



TESTING CENTER FOR OIL AND GAS LEMIGAS

SCIENTIFIC CONTRIBUTIONS OIL AND GAS (SCOG)

Vol. 49. No. 3, September: 537 – 565 - <https://journal.lemigas.esdm.go.id>

ISSN: 2089-3361, e-ISSN: 2541-0520

Published by:

LEMIGAS

Development of Machine Learning for Lost Production Opportunity Prediction and Production Optimization Through Workover Scheduling in Pertamina Hulu Sanga Sanga Field, Sanga-Sanga Block of East Kalimantan, Indonesia

Adinsya Yudha^{1, 2*}, Suranto², Frans Richard Kodong², and Muhammad Rahmanaji Soeparmono¹

¹PT Pertamina Hulu Indonesia Regional 3 Zona 9

Marsma R. Iswahyudi Street, Gunung Bahagia, South Balikpapan, Balikpapan City,
East Kalimantan 76114, Indonesia.

²Petroleum Engineering Department, Faculty of Mineral Technology and Energy,

Universitas Pembangunan Nasional "Veteran" Yogyakarta, SWK 104 Street, Condong Catur, Depok,
Sleman, Yogyakarta, Indonesia 55283, Indonesia

*Corresponding author: Adinsya Yudha (adinsya.yudha@gmail.com)

Adinsya Yudha: <https://orcid.org/0009-0000-2998-1621>

Received: 22 May 2026; Revised: in revised form 30 July 2026; Accepted: 20 August 2026

Available: online 30 September 2026.

ABSTRACT - Lost Production Opportunity (LPO) measures production that cannot be realized because a well or facility is constrained. In many field-planning workflows, LPO and workover schedules are still estimated from historical averages and a fixed production uplift for each job. Although practical, these assumptions do not represent the nonlinear interactions among production behavior, equipment failures, downtime, and intervention timing. This work examines whether those interactions can be learned from field data and used to improve the monthly scheduling of rigless workover jobs. Historical production, Planned Downtime (PDT), Unplanned Downtime (UPDT), and operational event records from Field 'YUD' were compiled at a monthly resolution. XGBoost regression models were trained using a chronological data split and evaluated using MAE, RMSE, sMAPE, and actual-versus-predicted plots. The resulting LPO forecasts and feature-importance rankings were then incorporated into a Decision Support System (DSS). By integrating these outputs with workover capacity, expected uplift, job cost, commodity price, and operational constraints, the DSS assigns a monthly risk level and recommends whether workover activity should be increased, selected cautiously, or reduced. The gas and oil models produced sMAPE values of 16.79% and 19.06%, respectively. In the 2026 simulation, DSS-based scheduling increased gas production from 83.43 to 84.56 MMscfd and oil production from 9,650 to 9,805 BOPD relative to the conventional schedule. These results suggest that field data can be used not only to estimate LPO, but also to place rigless workover jobs in months where their production benefit is less exposed to operational disruptions.

Keywords: lost production opportunity (LPO); unplanned downtime (UPDT); machine learning; workover; production planning.

How to cite this article:

Adinsya Yudha, Suranto, Frans Richard Kodong, and Muhammad Rahmanaji Soeparmono, 2026, Development of Machine Learning for Lost Production Opportunity Prediction and Production Optimization Through Workover Scheduling in Pertamina Hulu Sanga Sanga Field, Sanga-Sanga Block of East Kalimantan, Indonesia, Scientific Contributions Oil and Gas, 49 (3) pp. 537-565.



© 2026 The Author(s). Published by LEMIGAS under the Creative Commons Attribution-ShareAlike 4.0 International License (CC BY-SA 4.0)

<https://doi.org/10.29017/scog.v49i3.2150> | 537

1. Introduction

Maintaining production in mature oil and gas fields requires effective coordination between reservoir deliverability, well performance, and surface-facility availability. Production potential cannot be fully realized when wells or facilities experience operational constraints. The associated production shortfall, referred to as Lost Production Opportunity (LPO), is therefore an important consideration in production planning. In this study, LPO comprises losses associated with Planned Downtime (PDT), including scheduled maintenance, shutdowns, and well interventions, and Unplanned Downtime (UPDT), arising from unexpected equipment failures, well problems, and facility disruptions. Distinguishing these components is necessary because scheduled activities and unexpected disturbances require different planning responses. In Field 'PHSS', operated by Pertamina Hulu Indonesia Region 3 Zone 9, production planning must accommodate recurring operational disruptions alongside a limited capacity for rigless workover activities. Conventional planning relies on historical loss estimates, engineering judgment, and assumed production uplift from interventions. Although these inputs provide a practical planning basis, historical averages may not adequately represent changes in event frequency, downtime duration, affected equipment, and production conditions. Moreover, an intervention with promising production potential may deliver a smaller realized benefit when executed during a period of substantial facility constraints. The planning problem therefore involves both estimating future production losses and determining when available workover resources can be used most effectively.

Previous studies have applied data analytics to several aspects of this problem. [Nwaochei et al. \(2019\)](#) examined data-science approaches for prioritizing non-rig well-work candidates, while [Siahaan et al. \(2022\)](#) applied automated analytics and machine learning to generate well-intervention candidates. In a mature-field setting, [Maulana et al., \(2025\)](#) applied Apriori association-rule mining to identify relationships between well-performance indicators, treatment type, and workover success, demonstrating the value of data-driven screening for production optimization. [Popa et al. \(2023\)](#) investigated workover rig scheduling using genetic algorithms. In relation to production losses, [Mat Piah et al. \(2025\)](#) analyzed offshore production-deferment records and highlighted the difficulties of identifying consistent patterns in data affected by operational, mechanical, and external factors. Their findings also emphasized the importance of data quality for reliable interpretation. These studies provide complementary foundations for candidate selection, scheduling, and deferment analysis. However, applying these capabilities to PHSS requires an explicit connection between the monthly production-loss outlook and the timing of rigless interventions. The practical gap addressed in this study is the translation of LPO forecasts into monthly workover decisions that account for both planned and unplanned production-loss exposure. Candidate ranking alone does not establish whether the proposed execution month is favorable, while a loss forecast alone does not specify how intervention activity should be allocated. This connection is particularly relevant where candidate demand exceeds execution capacity and shared facilities constrain production from multiple wells. Considering UPDT without the accompanying PDT burden may also underestimate the production constraints already present in the operating plan.

