applied sciences

MDPI

Review

Development and Integration of Metocean Data Interoperability for Intelligent Operations and Automation Using Machine Learning: A Review

Kamaluddeen Usman Danyaro [1,*], Haizatul Hafizah Hussain [2], Mujaheed Abdullahi [1], M. S. Liew [3], Lim Eu Shawn [3] and Mustapha Yusuf Abubakar [4]

[1] Computer and Information Science Department, Universiti Teknologi PETRONAS, Seri Iskandar 32610, Perak, Malaysia; abdullahi_18001208@utp.edu.my
[2] Petroleum Engineering Department, Universiti Teknologi PETRONAS, Seri Iskandar 32610, Perak, Malaysia; haizatulhafizah.hus@utp.edu.my
[3] Civil and Environmental Engineering Department, Universiti Teknologi PETRONAS, Seri Iskandar 32610, Perak, Malaysia; shahir_liew@utp.edu.my (M.S.L.); shawn.lim@utp.edu.my (L.E.S.)
[4] Computer Science Department, School of Technology, Kano State Polytechnic, Kano 700231, Nigeria; mygawuna@kanopoly.edu.ng
* Correspondence: kamaluddeen.usman@utp.edu.my

Abstract: The current oil industry is moving towards digitalization, which is a good opportunity that will bring value to all its stakeholders. The digitalization of oil and gas discovery, which are production-based industries, is driven by enabling technologies which include machine learning (ML) and big data analytics. However, the existing Metocean system generates data manually using sensors such as the wave buoy, anemometer, and acoustic doppler current profiler (ADCP). Additionally, these data which appear in ASCII format to the Metocean system are also manual and silos. This slows down provisioning, while the monitoring element of the Metocean data path is partial. In this paper, we demonstrate the capabilities of ML for the development of Metocean data integration interoperability based on intelligent operations and automation. A comprehensive review of several research studies, which explore the needs of ML in oil and gas industries by investigating the in-depth integration of Metocean data interoperability for intelligent operations and automation using an ML-based approach, is presented. A new model integrated with the existing Metocean data system using ML algorithms to monitor and interoperate with maximum performance is proposed. The study reveals that ML is one of the crucial and key enabling tools that the oil and gas industries are now focused on for implementing digital transformation, which allows the industry to automate, enhance production, and have less human capacity. Lastly, user recommendations for potential future investigations are offered.

Keywords: data interoperability; Metocean; oil and gas; artificial intelligence; machine learning; automation

check for updates

Citation: Danyaro, K.U.; Hussain, H.H.; Abdullahi, M.; Liew, M.S.; Shawn, L.E.; Abubakar, M.Y. Development and Integration of Metocean Data Interoperability for Intelligent Operations and Automation Using Machine Learning: A Review. *Appl. Sci.* **2022**, *12*, 5690. https://doi.org/10.3390/app12115690

Academic Editor: Emanuele Carpanzano

Received: 9 April 2022
Accepted: 28 May 2022
Published: 3 June 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

According to the World Economic Forum, the estimated net benefits of the oil and gas industry, which is due to digital transformation, will be USD 945 billion over one decade (cumulative value 2016–2025). In addition, digital transformation is not realistic without artificial intelligence (AI) and data science, especially in industries such as oil and gas, finance, and the internet [1]. Conversely, due to the current oil and gas crash, digitalization is a good opportunity that will add value to all stakeholders. These opportunities show that there is a need to continue in the digitalization of the oil and gas sector to achieve the expected or more of the stated value by the year 2025. Digitalization of production-based industries is driven by enabling technologies which include machine learning (ML) and big data analytics [2,3]. This drives the oil and gas sector to be among the technological

advancements and convergence of industry 4.0 in Malaysia [4,5]. Thus, the application of AI and machine learning is necessary for any digitalization and automation to succeed in various sectors such as healthcare, finance, industry automation, transportation, and cybersecurity [6,7].

Machine learning (ML) is one of the crucial and key enabling tools that oil and gas industries are now focusing on to implement digital transformation. The application of AI and ML allows an industry to automate and enhance production and human capacity. For example, ML is used to understand and predict behaviour using an ice database for experiment with enhanced prediction for oil production and recovery [8–17]. The Metocean study, conversely, is used by oil and gas, particularly offshore industries, to estimate the environmental conditions for better and successful operations, which include prediction of gas–liquid pattern flow, groundwater anomaly detection, and pipe incident detection [18–21]. However, the existing Metocean system generates data manually using sensors such as wave buoy, anemometer and acoustic doppler current profiler (ADCP). Additionally, improving the existing techniques for Metocean datasets has been provided [22]. The input of these data, which appear in ASCII format to the Metocean system is also manual and silos. There is no fundamental method, model, or algorithm that allows the system to reason and automate intelligently. This slows down provisioning, while the monitoring element of Metocean data path is partial. Therefore, a study and model that can integrate the Metocean data as well as provide better decision making and reduce manual tasks is needed.

The main aim of this paper is to provide an in-depth review of the method for developing and integrating oil and gas data, using a particular case study on the Metocean data system using a machine learning algorithm.

Integration of Metocean Forecast Data and Automation

Metocean data interoperability development with integration using automation measures parameters such as surge height and wave height. This can also be addressed by using ML model techniques for efficient results. Qiao and Myers [23] proposed surrogate time-independent modelling to evaluate Metocean conditions during hurricanes, which include peak wave period, peak wave direction, and storm surge using a Metocean database. The model was developed using a recurrent neural network (RNN), gated recurrent unit (GTU), and multilayer perception (MLP). Orlandi et al. [24] proposed a prototype system that is capable of integrating Metocean model forecast and performance of ship data for modelling. In addition, the prototype allows for visualization of Metocean data after uploading from the system for evaluating the condition on each route.

Several research studies have explored the needs of ML in the oil and gas industries. In this study, we investigate in depth the integration of Metocean data interoperability for intelligent operations and automation using an ML-based approach.

- A comparative analysis of the current uses of ML algorithms in the oil and gas industries is presented.
- The transformation and integration of existing Metocean architecture for the use of oil and gas data operators through automation is meticulously reviewed.
- A new model to be integrated with the existing Metocean data system using a machine-learning algorithm to monitor and interoperate with maximum performance is proposed.

2. Meta-Survey

This section presents several works related to the automation and digitalization using machine learning in the area of control systems and decision making. Many workflows in business organizations prove to be difficult and monotonous due to the increase in access to organizational data. This makes the cost of production high, which forces industries, business, and organizations to automate their functions to reduce the cost of production and boost system efficiency [25–27]. An organization must choose to digitalize and automate its tools; otherwise, it will be obsolete. Therefore, to realize the vision of digitalization

in oil and gas or "digital oilfield", technology-centric solutions such as big data, cloud computing, the internet of things (IoT), and many more need to be considered in automation and decision-making processes [26]. Adding to this, Maroufkhani et al. [28] present a comprehensive systematic review of digital transformation in the resource and energy industries domain. What manifests the digitalization, automation, automated analysis, and data integration in oil and gas is machine learning and AI in general [29–31]. Therefore, adding value to cost-saving, automated operations, improved monitoring, mitigation of environmental hazards and improved decision making will be realized.

Pashali et al. [32] conducted a comprehensive survey of Metocean and ice for the development of the Russian Arctic continental shelf. The survey determines Metocean data for exploratory drilling processes with design data for offshore oil and gas facilities through round operations in freezing seas. A similar study by Buzin et al. [33] gives a comprehensive review for the period between 2012 and 2021 based on Metocean shelf development projects and their approaches and results. The study also provided an overview of the Arctic and Antarctic Research Institute (AARI) task scope chronology survey. Wang et al. [34] conducted a comprehensive review of current field monitoring development for offshore structure monitoring, which includes Metocean sensing and structural motions. The study also provides state-of-the-art development in offshore structure monitoring.

Although, the application of ML techniques has been utilized in various domains including manufacturing sectors. For example, Nasiri and Khosravani [35] investigated the additive manufacturing (AM) parameters and prediction of mechanical behaviour of 3D components using ML technique. Consequently, the authors focused on prediction of ML applications for mechanical behaviour. However, current challenges have been provided which include lacking huge datasets from 3D printing that can lead to low accuracy results. More recently, Verma and Verma [36], surveyed ML applications in the healthcare sector, where they play a vital role in several areas such as healthcare data analytics and medical data protection. However, medical records and disease forecasts have been analysed using ML applications. The authors also provided a research gap for efficient use of ML algorithms in the healthcare sector with opportunities and challenges.

Moreover, a systematic review has been presented which shows the potential of ML applications in oil and gas industries [37]. However, several authors have attempted implementing ML classifiers in various domains for prediction and enhancement such as acid-fracturing, CO_2 sequestration, and rock brittleness [38–40].

Comparison details of other related studies in the same domain is presented in Table 1. The studies provide an in-depth analysis of Metocean, oil and gas, and ML areas.

Table 1. Comparison of other related studies in the same domain: (✓: Yes, ×: No).

Authors	Year	Review Study	Oil and Gas	Metocean Data	ML
Verma and Verma [36]	2022	✓	×	×	✓
Buzin et al. [33]	2021	✓	✓	✓	×
Pashali et al. [32]	2021	✓	✓	✓	×
Nasiri and Khosravani [35]	2021	✓	×	×	✓
Tran et al. [26]	2020	×	✓	×	×
Hajizadeh [41]	2019	×	✓	×	✓
Wang et al. [34]	2018	✓	✓	✓	
Kohli and Johnson [25]	2011	×	✓	×	×
Aissani et al. [29]	2009	×	✓	×	✓
Hanga and Kovalchuk [42]	2019	✓	✓	×	✓
Barbosa et al. [43]	2019	✓	✓	×	✓
Our study	2022	✓	✓	✓	✓

3. Review Methodology

In this section, the method of review employed is described. Kitchenham and Charters [44] standard guidelines were applied. The literature review was performed using a relevant database based on the search strategy developed to identify relevant studies. The review flowchart for the data collection process is depicted in Figure 1.

