an ensemble of models sequentially, where each subsequent model focuses on instances that the preceding classifier misclassifies. While boosting typically yields more accurate results than bagging, it can be susceptible to overfitting.

Stacking involves partitioning the training set into N blocks. N-1 blocks are used to train a set of base models, while the Nth block, paired with outputs from the base classifiers (referred to as meta-features), trains another model. One limitation of the traditional stacking approach is the disparity between the meta-features in the training sample and the actual responses from specific regressors. In classical stacking, non-overlapping unique values may exist between training and testing meta-features. Our developed stacking model seeks to address this by deforming meta-features based on pairwise multiplication outcomes. These transformed features, combined with the training dataset, feed into a meta-model, mitigating weak predictor result correlations and enhancing model generalization.

Regression analysis utilized three condition indicators: RMS, Kurtosis, and E.I. Performance of each regression model was assessed using RMSE, R^2, and adjusted-R^2. Among the models tested, SVR, polynomial regression, and a single-hidden layer ANN with 12 neurons emerged as the superior weak predictors. Conversely, the tuned KNN model underperformed, as shown in Table 1.

Table 1.

Models' Performance Metrics.

Model/Metrics	RMSE	R2	Adj.R2
Polynomial regression	0.164	0.595	05.77
ANN	0.217	0.881	0.875
Knn	13.94	0.945	0.942
SVR	0.203	0.884	0.880

Proposed model	0.212	0.954	0.944
Bagged regression	0.321	0.942	0.893
Stacking (regression and knn +regression tree as metaregressor	0.932	0.921	0.921
Stacking (regression and knn +random forest as metaregressor	0.157	0.744	0.742

The results derived from our proposed ANN model showcase its capability to predict the remaining useful life of a bearing, a feat attributed to the synergy between the regression and ANN models through the optimal condition indicator. The DIAFS model also exhibits promising results, boasting the highest R-Squared and Adjusted R-Squared values, suggesting its potential in forecasting novel observations. Although the ANN model provides accurate predictions and is corroborated by other research [12, 16], its performance remains reliant on the regression model. This dependency underlines why ensemble models typically surpass standalone ANN models. The second stacking model boasts the lowest RMSE, signifying minimal residuals. However, its R^2 value doesn't mirror this superiority, underscoring the necessity of the regression model in optimizing the ANN's predictive efficacy.

In the implemented system, forecasts are pre-calculated using a pre-trained ML model for incoming data and stored in a dedicated database. Subsequent input requests access this database for predictions. Architecturally, this mirrors the Lambda pattern, blending "hot" real-time data processing with "cold" historical data processing, typically residing in a Data Lake on Apache Hadoop (Fig. 4).

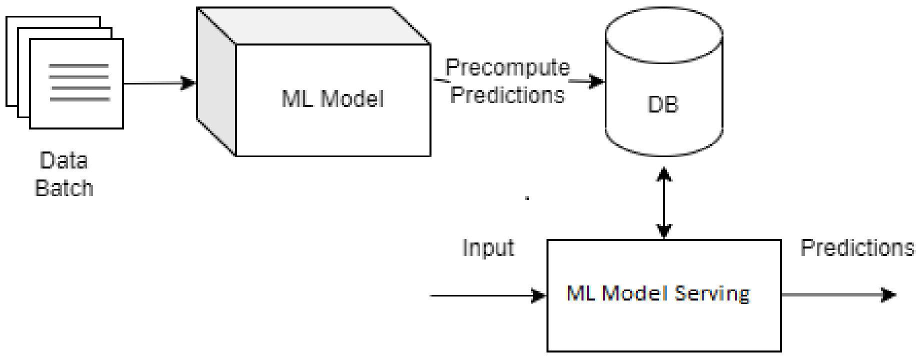

Figure 4: Lambda-architecture for ML-system in MLOps.

The Lambda Architecture encompasses both a conventional batch data pipeline for fast-streaming real-time data and a serving layer designated for query responses. Ingressed batch data populates a batch layer, prepping it for indexing. Several pre-trained ML models then analyze this data concurrently. The models listed in Table 1 are incorporated. The serving layer facilitates the pre-result voting process, selecting the best model or a suitable combination for the given data. The selected model subsequently processes incoming stream data.