Extreme gradient boosting (XGBoost) is selected to model nonlinear relationships and interactions within tabular production, downtime, and operational-event records. Its feature-importance outputs provide a basis for examining which recorded variables contribute to the predictions and for supporting engineering interpretation. The contribution of this study is an integrated, field-specific workflow that connects monthly LPO forecasting, the separate consideration of PDT and UPDT, and a monthly Risk Index within a Decision Support System (DSS). The DSS uses this information to guide rigless workover allocation while considering expected production uplift, candidate urgency, and operational capacity. Accordingly, this study aims to evaluate the accuracy of monthly oil and gas LPO forecasts, identify the operational variables most relevant to the predictions, and assess how incorporating these forecasts into a DSS affects rigless workover scheduling. The proposed schedule is evaluated against the conventional planning approach through a comparison of projected production under the stated operating assumptions. The intended contribution is a transparent and reproducible basis for aligning intervention timing with anticipated production-loss exposure, supporting engineering review and monthly production planning in PHSS.

1.1 Background

Rigless workover resources in Field ‘YUD’ are limited, while the number of candidate wells is considerably larger than the number of jobs that can be executed. A useful schedule must therefore consider both the expected production gain and the risk of carrying out a job during a period of high production loss. XGBoost was selected because tree-based boosting can represent nonlinear relationships in tabular operating data and can also rank the features that drive its predictions. Previous oil and gas applications have often treated production forecasting, downtime prediction, and intervention selection as separate problems. Here, they are linked in one DSS: the model forecasts LPO, the forecast is converted into a monthly risk indicator, and the indicator is used to guide rigless workover timing.

1.2 Research problems

The study addresses the following questions:

- How accurately can monthly lost production opportunity (LPO) be forecast from historical operating data using machine learning?
- Which forecast outputs and operational variables can be combined into a practical risk indicator for workover planning?
- How can LPO forecasts be incorporated into a Decision Support System (DSS) to improve rigless workover scheduling and the economic value of oil and gas production?

1.3 Research objectives

The primary objective is to develop a machine-learning model that forecasts LPO in Field ‘YUD’ while retaining the temporal patterns found in the operating data. A second objective is to identify the events and equipment most closely associated with UPDT, so the forecast can be interpreted by engineers rather than used as an unexplained numerical output. The study also converts the forecast into an operational risk indicator and embeds that indicator in a DSS for rigless workover scheduling. The final schedule is assessed by comparing its projected oil and gas production with the output of the conventional planning approach.

1.4 Research scope and limitations

The analysis covers historical oil and gas production, downtime, LPO, intervention, and selected facility records from one field over the period for which consistent data were available. Model development is restricted to XGBoost and does not include a coupled physical reservoir model. Only rigless workover activities are included; rig-based workovers fall outside the scheduling exercise. The resulting DSS is a planning prototype and is intended to complement, rather than replace, engineering review and the field's formal operational approval process.

1.5 Research benefits

The work contributes an applied workflow that connects LPO forecasting with intervention planning. For engineers, it provides a reproducible way to compare monthly workover windows using operational risk and production value. For planners and managers, it offers a transparent basis for reviewing the timing of the medium-term oil and gas program.

1.6 Research gap

Most published upstream applications address production forecasting, decline analysis, equipment reliability, or well-intervention ranking as separate tasks. Fewer studies examine PDT and UPDT together as sources of LPO and then carry the forecast forward into a monthly workover schedule. Field forecasts also continue to rely heavily on averages and manual adjustments, which can obscure changes in operational event patterns. This study addresses that gap by linking an LPO forecast derived from historical operating data to a risk-based scheduling rule.

2. Theoretical background

2.1 Concept of oil and gas production planning

An upstream production system extends from the reservoir to the point at which hydrocarbons are processed and transported. Reservoirs, completions, artificial-lift equipment, flowlines, manifolds, separators, compressors, and other surface facilities operate as one connected system. A restriction in any part of that chain can change the field rate, which is why reservoir behavior and surface-facility performance should be evaluated together (Allan et al., 2016; Natural Gas Market Module of the National Energy Modeling System: Model Documentation 2025).

From a reservoir-engineering perspective, production planning must balance fluid deliverability, recovery, pressure support, integrity, and facility capacity (Smithson 2016). Integrated Asset Modeling provides one way to examine these interactions when preparing forecasts and comparing development or operating scenarios (Allan et al., 2016). Planning then translates the technical potential into executable targets: long-term development volumes, annual and monthly commitments, and short-term decisions on downtime and well intervention (Hadi et al., 2008; Teh et al., 2008). The quality of those targets depends on the underlying data and on whether known subsurface, well, network, facility, and economic constraints are represented. For Field ‘YUD’, this makes PDT and UPDT part of the forecast itself rather than adjustments made only after the production target has been set.