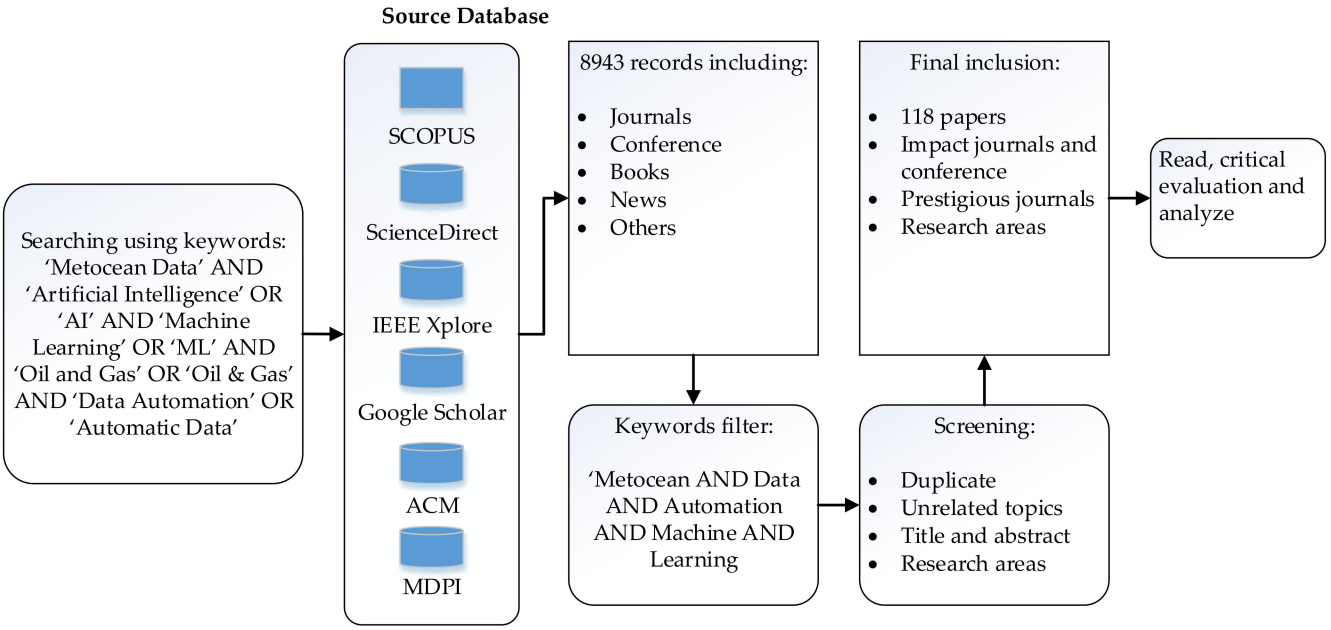


Figure 1. Flowchart for the data collection process.

Figure 2 demonstrates an overview structure of the review study which is categorized based on sections. The categorization approach provides a better understanding of the study.

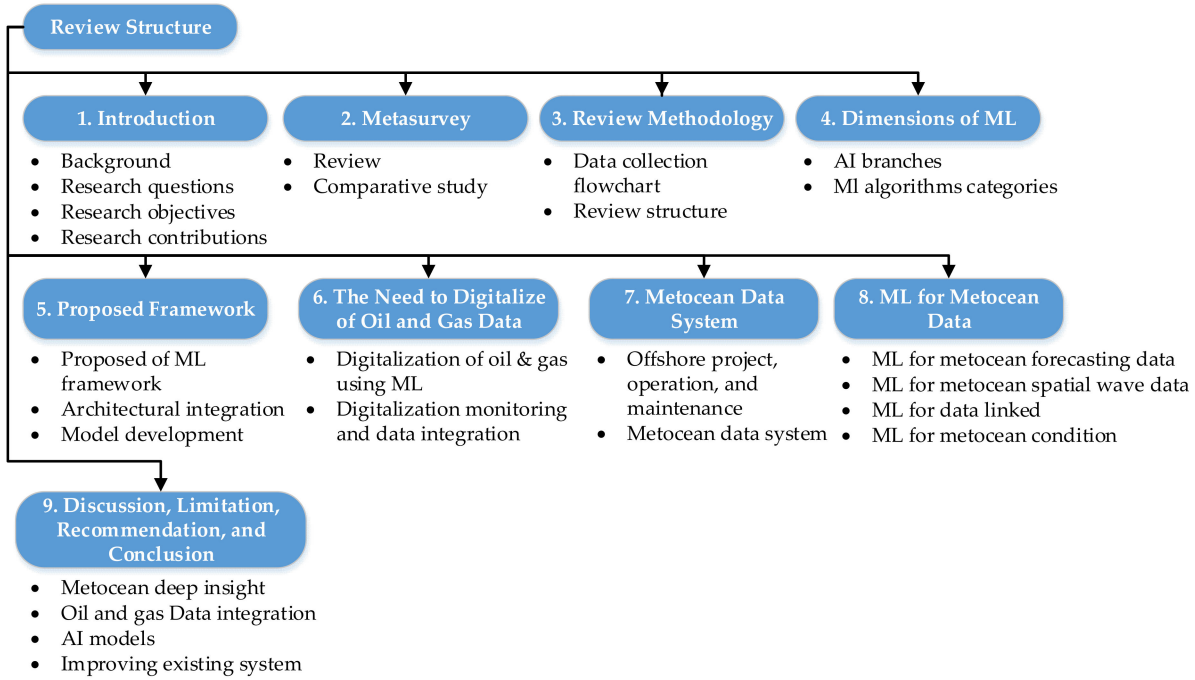


Figure 2. Structure of the review study.

4. Dimensions of Machine Learning Algorithms

In this section, the dimensions of ML are described. Machine learning (ML) is the study of mathematical models and algorithms that allow software applications to learn and improve automatically through experience. ML is a branch of artificial intelligence, where it deals with prediction or decision. ML algorithms use training data or mathematical model sample data without developing a new task. Additionally, machine learning focuses on data-driven modelling, which offers great value to oil and gas companies by digging out complex data [45], where structured or unstructured data may result in big data issues in oil and gas. Non-linearity and uncertainties make the oil and gas raw data actionable knowledge [45]. Moreover, ML algorithms have been used to develop a novel surrogate model using the random forest (RF) technique for estimating the simulation of the Simulating Waves Nearshore (SWAN) model [46].

The mathematical equation for ML is provided based on selected review studies. For example, k-means clustering approach is used for efficient dimensional reduction by minimizing the square error over all K clusters. With u^k as the centroid of the cluster, c_k is the main point of the cluster, and $|S_k|$ refers to the number of samples in the cluster c_k [47].

$$J = \sum_{k=1}^{K} \sum_{x^{(l)} \in c_k}^{M} ||x^{(l)} - \mu^k||^2, \; u^k = \frac{\sum^{x^{(l)} \in c_k} x^{(l)}}{|S_k|} \tag{1}$$

In addition, the silhouette index varies from -1 to 1; the value close to one refers to the appropriate data within its clusters. The silhouette (i) value can be determined using Equation (2).

$$s(i) = \frac{b(i) - a(i)}{max\{a(i), \; b(i)\}} \tag{2}$$

Another classifier (ANN) is among the most widely used ML algorithms due to its enhanced robustness. It also contains one input layer, one or more hidden layers, and one output layer. The signal process could be donated, as in Equation (3).

$$Y = G(\sum\nolimits_{i=1}^{m} w_i X_i + b) \tag{3}$$

To address the ML issue of negative samples from unlabelled datasets and classifiers, the training data is fit too tightly, which could lead to false results. The authors in [48] proposed MSE as the loss function L to overcome various penalty factors for positive and negative samples.

$$\begin{aligned} L(y, y', w, b) - 1/K \cdot (\sum\nolimits_{i=1}^{k} (yp\,(i) - y'p(i))^2 + \lambda 1 \cdot \sum\nolimits_{i=1}^{k} (yN(i) - y'N(i))^2 \\ + \lambda w \cdot \sum ||w||^2 + \lambda b \cdot \sum ||b||^2 \end{aligned} \tag{4}$$

ML can be generalized through training and the demonstration of models for data classification. In addition, normalization and standardization play an important role in ML modelling by rescaling values and data. Moreover, feature engineering improves ML prediction or unseen data after transforming raw data into features. However, to archive efficient integrity results, data quality assurance has to be considered. Figure 3 clearly provides an illustration of ML categorization based on a hierarchy flowchart.

4.1. Supervised Learning

Supervised learning algorithms are trained based on input data that have been labelled for output. They analyse the training data which can be used for new sample data [49]. The learning process can be categorized as classification and regression accordingly. The ML classification categorizes a given dataset into classes which include structured and unstruc-

tured data, where those classes can provide predication and detection after modelling. In Chen et al. [50], a new ML classification model based on a support vector machine (SVM) was proposed to calculate gas saturation in Shale reservoirs of the Chang 7 formation by comparing the result with laboratory-measured values. In addition, a hybrid ML model has been proposed which includes linear regression (LR) to predict and rate the efficiency of water flooding during oil production [51].

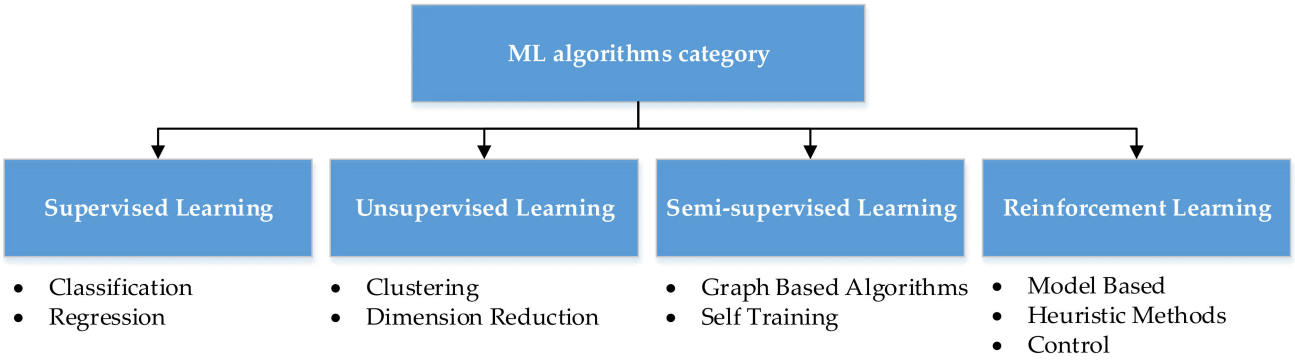


Figure 3. Illustration of ML categories.

4.2. Unsupervised Learning

Unsupervised learning algorithms learn from the availability of label class, where all the input samples are unlabelled. The learning process includes clustering and dimension reduction. For example, Liu et al. [52] proposed an ocean reconstruction front based on the K-means algorithm to archive hierarchical clustering sound speed profile (SSP). The study also provides a method to verify the feasibility method from the perspective of transmission loss (TL) calculations. Adding to this, a dimension reduction model has been proposed for uncertainty quantification and reservoir calibration with less computation time [47].

4.3. Semi-Supervised Learning

Semi-supervised learning algorithms involve few amounts of labelled classes with more unlabelled classes during training. It also classifies the problem of classification using both labelled and unlabelled data. Semi-supervised learning clustering can be identified as constraint clustering where there is an issue with creating a clustering based on labelled and unlabelled data [53]. Furthermore, semi-supervised learning has been proposed for long facies identification in carbonate reservoirs based on multiclass positive and unlabelled PU-learning ML [48]. Conclusively, a new approach has been proposed for oil prediction using a new graph, Laplacian, which is based on a semi-supervised clustering technique [54]. More recent evidence, as reported Salem et al. [55], indicates that semi-supervised learning plays a vital role in addressing diverse problems in oil and gas industry digitalization. The authors also identified the potential of semi-supervised learning in predicting well integrity failure after learning.