5. Conclusions

This research introduces a predictive model rooted in MLOps methodology to determine wind turbine bearings' Remaining Useful Life (RUL). The model efficiently detects degradation patterns in real time, adjusting its parameters in response to new data, and offers a fully automated configuration, enabling its deployment across multiple wind turbines. As such, it emerges as an indispensable tool for condition-based maintenance.

Our ensemble stacking model, incorporating regression, SVR, and random forest techniques, has demonstrated commendable generalization capabilities. The model's effectiveness is underscored by its proficiency in leveraging condition indicators, which were appraised using criteria like monotonicity and trendability.

The incorporation of MLOps in this research has facilitated the following: Enhanced innovation through holistic machine learning lifecycle management; Reproducible and robust model iterations tailored for enterprise settings; Efficient tracking using advanced dataset and model registries; Improved traceability and accountability through detailed logging; Streamlined model workflows ensuring consistent delivery; Generation of unbiased models emphasizing feature importance, assessed using uniform distribution metrics.

Our research further proposes new model for monitoring wind turbine bearing conditions. DIAFS's adaptability, anchored in an adaptive polynomial model, facilitates real-time refinement, promising accurate and timely predictions. This algorithm not only refines the precision of forecasting but also augments maintenance strategies. Through DIAFS, potential issues can be preemptively identified, optimizing resource allocation and reducing operational downtimes. Such a data-driven approach, focusing on empirical evidence, leads to significant cost savings, increasing wind turbines' overall efficiency and lifespan. Furthermore, the scalability inherent to DIAFS ensures its applicability within the expanding world of wind energy.

While the primary application of the ensemble model is in wind turbines, its adaptable architecture ensures relevance across varied industrial maintenance scenarios, such as managing Rolling Element Bearings (REBs) failures. Beyond the technical abilities mentioned above as a digital transformation driver, this study recognizes the pivotal role of workforce adaptation to these evolving digital tools.

Lastly, future research will focus on adapting the developed models to analyze raw data in real time. Additionally, it would be interesting to extend the MLOps pipeline to facilitate real-time monitoring of wind turbines, thereby enabling faster fault detection and the potential implementation of predictive maintenance strategies.

Acknowledgments

The paper was prepared as part of H2020 project ZEBAI, # 101138678

References

[1] Swanson, L. Linking maintenance strategies to performance. International journal of production economics, 2001, 70(3), 237-244.

[2] Rishel, Tracy D.; Canel, Cem. Using a maintenance contribution model to predict the impact of maintenance on profitability. Journal of Information and Optimization Sciences, 2006, 27.1: 21-34.

[3] Hofmann, M.; Seprstad, I.V. NOWIcob–A tool for reducing the maintenance costs of offshore wind farms. Energy Procedia, 2013, 35: 177-186.

[4] Fox, H., Pillai, A. C., Friedrich, D., Collu, M., Dawood, T., & Johanning, L. A review of predictive and prescriptive offshore wind farm operation and maintenance. Energies, 2022, 15(2), 504.

[5] Li, J., Zhang, X., Zhou, X., & Lu, L. Reliability assessment of wind turbine bearing based on the degradation-Hidden-Markov model. Renewable energy, 2019, 132, 1076-1087.

[6] Walker, C., & Coble, J. Wind turbine bearing fault detection using adaptive resampling and order tracking. International Journal of Prognostics and Health Management2018, 9(2).

[7] Lu, Y.; Sun, L., Zhang, X., Feng, F., Kang, J., & Fu, G. Condition based maintenance optimization for offshore wind turbine considering opportunities based on neural network approach. Applied Ocean Research, 2018, 74, 69-79.

[8] Márquez, F. P. G.; Tobias, A. M., Pérez, J. M. P., & Papaelias, M. Condition monitoring of wind turbines: Techniques and methods. Renewable energy, 2012, 46: 169-178.

[9] Kabir, M. J.; Oo, A. M., & Rabbani, M. A brief review on offshore wind turbine fault detection and recent development in Condition monitoring based maintenance system. In: 2015 Australasian Universities Power Engineering Conference (AUPEC). IEEE, 2015. p. 1-7.

[10] Shen, W.; Chen, X.; Qiu, J.; Hayward, J.A.; Sayeef, S.; Osman, P.; Meng, K.; Dong, Z.Y. A comprehensive review of variable renewable energy levelized cost of electricity. Renew. Sustain. Energy Rev. 2020, 133, 110301.

[11] Balischewski, S.; Hauer, I., Wolter, M., Wenge, C., Lombardi, P., & Komarnicki, P. Battery storage services that minimize wind farm operating costs: A case study. In: 2017 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT-Europe). IEEE, 2017. p. 1-6.