2.2 Lost production opportunity (LPO) and unplanned downtime (UPDT)

LPO is the difference between the production that could have been delivered under normal conditions and the volume actually realized while a well or facility was constrained. Its magnitude depends on the affected production capacity as well as the frequency and duration of the event. UPDT covers shutdowns that were not included in the operating plan, including equipment failure, instrument malfunction, process disturbance, well-integrity problems, and power-supply interruption. Because the same duration of downtime can have very different consequences at a high-rate and a low-rate asset, identifying the event cause and affected equipment is as important as recording the number of hours offline. This event-level view is consistent with the reliability studies reported by Asihono & Mulyono (2023) and Wisnu & Ros (2024).

2.3 Workover and well intervention in production context

Production from a mature field decline for several reasons, including depletion, changes in produced fluids, formation damage, integrity problems, and restrictions in the production system. New drilling is one response, but a large part of the remaining potential must be recovered from existing wells. Well intervention covers operations that change a well's mechanical, hydraulic, or flowing condition; a workover generally involves more extensive access to the wellbore or a material change to the completion or production system (von Flatern 2014). Typical candidates include wells affected by scale, sand, liquid loading, artificial-lift failure, tubing leaks, or completion damage (Okon 2021). Selecting a candidate is not simply a matter of choosing the steepest decline. Expected uplift, diagnosis, probability of success, job duration, resource availability, cost, safety, and the production deferred during execution all affect the decision (Popa et al., 2023). The LPO and UPDT forecasts in this work are used to add the timing of operational risk to that selection process (Siahaan et al., 2022).

2.4 Rigless workover

At PT Pertamina Hulu Sanga Sanga, the term rigless workover refers to interventions performed without a conventional workover rig. Depending on well condition and job complexity, this route can shorten mobilization and execution time and reduce cost, which is particularly useful in a mature field with many candidates (Wijaya et al., 2020). The activity may involve installing or replacing downhole equipment, diagnostic logging, well cleanout, selected remedial work, access to a productive interval, or a limited completion modification. Not every well is technically suitable for a rigless method; candidate screening is therefore retained in the scheduling workflow.

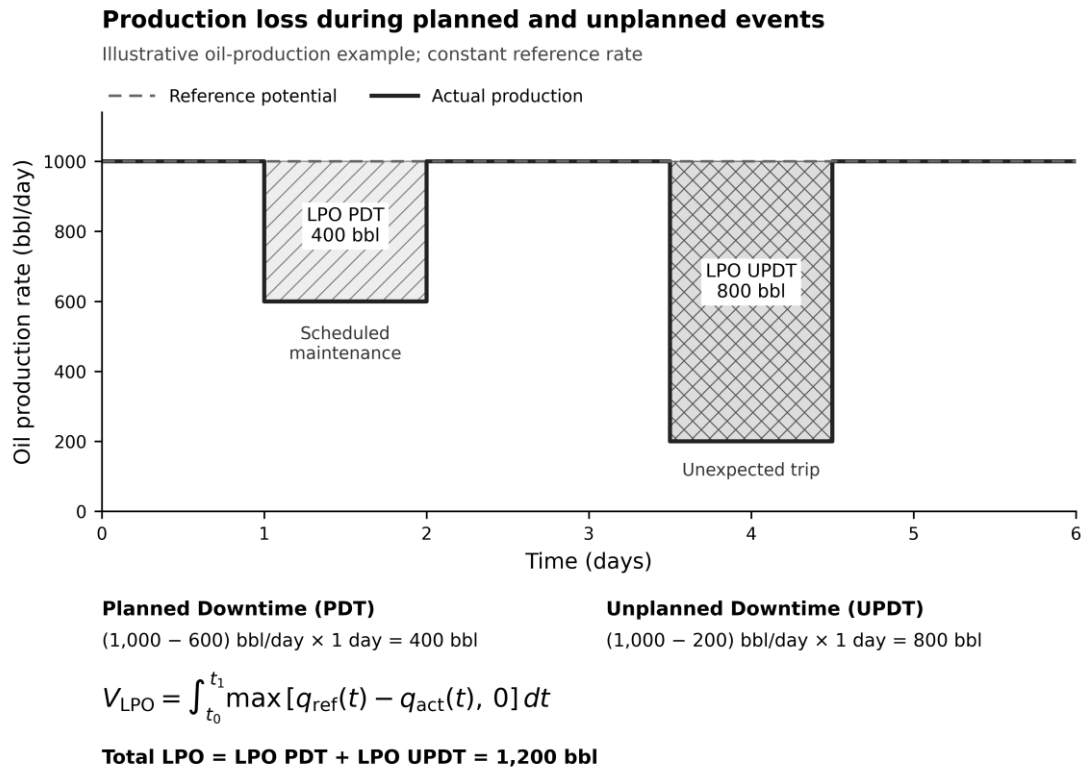


Figure 1.

Conceptual illustration of Lost Production Opportunity associated with Planned Downtime (LPO PDT) and Unplanned Downtime (LPO UPDT). Hatched areas represent the production volumes not realized during each event. Original schematic prepared for this study based on the production-loss concepts described by Feng et al. (2013)