4.4. Reinforcement Learning

Reinforcement learning algorithms involve a learning process with no label class but which consists of reward value. However, it can be somewhere between supervised and unsupervised learning. Dong et al. [56] proposed a deep reinforcement learning (DRL) for automatic curve matching to achieve well-testing interpretation through evaluation reservation parameters based on a double deep Q-network (DDQN). Gas turbine maintenance using an optimal part flow management approach has been proposed based on reinforcement learning to overcome a sequential problem [57]. A recent study has indicated the capabilities of reinforcement learning in the oil and gas industries. Nasir et al. [58] presented a deep reinforcement learning which provides optimization and development plans for reservoir models with less computation cost.

5. Proposed Framework

In this section, the proposed machine learning model will be developed using the workflow described in Figure 4: proposed research methodology. The work will start with the problem definition stage, which is studying the existing Metocean data system as well as the business needs/objectives. Each attribute of the data will be identified, and the Metocean data for the solution is prepared accordingly. The next stage will be gathering and preparation of data. This includes analysing and preparing for structuring and distributing of the data. The data will be organized and well prepared to be ready for the next level. Data conversion and aggregation will be conducted in this phase. The first training data set will be utilized in this phase. By using a Gaussian-means algorithm, the Metocean data will be classified into different distinct areas, while other data will then be chosen randomly, using Equation (5). Fundamentally, kernel density estimation (KDE) will be represented as K, with the σ as bandwidth. Thus, x as the weighing of the observation from a particular point can be expressed as:

$$p(x) = \frac{1}{N}\sum_{n=1}^{n}\left(\left|\left|\frac{x - x_n}{\sigma}\right|\right|^2\right),\ k(t) \propto e^{-t/2} \tag{5}$$

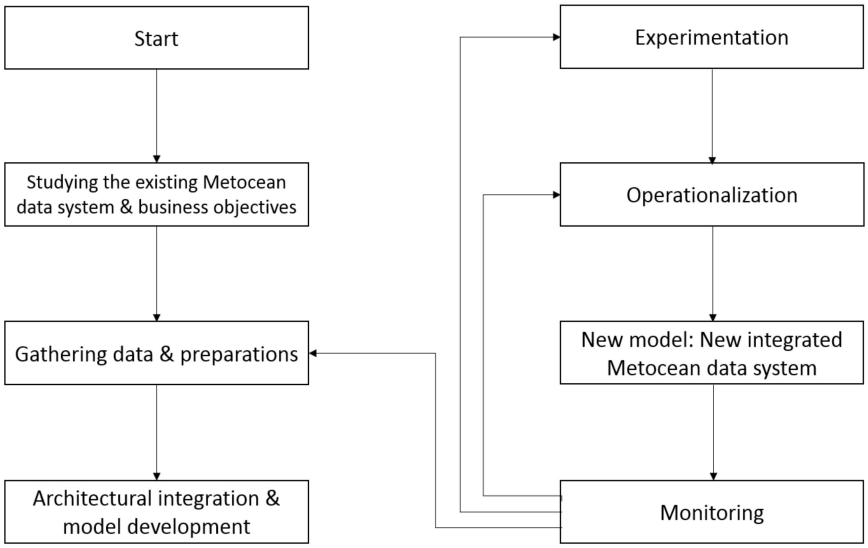


Figure 4. Proposed research methodology.

In the architectural integration and model development stage, algorithm development and evaluation will be conducted. This consists of exploring and selecting the algorithm and reporting the interpretation of the results. Testing of the dataset occur, after which the training dataset will be run again, plus subsequent tests. Python programming language and Hitachi Infrastructure Analytics Advisor will be used for application development. To fully deploy a scalable system, cloud infrastructure is suggested to be developed in this work, as the data will be collected from sensors in real time and non-real time. Therefore, Software as a Service (SaaS) will act as a leveraging platform, such that we will exploit and launch a better Metocean data system with AI Solutions. The proof of concept includes the Metocean parameters (input and output) and component integration, while the decision tree is the data-driven technique. Thus, a new predictive model will then be developed.

Conversely, training the data model in this work will rely on an unsupervised machine learning algorithm. The expectation–maximization (EM) algorithm will provide the fundamental concepts in graphical models and inference algorithms on graphs. It will simplify the iteration of parameters by optimizing the lower bound function. Estimating the ML parameters in this model can be one of the anticipated contributions of this research.

Therefore, for a model with latent/hidden variables x for the data points and y as the observed variables, the lower bound can be considered as follows:

$$L(\theta) = \log p(y|\theta) = \log \int p(x,\ y|\theta)dx \tag{6}$$

$$= \log \int q(x) \frac{p(x,y|\theta)}{q(x)} dx \tag{7}$$

$$\geq \int q(x) \log \frac{p(x,y|\theta)}{q(x)} dx \stackrel{def}{=} F(q,\theta) \tag{8}$$

where $q(x)$ is an arbitrary density function, while the lower bound F is the function of both the density $q(x)$ and the model parameters. As most of the Metocean data are the time series data generated from offshore sensors, these data are unstructured. This research will model the time series data using the state-space model (SSM). Equation (9) describes the observation factors and hidden state's joint probability. The sequence of observed data y_1, y_2, y_3, ... can be generated from a sequence of hidden state variables x_1, x_2, x_3 .... and X_1, X_2, ... and X_T is represented by X_{1-T}:

$$p(x_{1:T,}y_{1:T}|\theta) = \prod_{t=1}^{T} p(x_t|x_{t-1,}\theta)p(y_{t,}|x_t\theta) \tag{9}$$

During the stage of experimentation, a test harness will be employed, where test model, test data, and evaluation of the new ML algorithm will begin. The performance will also be determined in this phase. The incorporation of the new model with the existing Metocean system will be determined. The memory size and parametrization of the system will also be evaluated after confirming the robustness of the K-mean algorithm. Furthermore, clustering, neural networks, and anomaly detection are all common algorithms used in unsupervised machine learning. However, this project focuses on clustering (using K-Mean algorithm). This is mainly because of the quality nature of Metocean data.

Operationalization stage: this phase includes the data interpretation, reporting services, and generating results such as quality control (QC), environmental statistical analysis, spectral analysis, etc., by Metocean users. Confirmation of successful interoperability and data integration of the model will be determined in this operationalization stage. Finally, a new integrated Metocean data system will be deployed, tested, and monitored.

6. The Need to Digitalize, Automate, Integrate Oil and Gas Data

In this section, the need for oil and gas data digitalization is described. Digitalization acts as an enabler that brings value to all oil and gas stakeholders, especially at the time of downtime [59]. Digitalization can be referred to as a way of restructuring digital infrastructure and communications by organizations or industries [60]. Therefore, digitalization is a good business opportunity for the oil and gas industry, especially during the crush of oil prices. For this reason, it was estimated that the net benefits of the oil and gas industry due to digital transformation will be USD 945 billion over the decade to 2025 (cumulative value 2016–2025) [59]. The oil and gas sector has been one of the most significant and competitive economic sectors around the globe [61,62]. To have successful operations and exploration success, the industry needs to sustain its production. Nonetheless, one of the positive implications of an organization's performance is that the organization sustains its competitive advantage [63,64]. It is posited that "success in the future of oil and gas will require the continued adaptation of the complex business model to unforeseen challenges". On this account, the competitiveness and the demand for digitalization (especially in the AI and big data domains) have made the implementation necessary [65].

Digitalization of Oil and Gas Using Machine Learning

Reviews on digitalization were conducted using ML on recent developments in oil and gas. One of the findings revealed that oil and gas should leverage the new technological developments with focus, agility, and collaborative teams of big data [41]. Another study was conducted on evaluating the status of the data-driven approach in the oil and gas industry, where historical data is used [45]. It was found that the data-driven approach provides huge advantages in the industry over the conventional approach under certain conditions. Unfortunately, it was also discovered that the approach for many industry professionals remains fuzzy. Thus, it is clear that there is a need for further study of ML in the oil and gas domain. Figure 5 shows a real-time dashboard illustration for web application necessaries, functionality, analytics library, and web-based programming.

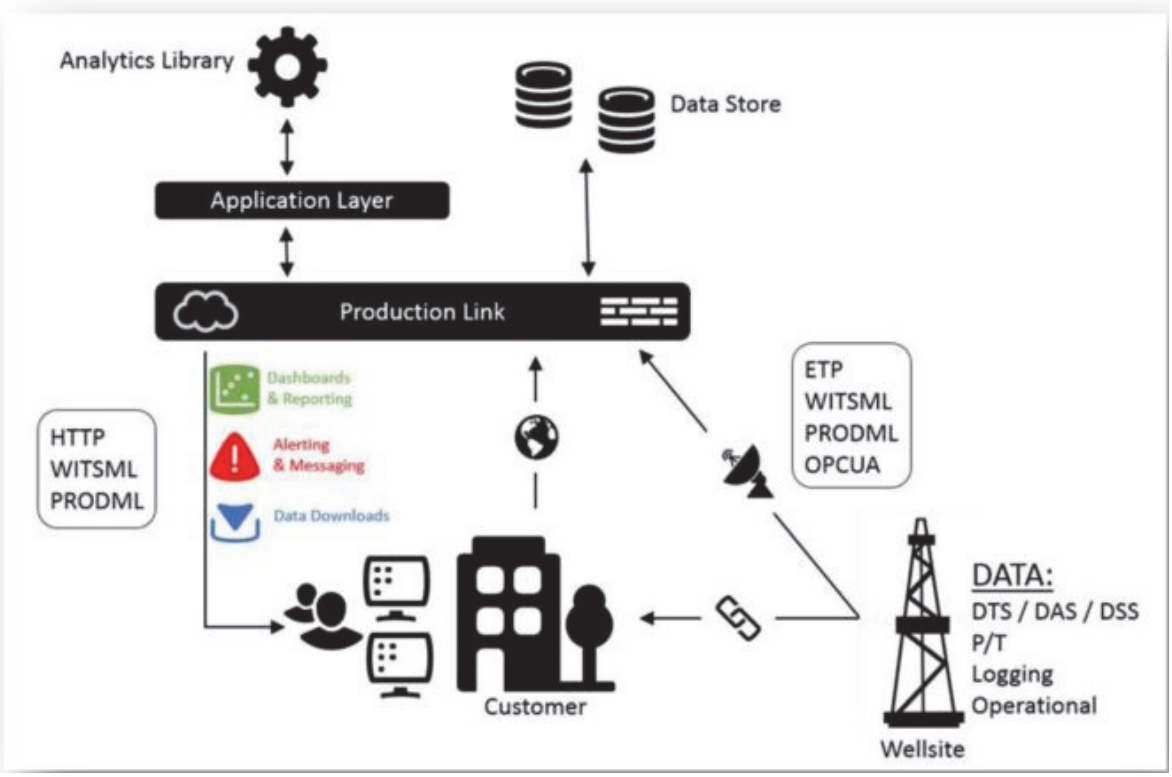


Figure 5. Data communication and flow between field device, cloud server, and customer [66].