[12] Strömbergsson, D.; Marklund, P., & Berglund, K. Increasing Wind Turbine Drivetrain Bearing Vibration Monitoring De-tectability Using an Artificial Neural Network Implementation. Applied Sciences, 2021, 11(8), 3588.

[13] Campos, J. Development in the application of ICT in condition monitoring and maintenance. Computers in industry, 2009, 60(1), 1-20.

[14] Hart, E.; Clarke, B.; Nicholas, G.; Kazemi Amiri, A.; Stirling, J.; Carroll, J.; Dwyer-Joyce, R.; McDonald, A.; Long, H. A review of wind turbine main bearings: design, operation, modelling, damage mechanisms and fault detection. Wind. Energy Sci. 2020, 5, 105–124.

[15] Liu, Z.; L. A, Zhang. Review of failure modes, condition monitoring and fault diagnosis methods for large-scale wind turbine bearings. Measurement, 2020, 149, 107002.

[16] Stetco, A., Dinmohammadi, F., Zhao, X., Robu, V., Flynn, D., Barnes, M., ... & Nenadic, G. Machine learning methods for wind turbine condition monitoring: A review. Renewable energy, 2019, 133, 620-635.

[17] Pagitsch, M., Jacobs, G., & Bosse, D. Remaining Useful Life Determination for Wind Turbines. In Journal of Physics: Conference Series, 2020, 1452 (1), 012052.

[18] Dinwoodie, Iain, et al. "Reference cases for verification of operation and maintenance simulation models for offshore wind farms." Wind Engineering 39.1 (2015): 1-14.

[19] Tavner, Peter. Offshore wind turbines: reliability, availability and maintenance. Vol. 13. IET, 2012.

[20] Ren, Zhengru, et al. Offshore wind turbine operations and maintenance: A state-of-the-art review. Renewable and Sustainable Energy Reviews, 2021, 144: 110886.

[21] Oluyisola, O. E., Sgarbossa, F., & Strandhagen, J. O. Smart production planning and control: Concept use-cases and sustaina-bility implications. Sustainability, 2020, 12, 3791.

[22] Rojas, R. A., & Garcia, M. A. R. (2020). Implementation of industrial internet of things and cyber-physical systems in smes for distributed and service-oriented control industry 4.0 for smes. Palgrave Macmillan.

[23] Sarker, I. H. Machine learning: Algorithms, real-world applications and research directions. S.N. Computer Science, 2021, 2(3), 1-21.

[24] Almounajjed, A., Sahoo, A. K., Kumar, M. K., & Alsebai, M. D. Investigation Techniques for Rolling Bearing Fault Diagnosis Using Machine Learning Algorithms. In 2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS) (2021, pp. 1290-1294). IEEE.

[25] Benbya, H.; Pachidi, S., & Jarvenpaa, S. (2021). Special issue editorial: Artificial intelligence in organizations: Implications for information systems research. Journal of the Association for Information Systems, 2021, 22(2), 10. Available at: https://aisel.aisnet.org/misqe/vol19/iss4/4

[26] Purarjomandlangrudi, A., Nourbakhsh, G., Ghaemmaghami, H., & Tan, A. (2014, October). Application of anomaly technique in wind turbine bearing fault detection. In IECON 2014-40th Annual Conference of the IEEE Industrial Electronics Society (pp. 1984-1988). IEEE.

[27] Ali, J. B., Saidi, L., Harrath, S., Bechhoefer, E., & Benbouzid, M. (2018). Online automatic diagnosis of wind turbine bearings progressive degradations under real experimental conditions based on unsupervised machine learning. Applied Acoustics, 132, 167-181.

[28] Sculley, D.; Holt, G., Golovin, D., Davydov, E., Phillips, T., Ebner, D. & Dennison, D. Hidden technical debt in machine learning systems. Advances in neural information processing systems, 2015, 28, 2503-2511.

[29] Kans, M.; Campos, J., Salonen, A., & Bengtsson, M. The thinking industry: an approach for gaining highest advantage of digitalization within maintenance. Journal of Maintenance Engineering, 2017, 2, 147-158.

[30] Bechhoefer, Eric, Brandon Van Hecke, and David He. "Processing for improved spectral analysis." Annual Conference of the Prognostics and Health Management Society, New Orleans, LA, Oct. 2013.