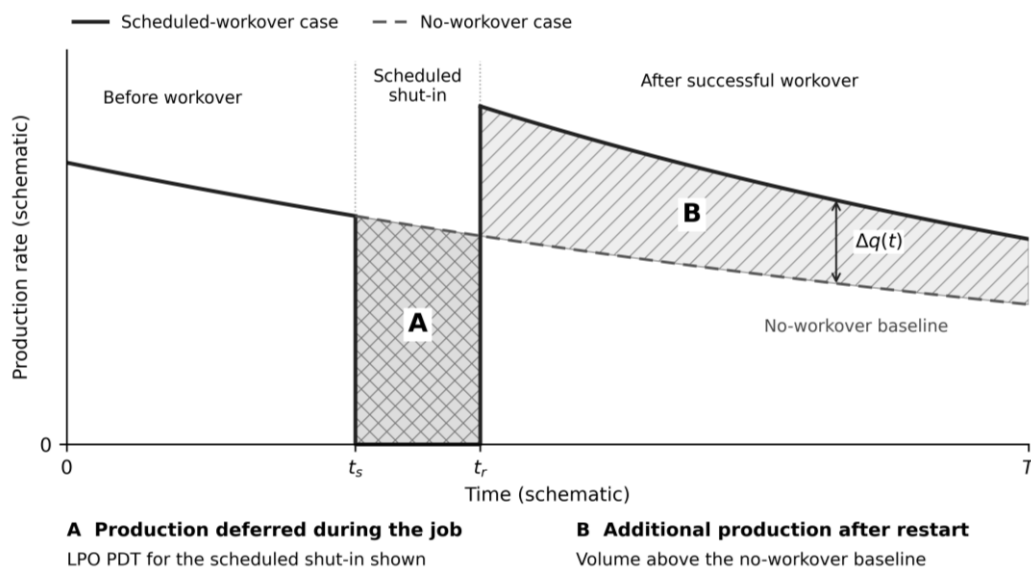


Figure 2.

Conceptual illustration of well intervention and workover to restore or improve production in mature fields (Popa et al., 2023).

2.5 Machine learning in the context of LPO prediction and oil and gas production DSS

Machine learning estimates a response from patterns present in historical data without requiring every relationship to be written as a single physical equation. Upstream applications include production forecasting,

equipment surveillance, predictive maintenance, and operational decision support (Hyoung et al., 2024). LPO is a suitable, although challenging, target because it is shaped by several interacting variables: recent well performance, event frequency, downtime duration, well age, intervention history, and facility condition. The influence of these variables can change over time and is rarely proportional. Rule-based estimates remain useful for screening, but they may miss such interactions when the operating environment changes (Mat et al., 2025b).

2.6 Time series analysis

Time-series observations preserve their order in time, and adjacent values are usually related. A monthly LPO value, for example, may carry information from events, deferment, and production behavior in preceding months. Randomly shuffling these observations would leak future information into model training. For this reason, the data in this study were split chronologically, and lagged or rolling variables were constructed only from information available before the forecast month (Woodward et al., 2017).

2.7 Exploratory data analysis

Exploratory data analysis (EDA) was used before model fitting to determine what the field records actually contained. Distributions and time plots revealed outliers, shutdown periods, and possible seasonal behavior, while boxplots, correlation maps, and autocorrelation plots helped assess relationships and temporal dependence. This step was also used to distinguish genuine high-loss events from missing or inconsistent entries. The findings informed data cleaning and the selection of operational features for the LPO model (Woodward et al., 2017).

2.8 XGBoost characteristics relevant to this study

XGBoost was selected for the following characteristics.

- Nonlinear relationships and variable interactions

The effect of an operational event on LPO is not constant. One additional day of downtime produces a larger loss at a high-rate well, and the uplift from a workover depends on the condition of the well before intervention. XGBoost does not require these relationships to be specified in advance. Its tree structure can split the data at different operating ranges, while boosting combines the splits into a model that represents interactions among event, equipment, and production variables.

- Handling missing values and noisy field data

Operational databases often contain incomplete measurements, inconsistent event descriptions, and changes in reporting practice. XGBoost can assign missing values to a learned default branch within each tree and uses regularization to limit unnecessary complexity. These properties do not remove the need for data quality control, but they make the algorithm practical for moderately noisy field data after the main inconsistencies have been resolved.

- Regularization to prevent overfitting

Unlike classical boosting algorithms, XGBoost explicitly includes L1 and L2 regularization in its objective function:

$$\mathcal{L} = \sum_i l(y_i, \hat{y}_i) + \sum_k \Omega(f_k) \quad (1)$$

with:

$$\Omega(f) = \gamma T + \frac{1}{2} \lambda \|w\|^2 \quad (2)$$

where T , w , γ and λ are the number of leaves, the leaf weight, and regularization parameters, respectively.

Regularization is important in this case because the number of monthly observations is limited relative to the number of possible operational features. Penalizing unnecessary leaves and extreme leaf weights reduces the tendency to reproduce isolated events in the training period. The objective is a model that retains useful patterns when it is applied to a later month, rather than one that only fits the historical sample.

- Computational efficiency and DSS implementation

The selected model also had to operate inside a planning tool. XGBoost trains quickly on the tabular dataset used here, is supported directly in Python, and can generate predictions without a large computing environment. The trained model could therefore be connected to the risk-index and scheduling calculations used by the DSS, rather than remaining a stand-alone academic exercise.

- Model interpretation using feature importance

Feature importance provides an initial view of which variables the fitted model uses most often or most effectively. Gain, cover, and split frequency can be inspected to determine whether the forecast is being driven by plausible production and equipment signals. In this application, those rankings are used to identify recurring LPO drivers and to direct attention to compressors or event types that warrant engineering review. They do not establish causality, but they make the model easier to challenge before its forecast is used in workover planning.

2.9 Decision support system (DSS): definition and decision architecture

A decision support system (DSS) combines information and analytical rules to assist with decisions that cannot be resolved by a fixed procedure alone. Its architecture is commonly described in terms of data-driven, model-driven, and knowledge-driven components. For the present workflow, historical LPO, UPDT events, production, and well status form the data layer. The XGBoost forecast and workover optimizer form the model layer. Operating rules - for example, limiting additional work during a high-risk month and honoring resource capacity - provide the knowledge layer.