Keerqinhu et al. [67] attempted to solve issues of reciprocating compressors in the petroleum industry. The findings suggested a system for fault diagnosis for reciprocating compressors using ML techniques based on a learned dictionary. The system evaluated 5-year operations collected from offshore oil corporations in the cloud environment. Significantly, their proposed system showed better results and indicated 80% accuracy, which can effectively diagnose potential faults in compressors. Similarly, research was conducted by applying a machine learning approach for big data, sizing of metal-loss defects, and failure risk analysis in oil and gas pipelines [68–70]. They adopted Levenberg–Marquardt's back-propagation learning algorithm [71]. They found a promising result with an estimated accuracy of 86% (±10% error tolerance) and 89% (±15% error tolerance). A recent work conducted by Yang et al. attempts to provide an elastically scalable cloud-based system to solve big data issues for the upstream oil and gas industry with high performance. Yang et al. [66] used machine learning and processed complex datasets in real time or near real time effectively. Figure 6 presents the thermal analysis based on distributed temperature sensing (DTS) measurements on total production. The proposed integrated system contains several services, including legacy oil and gas data. However, this work has not been integrated with meteorological or oceanographic data. There is no review

and approval process in the system, which reduces the quality of the proposed monitoring. Hence, a system that will incorporate digitalization, monitoring, and data integration of meteorological and oceanographic or Metocean data is necessary.

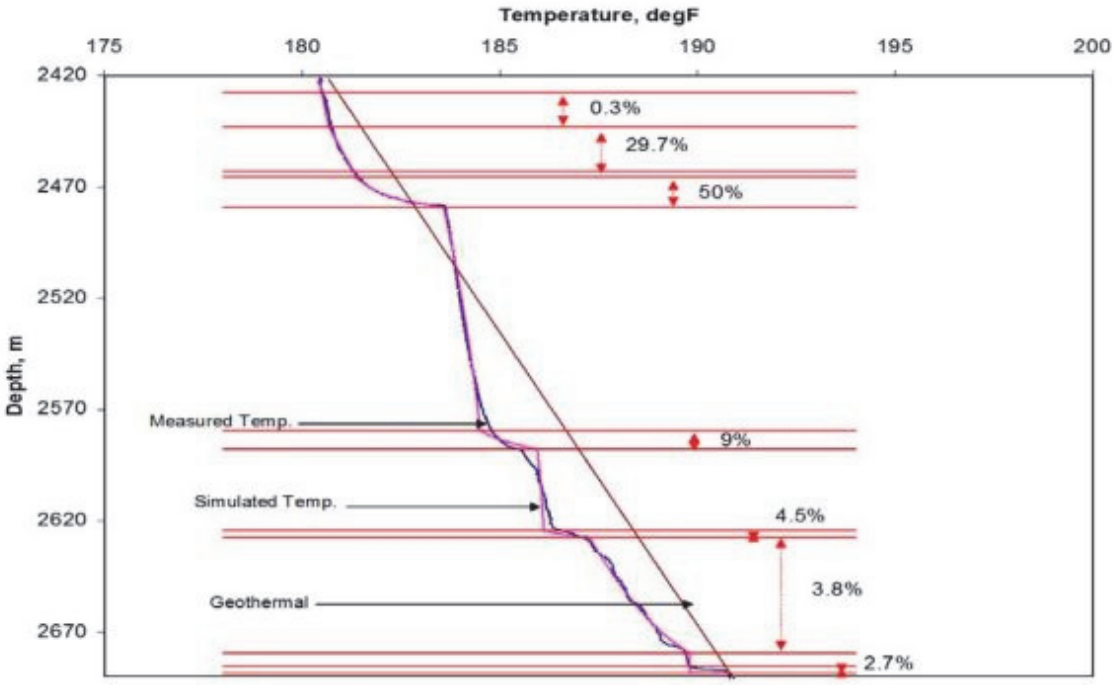


Figure 6. Gas production profiling using DTS data [66].

7. Metocean Data System

This section presents an overview of the Metocean data system. Metocean conditions are directly related to the offshore project, project operation, and maintenance [72].

Metocean Data System in Malaysia

Oil and gas companies use Metocean data as the main information source for determining the time and other weather factors. The users that benefit from these data comprise the participants of oil and gas companies, the partners, the contributors, as well as the research institutes, and are of extreme value [73–75]. The Metocean contributors are BP, Chevron, PETRONAS, Statoil, Total and research institutes. In Malaysia, demographically, the main Metocean contributor in Malaysia is PETRONAS, which is the Malaysian-owned oil and gas company that is among the Metocean contributors. It has enormous resources with diverse data for scientific research but they have not implemented AI or machine learning technologies in its Metocean data system. The information supplied by Metocean to users is primarily for oil spill response, satellite locations, environmental platforms and other types of drifting buoys [76]. The Metocean database has high quality of oil and gas datasets which are structured on uniform formats of ASCII, NetCDF and indexed upon the ISO 19115 metadata standard [73]. Metocean has become an industry that handles large amounts of data and metadata which yield immense benefits to oil and gas stakeholders. The results of the Metocean data are in the form of text or graph formats. Undeniably, the Metocean data are the resources for petroleum industries as well as human life activities, which are therefore necessary to be monitored, automated, and interoperable for the benefit of its users.

As PETRONAS remains the major oil and gas company in Malaysia, it officiously intends to diversify the energy sector beyond 2020 with a high competitive advantage [77,78]. More recently, as reported in the PETRONAS activity outlook 2019–2021, "PETRONAS is actively seeking ways to deploy technology in terms of digital, data analytics, automation, etc." [79]. Thus, PETRONAS must make the Metocean data system fully autonomous and digitalized to

achieve its goal effectively. In addition, some ML initiatives have been seen in the oil and gas sector. While the sector wants to transform into a fully digitally data-driven industry, it is not enough to solve the continuous problem of integrating all oil and gas data over the globe using AI or ML. This indicates the presence of limited integration and few empirical studies of ML, digitalization, and data integration in the oil and gas sector. Hence, the need to fill this gap is apparent. The Metocean data system flow information for assessment process based on risk inspection technique is illustrated in Figure 7.

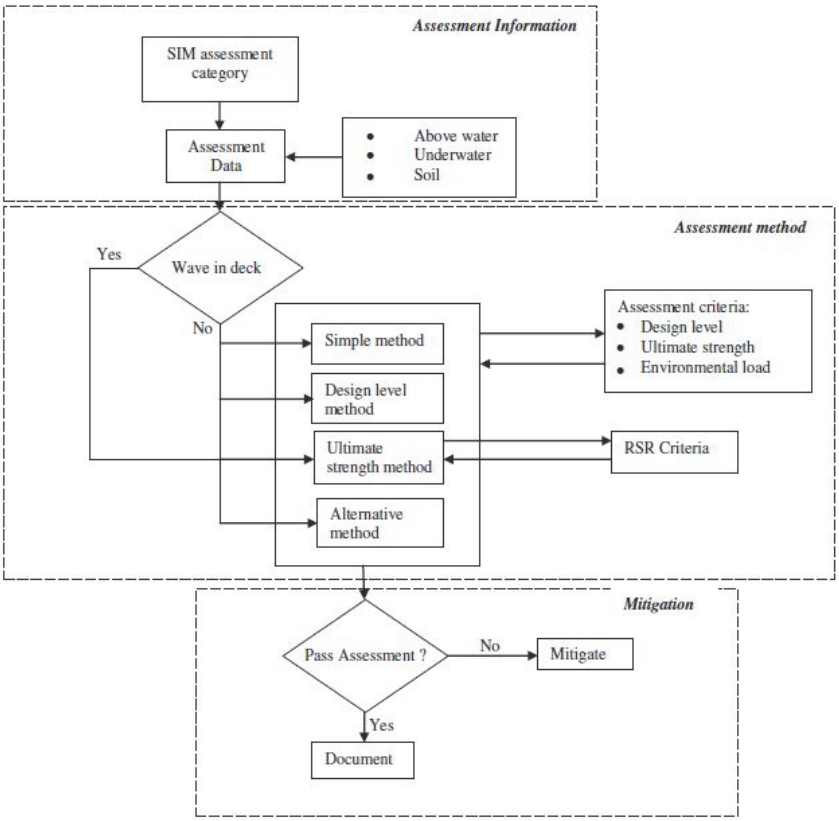


Figure 7. Metocean data assessment process for risk-based inspection technique [80].

A growing body of literature has investigated the Metocean data system in Malaysia. Lai et al. [81] investigated the effectiveness of genetic programming (GA) and support vector machine (SVM) learning models in predicting monthly sea-level variations by comparing model accuracy performance. Adding to this, the model has been validated using datasets from Tioman Island, Kerteh, and Tanjung sedili, Malaysia. Moreover, other methods have been applied in Metocean data systems by providing semantic web technologies for the oil and gas industries, which include an architecture for data integration. The application database utilized includes RDF query (D2RQ) for setting performance. The study data were obtained from Malaysia's oil and gas industries [82]. More recent studies have investigated Metocean data systems around the globe. Qiao et al. [83] implemented a numerical model using Mike 21 to estimate Metocean conditions by evaluating hurricanes and hindcasts on the United States Atlantic coast.

8. Machine Learning for Metocean Data Integration

In this section, the impact of ML in Metocean data integration is elucidated. ML algorithms play a vital role in Metocean data integration by improving accuracy with computational performance for forecasted wave conditions, wind parks, offshore modelling, and many more. Chen et al. [46] proposed the ML technique for Metocean data integration to derive spatial wave data. However, ML algorithms can be applied for Metocean data model comparison to predict offshore platform integrity, model structure integrity, and

geohazard data [84]. In Wyatt [85], the author investigated the Metocean data parameters towards HF radar wind speed and measurement using various ML techniques. However, SVM regression provides efficient results in wind speed estimations. The proposed method has been validated using different radar systems at different locations [85]. In addition, ML algorithms have been used in the prediction of Metocean data in Korea Strait, where feedforward neural network (FNN) and long-short term memory (LSTM) models have been used for the prediction of wave height. The proposed model has been validated using Metocean data from 2012–2020 from the Korea Institute of Ocean Science and Technology [86].

8.1. ML for Metocean Forecasters Data

Metocean forecasting data are models that provide in-depth reliable weather information in an ocean. In addition, they provide analysis, planning, and securing of offshore operations. ML-based techniques improve the performance of Metocean forecasting data. For example, Martinez-Perurena et al. [87] designed a forecast of Metocean data for a marine renewable energy system using hybrid ML techniques which include support vector regression and random forest. Moreover, a forecasting model has been proposed for Metocean prediction using linear regression and H2O auto-ML techniques for knowledge enrichment [88,89].

8.2. ML for Metocean Spatial Wave Data

Metocean spatial wave data provide an accurate spatial analysis using computational models to predict wave situations across the ocean. This prediction can be improved using ML-based techniques. Chen et al. [46] proposed ML techniques to drive spatial wave data using a novel surrogate SWAN numerical model based on random forest to replicate the spatial nearshore wave data. In addition, the model demonstrated ML abilities for correlating the spatial ocean waves through optimal spatial gridding and provided high-resolution distribution in real time [46].

8.3. ML for Metocean Data Linked

The Metocean data link provides structured interlinks with other data to become useful in semantic queries such as providing information to petroleum and research industries. However, Metocean data or oceanographic datasets are scattered when published, and processing around the globe due to their huge amount leads to linked data and semantic web capabilities [90]. ML capabilities can be used to improve data-linked processing. Colin et al. [91] proposed a semantic segmentation containing ten Metoceanic processing large-quality image-level ground truths using deep-learning techniques.

8.4. ML for Metocean Conditions during Hurricanes

Metocean conditions in hurricanes have historically damaged the environment, which can occur by multiple parameters such as significant surge height and wave height that can vary in time and cross-correlation [23]. However, Metocean systems are remote complexes which lead to hazards such as hurricanes that affect platforms [84]. For example, in 2014, significant damage to the Taylor energy oil platform was caused by Hurricane Ivan after being installed for 20 years [92]. ML techniques can be applied for hurricane predictions. Asthana et al. [93] proposed an ML model for the prediction of Atlantic hurricane activity based on a convolutional neural network (CNN) approach.

9. Discussion

In this section, we discuss the findings of this research study. The critical analysis and evaluation for the integration of Metocean data intelligent operations using ML models based on empirical studies have been revealed. The oil and gas industry's business objective is to digitalize its production and development, which cannot be achieved without the implementation of ML techniques and big data analytics. In addition, the transformation and integration of existing Metocean architecture for oil and gas data operation can be

enhanced with the advancement of ML techniques. The Metocean data development and integration significantly improved after applying ML techniques such as better prediction, providing efficient results, saving time, reducing system complexity, and reducing operational overheads. Moreover, the ML methods vary in different categories with its subsection known as DL in Metocean data analytics and integration.

As shown in Table 2, we indicated AI models that could fit in the Metocean data system. AI models have been the environment for creating, training, and implementing models based on available datasets for better decision making and predictions, for example, classification, regression, and clustering, where authors compare several ML models which include ANN, XGBoost, SVM, and statistical regression for ship power modelling with measuring Metocean condition and hull maintenance [94,95]. Convolutional neural network (CCN) models are part of deep neural networks which play vital roles in image classification problems through their multiple convolution layers with kernels to detect complex features. Adding to this, a novel CCN model has been applied for failure identification of mooring line and turret-moored FPSO systems [96]. Random forests (RF) are ML models based on trees assembled for prediction with high effective performance to solve regression and classification problems. For example, RF models have been developed to predict spatial distribution and frequency groundings of a ship [97]. As the modern technique provides an avenue for data-driven modelling using various methods, genetic programming (GP) techniques have evolved programs that can develop and compute solutions to human problems which cannot be solved directly. For example, the GP technique has been used for partial differential equations (PDE), discovery of Metocean processing, and acoustics function discovery [98]. The Bayesian framework for ML has been applied through Gaussian processing (GP) for efficient integrity management (IM) of steel lazy wave risers (SLWRs). However, GP is genetically supervised learning for probabilistic classification problems and regression [99]. Lastly, the state-of-art ML presented is efficient and effective in performance, which includes predictions, classifications, and decisions. All models discussed can fit the Metocean data system for better forecasting, integration, and intelligent operations.

As shown in Table 3, state-of-art studies have been listed with insight on Metocean data integration in the oil and gas domain using ML techniques. Notably, ANN, SVM, RF, and XGB are the most used models for Metocean data analytics and integration due to their efficient performance. However, ML performance towards the prediction of Metocean data interoperability for intelligent operations can be enhanced through hyperparameter tuning and feature selection techniques for better performance.

Table 2. List of some selected AI models that could fit the Metocean data system.

Authors	Year	Model	Application
Lang et al. [94]	2022	Artificial neural network (ANN)	Classification, clustering
		Support vector machine (SVM)	Classification
		Extreme gradient boosting (XGBoost)	Classification, regression
Janas et al. [96]	2021	Convolutional neural network (CNN)	Classification
Rawson et al. [97]	2021	Random forest (RF)	Classification, regression
Hvatov and Maslyaev [98]	2020	Genetic programming (GP)	Symbolic regression, classification
		Sparse regression	Classification
Hejazi et al. [99]	2020	Gaussian processes (GP)	Classification, regression
Park et al. [86]	2021	long-short term memory (LSTM)	Classification
Tadjer et al. [47]	2021	Gaussian processes (GP)	Classification, regression

Table 3. List of ML state-of-art studies with their pros and cons.

Authors	Year	Application	Technique	Pros	Cons
Syed et al. [100]	2022	Oil and gas production	GBRT	- Handles missing data - Optimization flexibility	- High computational complexity - Less interpretable
Dyer et al. [84]	2022	Metocean conditions	GBRT, ANN	- Handles missing data - Trustworthy predictions	- High computational complexity - High processing time for large NN
Mingyu et al. [101]	2021	Oil and gas development	DT, RF	- Optimization flexibility - Improved accuracy	- High computational complexity - High training time
García-Alba et al. [102]	2021	Metocean conditions	DNN	- Enhanced robustness	- Accuracy detection need enhancement
Maslyaev et al. [103]	2020	Metocean conditions	Sparse regression	- Works well on non-linear problems	- Compulsory to apply feature scaling
Alsaihati et al. [104]	2020	Oil and gas drilling	RF	- Improved accuracy - Works well with categorical and continuous values	- High computational power - High training time
Brown et al. [105]	2020	Oil and gas exploration	NN, XGBoost	- Handles missing data - Trustworthy predictions	- Trustworthy predictions - High computational complexity
Fan et al. [106]	2020	Metocean conditions	LSTM, SVM	- High dimensional space - Good in time series prediction	- Gradient exploding - Interpretation of result needs enhancement
Orrù et al. [107]	2020	Oil and gas reliability/maintenance	SVM, MLP	- Capable of fault tolerance - Parallel processing capability	- High computational complexity
Aldosari et al. [108]	2020	Oil and gas inspection	ANN, GPR	- Trustworthy predictions - Prediction is probabilistic	- High processing time for large NN - Efficiency lose in high dimensional spaces
Wang et al. [109]	2020	Oil and gas	LR	- Works well with classifying unknown records - Good in interpreting model coefficient	- Constructed linear boundaries - Limited to predict only discrete functions
Alharam et al. [110]	2020	Oil and Gas Inspection	RF	- Improved accuracy - Works well with categorical and continuous values	- High computational power - High training time
Deeva et al. [111]	2019	Metocean	Sparse regression	- Works well on non-linear problems	- Compulsory to apply feature scaling
Granado et al. [112]	2019	Metocean	BN	-	-
Abbas et al. [113]	2019	Oil and gas drilling	SVM	- Efficient memory - High dimensional space	- Interpretation of result needs enhancement
Khan et al. [114]	2019	Oil and gas production	ANN	- Trustworthy predictions	- High processing time for large NN
Sousa et al. [115]	2019	Oil and gas supply chain	ANN	- Trustworthy predictions	- High processing time for large NN

Table 3. *Cont.*

Authors	Year	Application	Technique	Pros	Cons
Paltrinieri et al. [116]	2019	Oil and gas risk management	DNN	- Enhanced robustness	- Accuracy detection needs enhancement
Tings et al. [117]	2019	Metocean	LR	- Works well with classifying unknown records - Good in interpreting model coefficient	- Construct linear boundaries - Limited to predict only discrete functions
Vychuzhanin et al. [118]	2018	Metocean	CNN	- High accuracy in image recognition - Weight sharing	- High training data are required

10. Limitation of the Study

Our review study comprehensively summarizes the deep insight on Metocean data interoperability for intelligent operations using ML techniques. Consequently, certain ML techniques have not been included, especially those under DL methods. After a critical analysis, it was revealed that some proposed models are weak due to the ML method and data analysis procedure used, which can lead to how the model performed.

In addition, each proposed technique has certain issues which have not yet been resolved. For example, some studies are having problems with datasets, feature reports, predictive features, and high-quality data. Our study has the potential for further investigation to provide a comparison for the state-of-art studies model performance and efficiency. Adding to this, several aspects of Metocean in oil and gas development are excluded. Lastly, the meta-survey studies have considered published articles between 2009 and 2022.

The current limitation can be overcome by implementing additional AI models based on DL algorithms such as CCN, RNN, GAN, and DBN to investigate Metocean data conditions. The proposed techniques/models can also be improved through hyperparameter turning or feature selection techniques. Adding to this, data quality assurance has to be considered in selecting datasets. However, relevant related studies before 2019 can be referred to, and other database sources can be explored.

11. Recommendation for Future Study

Owing to the digital transformation of the oil and gas sector, which includes Metocean data interoperability for intelligent operations using AI technique, further investigation of existing ML approaches is recommended. Additionally, AI algorithms are moving towards statistical reasoning by introducing next-generation approaches to automate the Metocean, including oil and gas development. Our result revealed that different AI techniques have been proposed and implemented based on ML models which include hybrid methods of different algorithms.

However, there is a need to improve the existing intelligent framework by introducing different combinations of AI models for better performance. Therefore, further work needs to concentrate on selecting suitable preferred ML models, high-quality data, real-world datasets, and improving result accuracy. Lastly, problems of imbalanced data need to be considered due to the huge amount of Metocean data.

12. Conclusions

In this paper, a comprehensive review of development for Metocean data integration, interoperability based on intelligent operations, and automation using ML techniques is presented. The existing Metocean system generates data using sensors such as wave buoy, anemometer and acoustic doppler current profiler (ADCP) manually. Additionally, these data which appear in ASCII format to the Metocean system are also manual and silos. There is no fundamental method, model, or algorithm that allows the system to reason and

automate intelligently. This slows down provisioning while the monitoring element of the Metocean data path is partial.

However, with the current oil and gas crash, digitalization is a good opportunity that will bring value to all its stakeholders, which shows that there is a need to continue in digitalization of the oil and gas sector to achieve the expected, or more of the, stated value by the year 2025. Moreover, the digitalization of the oil and gas production-based industries is driven by enabling technologies which include ML and big data analytics. ML is one of the crucial and key enabling tools that oil and gas industries are now focusing on for digital transformation implementation which allows the industry to automate, enhance production, and have less human capacity.

We summarized several research studies exploring the needs of ML in oil and gas industries by investigating in depth the integration of Metocean data interoperability for intelligent operations and automation using an ML-based approach. In addition, a comparative analysis of the current uses of ML algorithms in the oil and gas industries was provided, and we reviewed the transformation and integration of existing Metocean architecture for the use of oil and gas data operators through automation. Lastly, a new model and its integration with the existing Metocean data system using an ML algorithm to monitor and interoperate with maximum performance was proposed.