2.10 Workover scheduling considering risk index

Workover scheduling determines which candidate is executed and when. In a mature field, the candidate list normally exceeds available crews, equipment, and budget. Production decline is therefore only one part of the ranking; expected value, probability of success, deferment during the job, and execution risk must also be considered. Bringing the forecast and candidate ranking into the same workflow can make the monthly plan more consistent and easier to review (Popa et al., 2023).

The Risk Index used here summarizes several of those considerations in one monthly indicator. Following the usual definition of risk as a combination of likelihood and consequence, the index separates relatively high-, medium-, and low-risk workover windows (Dethlefs & Chastain 2012).

Its inputs are forecasted LPO, PDT, and UPDT; deferment and decline history; expected workover uplift; urgency; job duration; resource demand; and available integrity or equipment information. Together, these variables describe the likely production loss, the value of intervention, the urgency of the candidate, and whether the job is practical within the available window. The index provides an initial ranking, not an automatic authorization to execute a job. The final schedule must still be checked against unit availability, materials, job type, location, simultaneous operations, and other field restrictions. This distinction is important in a mature asset, where a statistically attractive candidate may not be operationally ready (Cio Mario et al., 2019; Nwaochei et al., 2019).

3. Methodology

The research workflow consists of seven main stages. First, historical production, LPO, PDT, UPDT, intervention, and equipment-event data were collected and integrated. Second, the data were cleaned, standardized, and aligned chronologically. Third, relevant model variables were constructed through feature engineering. Fourth, the dataset was divided chronologically into training and testing periods. Fifth, XGBoost regression models were developed and evaluated using MAE, RMSE, sMAPE, and actual-versus-predicted plots. Sixth, the resulting LPO forecasts were converted into monthly Risk Indices. Finally, the Risk Indices were incorporated into the DSS to support rigless workover scheduling and compare the resulting production profile with the conventional schedule.

The workflow in Figure 3 begins with separate treatment of PDT and UPDT, the two components used to construct total LPO. Field records are collected, cleaned, aligned in time, and transformed into model variables. XGBoost is then trained and evaluated on a later test period. The predicted components are combined into total LPO and passed to the Risk Index, which guides the monthly placement of rigless workover activity. Machine learning is therefore one element of the DSS rather than the final output of the study. The workflow continues from prediction to operational interpretation, scheduling, and a comparison of the production value obtained under the proposed and conventional plans.

Figure 4 presents the detailed workflow that extends the general model-development architecture shown in Figure 3. Historical production, PDT- and UPDT-related LPO, and operational-event data are first collected, cleaned, standardized, and chronologically aligned. The processed data are then used to develop separate XGBoost models for gas and oil UPDT-LPO forecasting. After determining the forecasting window and model parameters, the data are divided chronologically into training and testing periods. Model performance is evaluated using MAE, RMSE, sMAPE, and actual-versus-predicted plots. Models with unacceptable errors are recalibrated, whereas validated models are used to generate monthly UPDT-LPO forecasts and identify the most influential recurring operational and equipment-related factors.

General architecture of the study

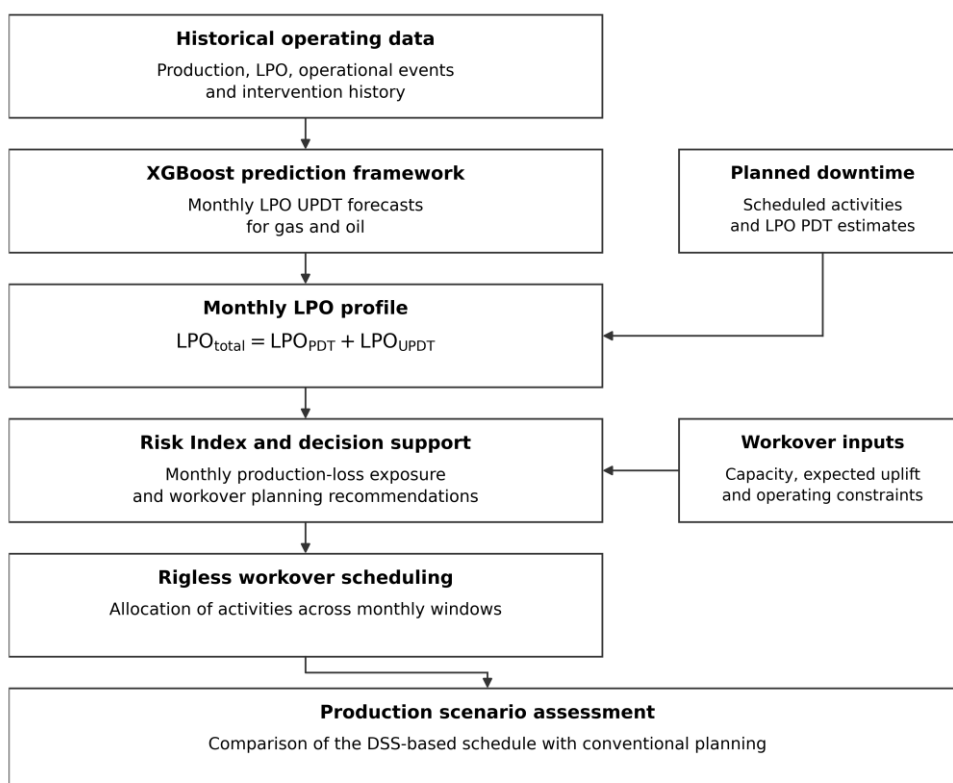


Figure 3.
General analytical architecture of the study, linking historical operating data, LPO forecasting, monthly risk assessment, and rigless workover planning.