Author Contributions: Conceptualization, K.U.D.; methodology, M.A. and K.U.D.; software, K.U.D., M.A. and M.Y.A.; validation, K.U.D., M.S.L. and L.E.S.; formal analysis, K.U.D. and H.H.H.; investigation, K.U.D. and M.A.; resources, H.H.H. and M.S.L.; data curation, K.U.D., M.A. and L.E.S.; writing—original draft preparation, K.U.D. and H.H.H.; writing—review and editing, M.A. and M.Y.A.; visualization, K.U.D. and M.A.; supervision, K.U.D. and M.S.L.; project administration, K.U.D., H.H.H. and L.E.S.; funding acquisition, H.H.H. and M.S.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Short Term Internal Research Funding (STIRF)grant, grant numbers: 015LA0-026.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: We would like to acknowledge M. Nasir B Abdullah, the Head Metocean, Group Technical Solutions (GTS) for providing the necessary support in terms of Metocean.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

ADCP	Acoustic doppler current profiler
AI	Artificial intelligence
AARI	Antarctic Research Institute
ANN	Artificial neural network
DL	Deep learning
DNN	Deep neural network
DBN	Deep belief network
CNN	Convolutional neural network
DDQN	Double deep Q-network
DT	Decision tree
EM	Expectation maximization
FNN	Feedforward neural network
GA	Genetic algorithm
GAN	Generative adversarial network
GTU	Gated recurrent unit

GP	Genetic programming
GP	Gaussian processing
GBRT	Gradient Boosted Regression Trees
KDE	Kernel density estimation
LR	Linear regression
LR	Logistic regression
LSTM	Long short-term memory
ML	Machine learning
MLP	Multilayer perception
NN	Neural network
PDE	Partial differential equations
QC	Quality control
US	United States
RNN	Recurrent neural network
RF	Random forest
SWAN	Simulating Waves Nearshore
SVM	Support vector machine
SaaS	Software as a Service
SSM	State-space model
SLWRs	Steel lazy wave risers
TL	Transmission loss
XGBoost	Extreme Gradient Boosting

References

1. Tiwari, S. Data is the New Oil, even in the Oil and Gas Industry. Available online: https://medium.com/@shubham0189 (accessed on 9 March 2022).
2. Navarro, A.M.G.; Capozzoli, A.; Rocca, V.; Romagnoli, R. A glance about the Big Data Analytics in the Oil & Gas industry. *Geoing. Ambient. Min.* **2020**, *160*, 36–43. [CrossRef]
3. Li, Y.; Han, Y. Decline curve analysis for production forecasting based on machine learning. In Proceedings of the SPE Symposium: Production Enhancement and Cost Optimisation, Kuala Lumpur, Malaysia, 7–8 November 2017. [CrossRef]
4. Ministry of International Trade and Industry. MyGOV—Capacities for Digital Transformation | Industry4WRD: National Policy on Industry 4.0. Available online: https://www.malaysia.gov.my/portal/content/30610 (accessed on 9 March 2022).
5. Flichy, P.; Baudoin, C. The Industrial IoT in Oil & Gas: Use Cases. In Proceedings of the SPE Annual Technical Conference and Exhibition, Dallas, TX, USA, 24–26 September 2018.
6. Abdullahi, M.; Baashar, Y.; Alhussian, H.; Alwadain, A.; Aziz, N.; Capretz, L.F.; Abdulkadir, S.J. Detecting Cybersecurity Attacks in Internet of Things Using Artificial Intelligence Methods: A Systematic Literature Review. *Electronics* **2022**, *11*, 198. [CrossRef]
7. Zu, Y.; Siréta, F.X.; Zhang, D.; Liu, X.; Sun, H. Machine learning for motion prediction during dry transportation. *Proc. Int. Offshore Polar Eng. Conf.* **2019**, *1*, 936–942.
8. Kellner, L.; Stender, M.; von Bock und Polach, R.U.F.; Herrnring, H.; Ehlers, S.; Hoffmann, N.; Høyland, K.V. Establishing a common database of ice experiments and using machine learning to understand and predict ice behavior. *Cold Reg. Sci. Technol.* **2019**, *162*, 56–73. [CrossRef]
9. Krasnov, F.; Glavnov, N.; Sitnikov, A. A machine learning approach to enhanced oil recovery prediction. *Lect. Notes Comput. Sci. Incl. Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinform.* **2018**, *10716*, 164–171. [CrossRef]
10. Tewari, S.; Dwivedi, U.D.; Biswas, S. A novel application of ensemble methods with data resampling techniques for drill bit selection in the oil and gas industry. *Energies* **2021**, *14*, 432. [CrossRef]
11. Maucec, M.; Garni, S. Application of automated machine learning for multi-variate prediction of well production. In Proceedings of the SPE Middle East Oil and Gas Show and Conference, Manama, Bahrain, 18–21 March 2019. [CrossRef]
12. Onwuchekwa, C. Application of machine learning ideas to reservoir fluid properties estimation. In Proceedings of the SPE Nigeria Annual International Conference and Exhibition, Lagos, Nigeria, 6–8 August 2018. [CrossRef]
13. Hou, X.; Yang, J.; Yin, Q.; Chen, L.; Cao, B.; Xu, J.; Meng, L.; Zhang, Y.; Xu, D.; Zhao, X. Automatic gas influxes detection in offshore drilling based on machine learning technology. In Proceedings of the SPE Gas & Oil Technology Showcase and Conference, Dubai, United Arab Emirates, 21–23 October 2019. [CrossRef]
14. Hazbeh, O.; Aghdam, S.K.Y.; Ghorbani, H.; Mohamadian, N.; Ahmadi Alvar, M.; Moghadasi, J. Comparison of accuracy and computational performance between the machine learning algorithms for rate of penetration in directional drilling well. *Pet. Res.* **2021**, *6*, 271–282. [CrossRef]
15. Rachman, A.; Ratnayake, R.M.C. Corrosion loop development of oil and gas piping system based on machine learning and group technology method. *J. Qual. Maint. Eng.* **2020**, *26*, 349–368. [CrossRef]

16. Hegde, C.; Gray, K. Evaluation of coupled machine learning models for drilling optimization. *J. Nat. Gas Sci. Eng.* **2018**, *56*, 397–407. [CrossRef]
17. Van Pham, V.; Fathi, E.; Belyadi, F. New Hybrid Approach for Developing Automated Machine Learning Workflows: A Real Case Application in Evaluation of Marcellus Shale Gas Production. *Fuels* **2021**, *2*, 286–303. [CrossRef]
18. Mask, G.; Wu, X.; Ling, K. An improved model for gas-liquid flow pattern prediction based on machine learning. *J. Pet. Sci. Eng.* **2019**, *183*, 106370. [CrossRef]
19. Sabah, M.; Mehrad, M.; Ashrafi, S.B.; Wood, D.A.; Fathi, S. Hybrid machine learning algorithms to enhance lost-circulation prediction and management in the Marun oil field. *J. Pet. Sci. Eng.* **2021**, *198*, 108125. [CrossRef]
20. Liu, J.; Gu, J.; Li, H.; Carlson, K.H. Machine learning and transport simulations for groundwater anomaly detection. *J. Comput. Appl. Math.* **2020**, *380*, 112982. [CrossRef]
21. Alshaikh, A.; Magana-Mora, A.; Al Gharbi, S.; Al-Yami, A. Machine learning for detecting stuck pipe incidents: Data analytics and models evaluation. In Proceedings of the International Petroleum Technology Conference, Beijing, China, 26–28 March 2019. [CrossRef]
22. Quaeghebeur, E.; Zaaijer, M.B. How to improve the state of the art in metocean measurement datasets. *Wind Energy Sci.* **2020**, *5*, 285–308. [CrossRef]
23. Qiao, C.; Myers, A.T. Surrogate modeling of time-dependent metocean conditions during hurricanes. *Nat. Hazards* **2022**, *110*, 1545–1563. [CrossRef]
24. Orlandi, A.; Cappugi, A.; Mari, R.; Pasi, F.; Ortolani, A. Meteorological navigation by integrating metocean forecast data and ship performance models into an ecdis-like e-navigation prototype interface. *J. Mar. Sci. Eng.* **2021**, *9*, 502. [CrossRef]
25. Kohli, R.; Johnson, S. Digital Transformation in Latecomer. *MIS Q. Exec.* **2011**, *10*, 141–156.
26. Tran, T.; Tran, T.; Ngo, H.; Thanh, T. Digital Transformation in Oil and Gas Companies–A case study of Bien Dong POC. *Petrovietnam J.* **2020**, *2020*, 68–81. [CrossRef]
27. Rahman, M.S.A.; Jamaludin, N.A.A.; Zainol, Z.; Sembok, T.M.T. Machine Learning Algorithm Model for Improving Business Decisions Making in Upstream Oil Gas. In Proceedings of the 2021 International Congress of Advanced Technology and Engineering (ICOTEN), Taiz, Yemen, 4–5 July 2021; pp. 1–5. [CrossRef]
28. Maroufkhani, P.; Desouza, K.C.; Perrons, R.K.; Iranmanesh, M. Digital transformation in the resource and energy sectors: A systematic review. *Resour. Policy* **2022**, *76*, 102622. [CrossRef]
29. Aissani, N.; Beldjilali, B.; Trentesaux, D. Dynamic scheduling of maintenance tasks in the petroleum industry: A reinforcement approach. *Eng. Appl. AI* **2009**, *22*, 1089–1103. [CrossRef]
30. Parapuram, G.; Mokhtari, M.; Ben Hmida, J. *Prediction and Analysis of Geomechanical Properties of the Upper Bakken Shale Utilizing Artificial Intelligence and Data Mining*; Unconventional Resources Technology Conference: Austin, TX, USA, 2017.
31. Castiñeira, D.; Toronyi, R.; Saleri, N. Machine learning and natural language processing for automated analysis of drilling and completion data. In Proceedings of the SPE Kingdom of Saudi Arabia Annual Technical Symposium and Exhibition, Dammam, Saudi Arabia, 23–26 April 2018. [CrossRef]
32. Pashali, A.A.; Boldyrev, M.L.; Kornishin, K.A.; Efimov, Y.O.; Tarasov, P.A.; Verbitskaya, O.A.; Gudoshnikov, Y.P.; Mironov, E.U.; Nesterov, A.V.; Kovalev, S.M.; et al. Ice and metocean survey for development of the Russian Arctic continental shelf (Russian). *Neft. Khozyaystvo-Oil Ind.* **2021**, *2021*, 8–12. [CrossRef]
33. Buzin, I.V.; Gudoshnikov, Y.P.; Nesterov, A.V.; Naumov, A.K.; Skutina, E.A.; Smirnov, K.G.; Andreev, O.M.; Vinogradov, R.A.; Novikov, S.A.; Skutin, A.A. Study of the metocean conditions for the Russian Arctic Shelf development projects approaches and results: Review for the period of 2012–2021. *IOP Conf. Ser. Mater. Sci. Eng.* **2021**, *1201*, 012057. [CrossRef]
34. Wang, P.; Tian, X.; Peng, T.; Luo, Y. A review of the state-of-the-art developments in the field monitoring of offshore structures. *Ocean Eng.* **2018**, *147*, 148–164. [CrossRef]
35. Nasiri, S.; Khosravani, M.R. Machine learning in predicting mechanical behavior of additively manufactured parts. *J. Mater. Res. Technol.* **2021**, *14*, 1137–1153. [CrossRef]
36. Verma, V.K.; Verma, S. Machine learning applications in healthcare sector: An overview. *Mater. Today Proc.* **2022**, *57*, 2144–2147. [CrossRef]
37. Tariq, Z.; Aljawad, M.S.; Hasan, A.; Murtaza, M.; Mohammed, E.; El-Husseiny, A.; Alarifi, S.A.; Mahmoud, M.; Abdulraheem, A. *A Systematic Review of Data Science and Machine Learning Applications to the Oil and Gas Industry*; Springer International Publishing: Berlin/Heidelberg, Germany, 2021; Volume 11, ISBN 1320202101.
38. Hassan, A.; Chan, S.; Aljawad, M.; Humphrey, J.; Abdulraheem, A. Artificial Intelligence-Based Model of Mineralogical Brittleness Index Based on Rock Elemental Compositions. *Arab. J. Sci. Eng.* **2022**, *47*. [CrossRef]
39. Aljawad, M.S.; Desouky, M.; Tariq, Z.; Alhoori, H.; Mahmoud, M.; AlShehri, D. Data-driven acid fracture conductivity correlations honoring different mineralogy and etching patterns. *ACS Omega* **2020**, *5*, 16919–16931. [CrossRef]
40. Gupta, S.; Li, L. The Potential of Machine Learning for Enhancing CO2 Sequestration, Storage, Transportation, and Utilization-based Processes: A Brief Perspective. *JOM* **2022**, *74*, 414–428. [CrossRef]
41. Hajizadeh, Y. Machine learning in oil and gas; a SWOT analysis approach. *J. Pet. Sci. Eng.* **2019**, *176*, 661–663. [CrossRef]
42. Hanga, K.M.; Kovalchuk, Y. Machine learning and multi-agent systems in oil and gas industry applications: A survey. *Comput. Sci. Rev.* **2019**, *34*, 100191. [CrossRef]

43. Barbosa, L.F.F.M.; Nascimento, A.; Mathias, M.H.; de Carvalho, J.A. Machine learning methods applied to drilling rate of penetration prediction and optimization—A review. *J. Pet. Sci. Eng.* **2019**, *183*, 106332. [CrossRef]
44. Kitchenham, B.A.; Charters, S. *Guidelines for Performing Systematic Literature Reviews in Software Engineering*; EBSE Technical Report EBSE-2007-01; School of Computer Science and Mathematics, Keele University: Newcastle, UK, 2007; pp. 1–57.
45. Balaji, K.; Rabiei, M.; Canbaz, C.H.; Agharzayeva, Z.; Tek, S.; Bulut, U.; Temizel, C.; Suicmez, V. Status of Data-Driven Methods and their Applications in Oil and Gas Industry. In Proceedings of the SPE Europec featured at 80th EAGE Conference and Exhibition, Copenhagen, Denmark, 11–14 June 2018.
46. Chen, J.; Pillai, A.C.; Johanning, L.; Ashton, I. Using machine learning to derive spatial wave data: A case study for a marine energy site. *Environ. Model. Softw.* **2021**, *142*, 105066. [CrossRef]
47. Tadjer, A.; Bratvold, R.B.; Hanea, R.G. Efficient dimensionality reduction methods in reservoir history matching. *Energies* **2021**, *14*, 3137. [CrossRef]
48. Lan, X.; Zou, C.; Kang, Z.; Wu, X. Log facies identification in carbonate reservoirs using multiclass semi-supervised learning strategy. *Fuel* **2021**, *302*, 121145. [CrossRef]
49. Liang, F.; Hatcher, W.G.; Liao, W.; Gao, W.; Yu, W. Machine Learning for Security and the Internet of Things: The Good, the Bad, and the Ugly. *IEEE Access* **2019**, *7*, 158126–158147. [CrossRef]
50. Chen, S.; Yu, H.; Lu, M.; Lebedev, M.; Li, X.; Yang, Z.; Cheng, W.; Yuan, Y.; Ding, S.; Johnson, L. A New Approach To Calculate Gas Saturation in Shale Reservoirs. *Energy Fuels* **2022**, *36*, 1904–1915. [CrossRef]
51. Makhotin, I.; Orlov, D.; Koroteev, D. Machine Learning to Rate and Predict the Efficiency of Waterflooding for Oil Production. *Energies* **2022**, *15*, 1199. [CrossRef]
52. Liu, Y.; Chen, W.; Chen, Y.; Chen, W.; Ma, L.; Meng, Z. Ocean front reconstruction method based on k-means algorithm iterative hierarchical clustering sound speed profile. *J. Mar. Sci. Eng.* **2021**, *9*, 1233. [CrossRef]
53. Bhowmick, K.; Narvekar, M. A Comprehensive Study and Analysis of Semi Supervised Learning Techniques. *Int. J. Eng. Tech. Res.* **2019**, *8*, 810–816.
54. Boesen, T.; Haber, E.; Hoversten, G.M. Data-driven semi-supervised clustering for oil prediction. *Comput. Geosci.* **2021**, *148*, 104684. [CrossRef]
55. Salem, A.M.; Yakoot, M.S.; Mahmoud, O. Addressing Diverse Petroleum Industry Problems Using Machine Learning Techniques: Literary Methodology—Spotlight on Predicting Well Integrity Failures. *ACS Omega* **2022**, *7*, 2504–2519. [CrossRef] [PubMed]
56. Dong, P.; Chen, Z.M.; Liao, X.W.; Yu, W. A deep reinforcement learning (DRL) based approach for well-testing interpretation to evaluate reservoir parameters. *Pet. Sci.* **2022**, *19*, 264–278. [CrossRef]
57. Compare, M.; Bellani, L.; Cobelli, E.; Zio, E.; Annunziata, F.; Carlevaro, F.; Sepe, M. A reinforcement learning approach to optimal part flow management for gas turbine maintenance. *Proc. Inst. Mech. Eng. Part O J. Risk Reliab.* **2020**, *234*, 52–62. [CrossRef]
58. Nasir, Y.; He, J.; Hu, C.; Tanaka, S.; Wang, K.; Wen, X.H. Deep Reinforcement Learning for Constrained Field Development Optimization in Subsurface Two-phase Flow. *Front. Appl. Math. Stat.* **2021**, *7*, 689934. [CrossRef]
59. Cheryl, M.; Snab, J.H.; Pierre, N. *Digital Transformation Initiative in Collaboration with Accenture*; World Economic Forum: Geneva, Switzerland, 2018; pp. 1–75.
60. Brennen, J.S.; Kreiss, D. Digitalization. *Int. Encycl. Commun. Theory Philos.* **2016**, 1–11. [CrossRef]
61. Olah, G.A. Beyond Oil and Gas: The Methanol Economy. *Angew. Chemie Int. Ed.* **2005**, *44*, 2636–2639. [CrossRef]
62. Olah, G.A.; Goeppert, A.; Prakash, G.K.S. Beyond Oil and Gas: The Methanol Economy (Google eBook). *ECS Trans.* **2011**, *35*, 31.
63. Barney, J. Firm Resources and Sustained Competitive Advantage. *J. Manag.* **1991**, *17*, 99–120. [CrossRef]
64. Peteraf, M.A. The cornerstones of competitive advantage: A resource-based view. *Strateg. Manag. J.* **1993**, *14*, 179–191. [CrossRef]
65. Aleklett, K.; Campbell, C.J. The Peak and Decline of World Oil and Gas Production. *Miner. Energy-Raw Mater. Rep.* **2003**, *18*, 5–20. [CrossRef]
66. Yang, X.; Bello, O.; Yang, L.; Bale, D.; Failla, R. Intelligent Oilfield—Cloud Based Big Data Service in Upstream Oil and Gas. In Proceedings of the the International Petroleum Technology Conference, Beijing, China, 26–28 March 2019.
67. Keerqinhu; Qi, G.; Tsai, W.-T.; Hong, Y.; Wang, W.; Hou, G.; Zhu, Z. Fault-Diagnosis for Reciprocating Compressors Using Big Data. In Proceedings of the 2016 IEEE Second International Conference on Big Data Computing Service and Applications (BigDataService), Oxford, UK, 29 March–1 April 2016; pp. 72–81.
68. Layouni, M.; Hamdi, M.S.; Tahar, S. Detection and sizing of metal-loss defects in oil and gas pipelines using pattern-adapted wavelets and machine learning. *Appl. Soft Comput. J.* **2017**, *52*, 247–261. [CrossRef]
69. Mazumder, R.K.; Salman, A.M.; Li, Y. Failure risk analysis of pipelines using data-driven machine learning algorithms. *Struct. Saf.* **2021**, *89*, 102047. [CrossRef]
70. Pandya, D.; Srivastava, A.; Doherty, A.; Sundareshwar, S.; Needham, C.; Chaudry, A.; Krishnalyer, S. Increasing production efficiency via compressor failure predictive analytics using machine learning. *Proc. Annu. Offshore Technol. Conf.* **2018**, *1*, 47–55. [CrossRef]
71. Mohamed, A.; Hamdi, M.; Tahar, S. A Machine Learning Approach for Big Data in Oil and Gas Pipelines. In Proceedings of the 2015 3rd International Conference on Future Internet of Things and Cloud, Rome, Italy, 24–26 August 2015.
72. SIngh, S.S.P.; Agarwal, J. *Offshore Operations and Engineering*; CRC Press: Boca Raton, FL, USA, 2019; ISBN 9780429354533.
73. Schaap, D.M.A. SIMORC-System of industry metocean data for the offshore and research communities. In Proceedings of the OCEANS 2007-Europe, Aberdeen, UK, 18–21 June 2007; pp. 1–2.