The forecasted UPDT-LPO is subsequently combined with scheduled PDT to construct a monthly Operational Risk Index. This index, together with the annual workover plan, expected production uplift, and operational constraints, becomes the input to the Decision Support System. The system reallocates rigless workover activities toward lower-risk months and limits additional interventions during periods of high production-loss exposure. The resulting schedule is translated into three operational recommendations—Boost WO, Normal/Selective WO,

and Hold/Reduce WO—while remaining subject to engineering review, candidate readiness, resource availability, and field-operating constraints. 3 presents the overall analytical architecture of the study, showing the relationship between data preparation, LPO forecasting, Risk Index development, and workover scheduling. Figure 4 provides a more detailed representation of the same workflow, including the individual data-processing, model-validation, forecasting, and decision-support stages.

3.1 Data and field characteristics

The dataset represents a mature oil and gas field and contains production, LPO, UPDT, deferment, intervention history, and selected facility or equipment records. Each source describes a different part of field behavior, so the data were aligned by date, well or operating unit, and production stream before modeling. The assembled series includes routine operating periods as well as shutdowns and unusually high-loss events. Only records with sufficient completeness and consistent definitions were retained for the analysis period. This screening was necessary because reporting formats and event labels changed over time, and some measurements were missing or contained outliers. The final dataset was organized at daily and monthly resolution to preserve event detail while providing the monthly observations required for the workover plan. Production behavior, facility capacity, equipment reliability, and intervention history were then converted into candidate predictors for LPO and the Risk Index.

3.2 Data processing and preprocessing stages

Preprocessing followed the sequence shown in Figure 4: source collection, integration, cleaning, treatment of missing values, standardization of formats and units, transformation, feature engineering, and final validation. The purpose was not to smooth away large operating events, since those events contain the signal of interest, but to separate them from duplicate, incomplete, or incorrectly coded records. Production, PDT, UPDT, deferment, intervention, and available equipment-condition records were obtained from internal field sources. Dates, well names, field identifiers, and operating-unit labels were standardized before the tables were joined. Events were then assigned to the corresponding daily or monthly production period. This alignment made it possible to examine whether a change in production was preceded or accompanied by a recorded disruption or workover activity.

3.3 Research variables and operational definitions

Variables were organized according to their role in the workflow. The first group contains the prediction targets: LPO and its PDT and UPDT components. The second group contains model inputs derived from production, deferment, decline, disruption, and intervention history. The third group contains the operational variables required by the workover Risk Index, including uplift, urgency, duration, resource demand, and integrity or equipment condition. A consistent definition and unit were assigned to each variable before the daily data were aggregated. Total LPO represents the production that was not realized because of a technical, operating, or facility constraint. PDT covers losses associated with activities already included in the plan, such as maintenance, shutdowns, and scheduled well work. UPDT covers losses caused by events that were not scheduled, including equipment failures, well problems, and unexpected facility restrictions.

Predictors include recent production and decline behavior, historical deferment, the frequency and duration of disruptions, and previous well-service or workover activity. Lagged and rolling features were added where they could be calculated without using future information. These variables allow the model to distinguish a recurring loss pattern from an isolated change in rate. The workover Risk Index uses the predicted LPO components together with expected uplift, urgency, job duration, and resource requirements. The variables were chosen to represent the consequence of delaying a job, the potential benefit of executing it, and the practical effort needed to complete it. The two variable groups serve different purposes. Historical operating variables are used to estimate the future LPO profile, whereas candidate and resource variables are introduced only when the forecast is converted into a workover decision. Keeping these stages separate prevents a scheduling preference from being treated as evidence about the physical occurrence of downtime.

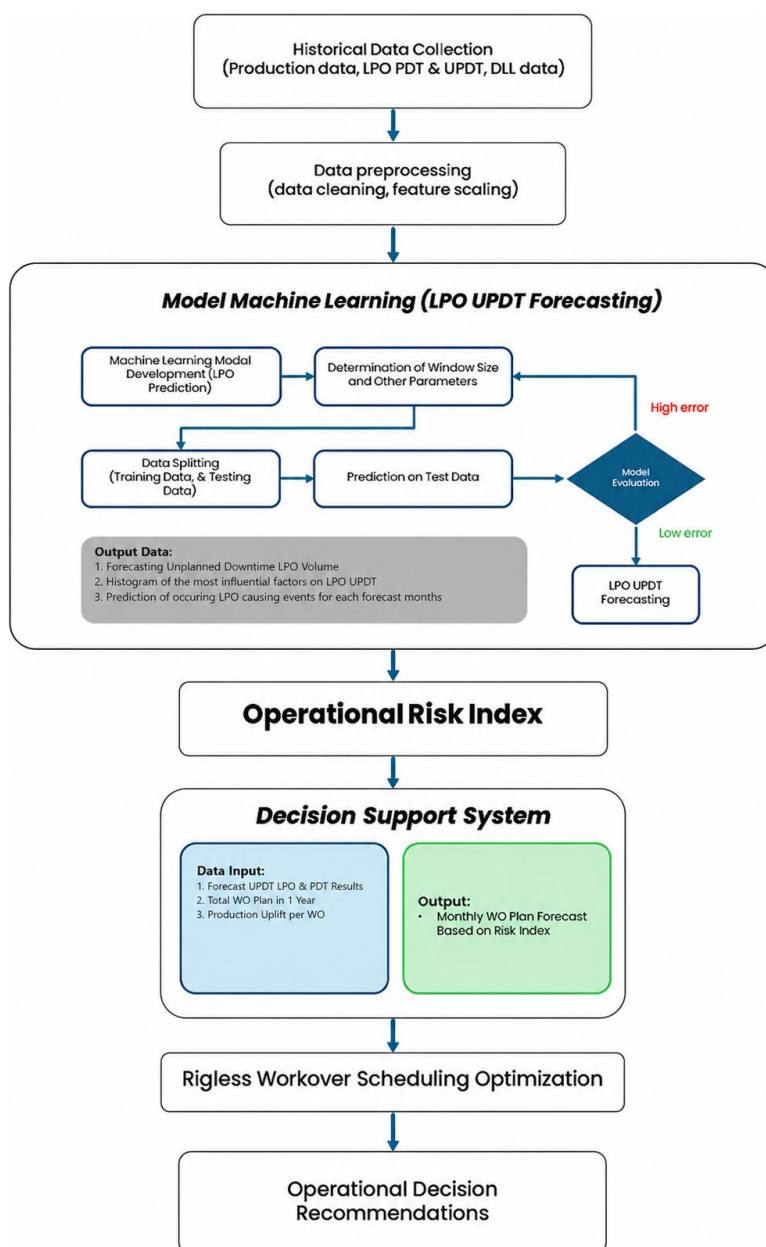


Figure 4.