74. Vladymyrov, V. EC Project: System of Industry Metocean Data for the Offshore and Research Communities (SIMORC). In Proceedings of the Nineenth Session of IOC Committee on International Oceanographic Data and Information Exchange (IODE-XIX), Trieste, Italy, 12–16 March 2007.
75. Park, S.B.; Shin, S.Y.; Shin, D.G.; Jung, K.H.; Choi, Y.H.; Lee, J.; Lee, S.J. Extreme Value Analysis of Metocean Data for Barents Sea. *J. Ocean Eng. Technol.* **2020**, *34*, 26–36. [CrossRef]
76. MetOcean Telematics. Track, Monitor, and Explore without Limits. Available online: https://metocean.com/ (accessed on 20 March 2022).
77. Performance Management and Delivery Unit (PEMANDU) Economic Transformation Programme Report. Econ. *Plan. Unit Malaysia 2010, 1–570.* Available online: https://policy.asiapacificenergy.org/sites/default/files/ETP.pdf (accessed on 2 February 2022).
78. Isa, E.M.; Salim, S.M.; Ghani, A.M.; Karim, I.I. Digital Transformation of PETRONAS' Hydrocarbon Production Allocation System (HyPAS). In Proceedings of the SPE Kingdom of Saudi Arabia Annual Technical Symposium and Exhibition, Dammam, Saudi Arabia, 23–26 April 2018.
79. PETRONAS. *Petronas Activity Outlook 2019–2021*; 2018: PETRONAS Carigali Samarang Platform, Offshore Sabah, Malaysia; PETRONAS Global: Kuala Lumpur, Malaysia, 2021.
80. El-Reedy, M.A. Chapter 8—Risk-Based Inspection Technique. In *Offshore Structure Design, Construction and Maintenance*; El-Reedy, M.A., Ed.; Gulf Professional Publishing: Boston, MA, USA, 2012; pp. 563–634. ISBN 978-0-12-385475-9.
81. Lai, V.; Najah, A.-M.; Abdul Malek, M.; Afan, H.; Ibrahim, R.; El-Shafie, A.; El-Shafie, A. Modeling the Nonlinearity of Sea Level Oscillations in the Malaysian Coastal Areas Using Machine Learning Algorithms. *Sustainability* **2019**, *11*, 4643. [CrossRef]
82. Danyaro, K.U.; Liew, M.S. Semantic Web for Meteorological and Oceanographic Data. In Proceedings of the 2022 2nd International Conference on Computing and Information Technology (ICCIT), Tabuk, Saudi Arabia, 25–27 January 2022; pp. 304–309.
83. Qiao, C.; Myers, A.T.; Arwade, S.R. Validation of a 37-year Metocean Hindcast along the U.S. Atlantic Coast. In Proceedings of the 13th International Conference on Applications of Statistics and Probability in Civil Engineering(ICASP13), Seoul, Korea, 26–30 May 2019; pp. 1–8.
84. Dyer, A.S.; Zaengle, D.; Nelson, J.R.; Duran, R.; Wenzlick, M.; Wingo, P.C.; Bauer, J.R.; Rose, K.; Romeo, L. Applied machine learning model comparison: Predicting offshore platform integrity with gradient boosting algorithms and neural networks. *Mar. Struct.* **2022**, *83*, 103152. [CrossRef]
85. Wyatt, L.R. Progress towards an HF Radar Wind Speed Measurement Method Using Machine Learning. *Remote Sens.* **2022**, *14*, 2098. [CrossRef]
86. Park, S.B.; Shin, S.Y.; Jung, K.H.; Lee, B.G. Prediction of Significant Wave Height in Korea Strait Using Machine Learning. *J. Ocean Eng. Technol.* **2021**, *35*, 336–346. [CrossRef]
87. Martinez-Perurena, A.; Aizpurua Unanue, J.; Penalba, M. Machine-Learning-based Long-Term Forecasting of Metocean Data for the Design of Marine Renewable Energy Systems. In Proceedings of the 16th Conference on Sustainable Development of Energy, Water and Environment Systems (SDEWES), Dubrovnik, Croatia, 10–15 October 2021.
88. Ricky, S.K.; Rahim, L.A. Metocean Prediction using Hadoop, Spark & R. In Proceedings of the 2021 International Conference on Computer & Information Sciences (ICCOINS), Kuching, Malaysia, 13–15 July 2021; pp. 312–315.
89. Kalyuzhnaya, A.V.; Nikitin, N.O.; Vychuzhanin, P.; Hvatov, A.; Boukhanovsky, A. Automatic Evolutionary Learning of Composite Models with Knowledge Enrichment. In Proceedings of the 2020 Genetic and Evolutionary Computation Conference Companion, Cancun, Mexico, 8–12 July 2020; Association for Computing Machinery: New York, NY, USA, 2020; pp. 43–44, ISBN 9781450371278.
90. Danyaro, K.U.; Jaafar, J.; Liew, M.S. MetOcean data to linked data. In Proceedings of the 2014 International Conference on Computer and Information Sciences (ICCOINS), Kuala Lumpur, Malaysia, 3–5 June 2014; pp. 1–5.
91. Colin, A.; Fablet, R.; Tandeo, P.; Husson, R.; Peureux, C.; Longépé, N.; Mouche, A. Semantic Segmentation of Metoceanic Processes Using SAR Observations and Deep Learning. *Remote Sens.* **2022**, *14*, 851. [CrossRef]
92. Bureau of Safety and Environmental Enforcement (BSEE); Bureau of Ocean Energy Management (BOEM). Incident Archive—Taylor Energy Oil Discharge at MC-20 Site and Ongoing Response Efforts. Available online: https://www.bsee.gov/newsroom/library/incident-archive/taylor-energy-mississippi-canyon/ongoing-response-efforts (accessed on 25 March 2022).
93. Asthana, T.; Krim, H.; Sun, X.; Roheda, S.; Xie, L. Atlantic hurricane activity prediction: A machine learning approach. *Atmosphere* **2021**, *12*, 455. [CrossRef]
94. Lang, X.; Wu, D.; Mao, W. Comparison of supervised machine learning methods to predict ship propulsion power at sea. *Ocean Eng.* **2022**, *245*, 110387. [CrossRef]
95. Coraddu, A.; Lim, S.; Oneto, L.; Pazouki, K.; Norman, R.; Murphy, A.J. A novelty detection approach to diagnosing hull and propeller fouling. *Ocean Eng.* **2019**, *176*, 65–73. [CrossRef]
96. Janas, K.; Milne, I.A.; Whelan, J.R. Application of a convolutional neural network for mooring failure identification. *Ocean Eng.* **2021**, *232*, 109119. [CrossRef]
97. Rawson, A.; Sabeur, Z.; Brito, M. Intelligent Geospatial Maritime Risk Analytics Using The Discrete Global Grid System. *Big Earth Data* **2021**, 1–29. [CrossRef]
98. Hvatov, A.; Maslyaev, M. The data-driven physical-based equations discovery using evolutionary approach. In Proceedings of the GECCO 2020 Companion—Genetic and Evolutionary Computation Conference Companion, Cancun, Mexico, 8–12 July 2020; pp. 129–130.

99. Hejazi, R.; Grime, A.; Randolph, M.; Efthymiou, M. A Bayesian Machine Learning Approach for Efficient Integrity Management of Steel Lazy Wave Risers. In Proceedings of the ASME 2020 39th International Conference on Ocean, Offshore and Arctic Engineering, Virtual, 3–7 August 2020.
100. Syed, F.I.; Alnaqbi, S.; Muther, T.; Dahaghi, A.K.; Negahban, S. Smart shale gas production performance analysis using machine learning applications. *Pet. Res.* **2022**, *7*, 21–31. [CrossRef]
101. Mingyu, S.; Jianjun, W.; Chenggao, Y.; Rong, C.; Yuanyuan, H. Study of Forecasting and Estimation Methodology of Oilfield Development Cost Based on Machine Learning. *Chem. Technol. Fuels Oils* **2021**, *56*, 1000–1019. [CrossRef]
102. García-Alba, J.; Bárcena, J.F.; Pedraz, L.; Fernández, F.; García, A.; Mecías, M.; Costas-Veigas, J.; Samano, M.L.; Szpilman, D. SOSeas Web App: An assessment web-based decision support tool to predict dynamic risk of drowning on beaches using deep neural networks. *J. Oper. Oceanogr.* **2021**, *15*, 1–20. [CrossRef]
103. Maslyaev, M.; Hvatov, A.; Kalyuzhnaya, A. Discovery of the data-driven models of continuous metocean process in form of nonlinear ordinary differential equations. *Procedia Comput. Sci.* **2020**, *178*, 18–26. [CrossRef]
104. Alsaihati, A.; Elkatatny, S.; Mahmoud, A.A.; Abdulraheem, A. Use of Machine Learning and Data Analytics to Detect Downhole Abnormalities while Drilling Horizontal Wells, With Real Case Study. *J. Energy Resour. Technol.* **2020**, *143*, 043201. [CrossRef]
105. Brown, N.; Roubíčková, A.; Lampaki, I.; Macgregor, L.; Ellis, M.; Newton, P. Machine learning on Crays to optimize petrophysical workflows in oil and gas exploration. *Concurr. Comput. Pract. Exp.* **2020**, *32*, e5655. [CrossRef]
106. Fan, S.; Xiao, N.; Dong, S. A novel model to predict significant wave height based on long short-term memory network. *Ocean Eng.* **2020**, *205*, 107298. [CrossRef]
107. Orrù, P.F.; Zoccheddu, A.; Sassu, L.; Mattia, C.; Cozza, R.; Arena, S. Machine Learning Approach Using MLP and SVM Algorithms for the Fault Prediction of a Centrifugal Pump in the Oil and Gas Industry. *Sustainability* **2020**, *12*, 4776. [CrossRef]
108. Aldosari, H.; Elfouly, R.; Ammar, R. Evaluation of Machine Learning-Based Regression Techniques for Prediction of Oil and Gas Pipelines Defect. In Proceedings of the 2020 International Conference on Computational Science and Computational Intelligence (CSCI), Las Vegas, NV, USA, 16–18 December 2020; pp. 1452–1456.
109. Wang, J.; Nadarajah, S.; Wang, J.; Ravikumar, A. A Machine Learning Approach to Methane Emissions Mitigation in the Oil and Gas Industry. In Proceedings of the NeurIPS 2020 Workshop Tackling Climate Change with Machine Learning, Virtual, 11 December 2020.
110. Alharam, A.; Almansoori, E.; Elmadeny, W.; Alnoiami, H. Real Time AI-Based Pipeline Inspection using Drone for Oil and Gas Industries in Bahrain. In Proceedings of the 2020 International Conference on Innovation and Intelligence for Informatics, Computing and Technologies (3ICT), Sakheer, Bahrain, 20–21 December 2020; pp. 1–5.
111. Deeva, I.; Nikitin, N.O.; Kaluyzhnaya, A. V Pattern Recognition in Non-Stationary Environmental Time Series Using Sparse Regression. *Procedia Comput. Sci.* **2019**, *156*, 357–366. [CrossRef]
112. Granado, I.; Basurko, O.C.; Rubio, A.; Ferrer, L.; Hernández-González, J.; Epelde, I.; Fernandes, J.A. Beach litter forecasting on the south-eastern coast of the Bay of Biscay: A bayesian networks approach. *Cont. Shelf Res.* **2019**, *180*, 14–23. [CrossRef]
113. Abbas, A.; Bashikh, A.; Abbas, H.; Mohammed, H. Intelligent decisions to stop or mitigate lost circulation based on machine learning. *Energy* **2019**, *183*, 1104–1113. [CrossRef]
114. Khan, M.R.; Alnuaim, S.; Tariq, Z.; Abdulraheem, A. Machine Learning Application for Oil Rate Prediction in Artificial Gas Lift Wells. In Proceedings of the SPE Middle East Oil and Gas Show and Conference, Manama, Bahrain, 18–21 March 2019.
115. Sousa, A.L.; Ribeiro, T.P.; Relvas, S.; Barbosa-Póvoa, A. Using Machine Learning for Enhancing the Understanding of Bullwhip Effect in the Oil and Gas Industry. *Mach. Learn. Knowl. Extr.* **2019**, *1*, 994–1012. [CrossRef]
116. Paltrinieri, N.; Comfort, L.; Reniers, G. Learning about risk: Machine learning for risk assessment. *Saf. Sci.* **2019**, *118*, 475–486. [CrossRef]
117. Tings, B.; Bentes, C.; Velotto, D.; Voinov, S. Modelling ship detectability depending on TerraSAR-X-derived metocean parameters. *CEAS Space J.* **2019**, *11*, 81–94. [CrossRef]
118. Vychuzhanin, P.; Hvatov, A.; Kalyuzhnaya, A.V. Anomalies Detection in Metocean Simulation Results Using Convolutional Neural Networks. *Procedia Comput. Sci.* **2018**, *136*, 321–330. [CrossRef]