Detailed flowchart for LPO forecasting and rigless workover scheduling, covering data preprocessing, feature engineering, chronological training–testing separation, XGBoost model development, and performance evaluation

3.4 Prediction model development

PDT and UPDT were modeled separately because their drivers are different. Scheduled maintenance and planned well activity largely determine PDT, while equipment failure, well problems, and unplanned facility restrictions dominate UPDT. Separate predictions also allow planners to see whether a high monthly LPO value is already embedded in the plan or is associated with additional operating uncertainty. The two forecasts were added only after model evaluation to obtain total LPO. The model matrix contained cleaned production and deferment history, decline indicators, event and intervention records, and time-based variables such as lags and moving averages. Training observations preceded the test observations in time; no random split was used. XGBoost regression was fitted to the training period, and its predictions were evaluated with MAE, RMSE, sMAPE, and actual-versus-predicted plots. Feature importance was reviewed alongside the error metrics to determine whether the model relied on credible operating signals. Once validated, the PDT and UPDT forecasts were combined and supplied to the workover Risk Index.

3.5 Model evaluation and workover risk index development

Evaluation was performed before the forecasts entered the scheduling calculation. PDT and UPDT outputs were examined separately, then combined to assess the error and operating meaning of total LPO. This sequence avoids masking a weak component model behind an apparently reasonable aggregate forecast. MAE describes the typical absolute error, RMSE places more weight on large misses, and sMAPE expresses error relative to the scale of actual and predicted values. R^2 was retained as a supplementary measure where appropriate. Plots of actual and predicted values were inspected for missed peaks, timing shifts, and systematic bias. The accepted forecast was then combined with deferment and decline history, uplift, urgency, duration, resource demand, and integrity or equipment status. After scoring and weighting, the resulting index was used to compare candidate months, subject to field constraints and engineering review.

3.6 Model implementation stages

Implementation follows the same sequence used during model development so that every scheduling recommendation can be traced to its inputs. A validated dataset is passed to the fitted XGBoost models, the predicted components are converted to total LPO, and the result is joined to operational and economic workover variables. The DSS then produces a monthly risk class and an associated planning action. For each forecast month, the system first generates PDT and UPDT estimates and records their contribution to total LPO. The Risk Index combines this profile with the candidate-workover information and checks the proposed allocation against annual and monthly capacity. The final step calculates the projected production response and compares it with the conventional schedule, allowing the recommendation to be reviewed in both operational and economic terms.

3.7 Research scenarios and analysis assumptions

Three event-load scenarios were used to test how the forecast responds to changes in disturbance-related inputs. The Base case retains the central input values, while the High and Low cases alter selected event drivers. They are sensitivity cases rather than complete simulations of every field condition; production history, affected units, and other predictors are not all shifted together. The scheduling exercise assumes that total LPO is the sum of separately forecasted PDT and UPDT. It also assumes a finite annual and monthly capacity for rigless workover jobs. Candidate placement therefore seeks a balance among loss risk, expected uplift, execution feasibility, and economic value. Because readiness, integrity, and actual facility conditions can change after the forecast is issued, the DSS output is treated as a planning recommendation rather than an automatic field instruction.

3.8 Results analysis method

Results were reviewed at three levels. First, predictive performance was assessed for the gas and oil models. Second, feature importance and scenario behavior were examined to understand what drove the forecasts. Third, the monthly Risk Index and workover allocation were evaluated against production and capacity outcomes. Model performance was evaluated using MAE, RMSE, and sMAPE were interpreted relative to the scale and unit of each target variable, with lower values indicating better predictive performance (Hyndman & Koehler 2006). For percentage-based forecast errors, as summarized by Montañó Moreno et al. 2013, classified values below 10% as highly accurate, values between 10% and 20% as good, values between 20% and 50% as reasonable, and values above 50% as inaccurate.

Although this classification was originally proposed for MAPE, the present study adopted the 20% threshold as a practical interpretive benchmark for sMAPE rather than as a universal machine-learning standard. Accordingly, the gas and oil models, with sMAPE values of 16.79% and 19.06%, respectively, were considered to provide good predictive accuracy for the monthly planning application. Actual-versus-predicted plots were also examined to identify systematic bias, timing deviations, and missed LPO peaks. Forecast result PDT and UPDT were then aggregated to total LPO, and high-loss months were compared with the event and equipment features ranked by the model. The recommended schedule was finally compared with the conventional plan using projected gas and oil production. This last comparison tests whether an accurate forecast leads to a useful operating decision, rather than treating model error as the only measure of success.

4. Result and discussion

4.1 Overview of research results

The working dataset contained 1,096 daily observations from DailyRate_LPO and 1,489 records from UPDT_Events. After cleaning and monthly aggregation, 36 monthly observations covering January 2023 through December 2025 were available for XGBoost development and testing. The analysis generated four outputs: test-period performance for gas and oil LPO, a 2026 forecast, a monthly Risk Index with workover actions, and a production comparison between the DSS and conventional schedules. The results are reported in that order so the operational recommendation can be traced back to model performance and its dominant inputs.

4.2 XGBoost model validation results

- Gas prediction model performance

For gas LPO, the test-period MAE was 3.13, RMSE was 3.80, and sMAPE was 16.79%. The prediction follows most of the month-to-month movement in the observed series, although the largest misses coincide with abrupt changes in June, August, and November 2025. The error level is acceptable for the monthly planning exercise, but the peak deviations show that isolated disruption events remain difficult to anticipate.

- Oil prediction model performance

The oil model returned an MAE of 4.02, an RMSE of 6.19, and an sMAPE of 19.06%. Its error is higher than that of the gas model because the oil series contains sharper fluctuations and several values close to zero, which also make percentage error more sensitive. The largest departures between actual and predicted LPO occur in December 2024, April 2025, and November 2025.

- Comparison of gas and oil model performance

All three metrics are lower for gas than for oil (Table 1). The difference is consistent with the smoother gas-LPO pattern in the test period and the more intermittent behavior of oil LPO. Both sMAPE values remain below the 20% threshold adopted for this study, so both forecasts were carried forward to the 2026 scenario and scheduling analysis.

Table 1.
XGBoost Model Error Summary.

Component	MAE	RMSE	sMAPE
Gas	3,13	3,80	16,79%
Oil	4,02	6,19	19,06%

4.3 Dominant factors influencing LPO prediction

The gas model relies strongly on event and compressor-condition variables. Low Suction Pressure, Vibration, and Discharge Gas Pressure appear among the leading issue-level features (Figure 6), together with event counts and trip or shutdown indicators. This pattern is operationally plausible because a compressor disturbance can immediately restrict the volume that the gas system is able to deliver. Figure 5 compares the actual and predicted monthly gas UPDT-LPO values using a chronological training–testing split, with the red dashed line indicating the beginning of the out-of-sample testing period. During training, the predictions closely follow the observed monthly fluctuations, indicating that the model captures the recurring temporal and operational patterns in the historical data. In the testing period, the model continues to reproduce the general gas-LPO trend, although larger deviations occur during abrupt changes in June, August, and November 2025. These deviations indicate that isolated operational disruptions remain more difficult to predict than recurring loss patterns. Overall, the gas model achieved an MAE of 3.13 MMSCF, an RMSE of 3.80 MMSCF, and an sMAPE of 16.79%, demonstrating good predictive accuracy for monthly planning based on the adopted error criterion.

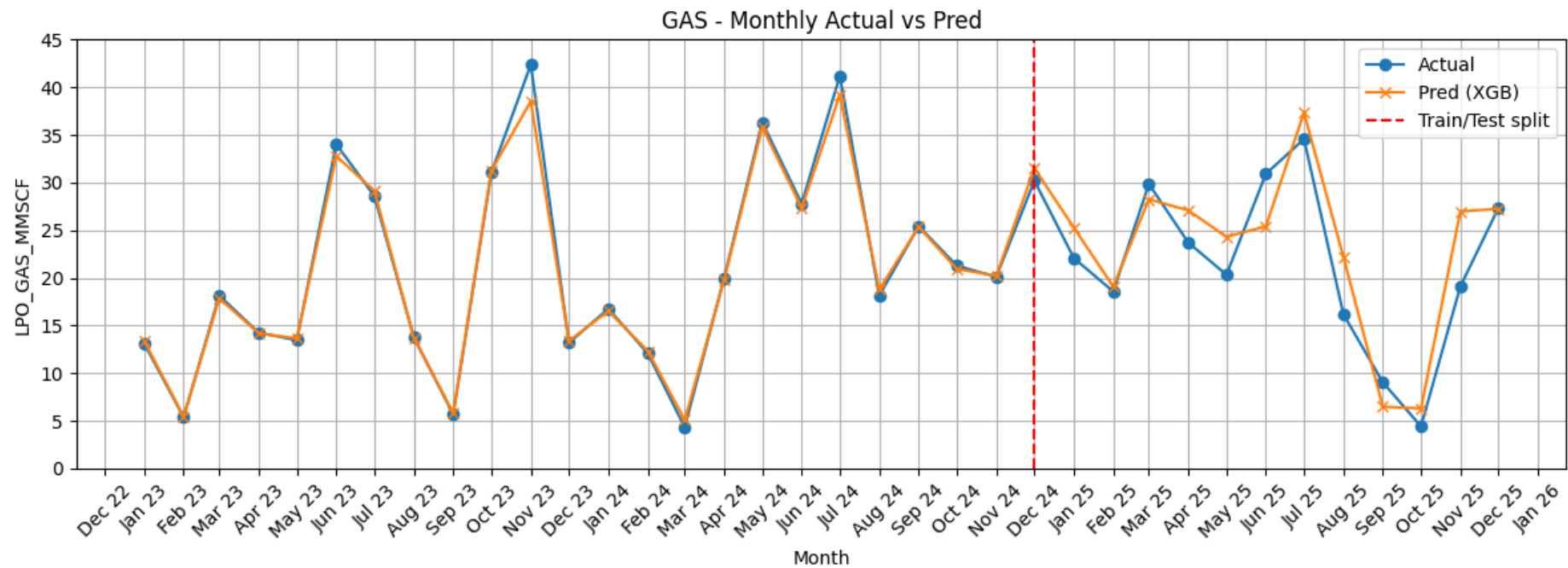


Figure 5.

Comparison of actual and XGBoost-predicted monthly gas LPO associated with unplanned downtime (UPDT). The vertical dashed line separates the training and testing periods, showing the model's fit to historical observations and its predictive performance on subsequent unseen data.

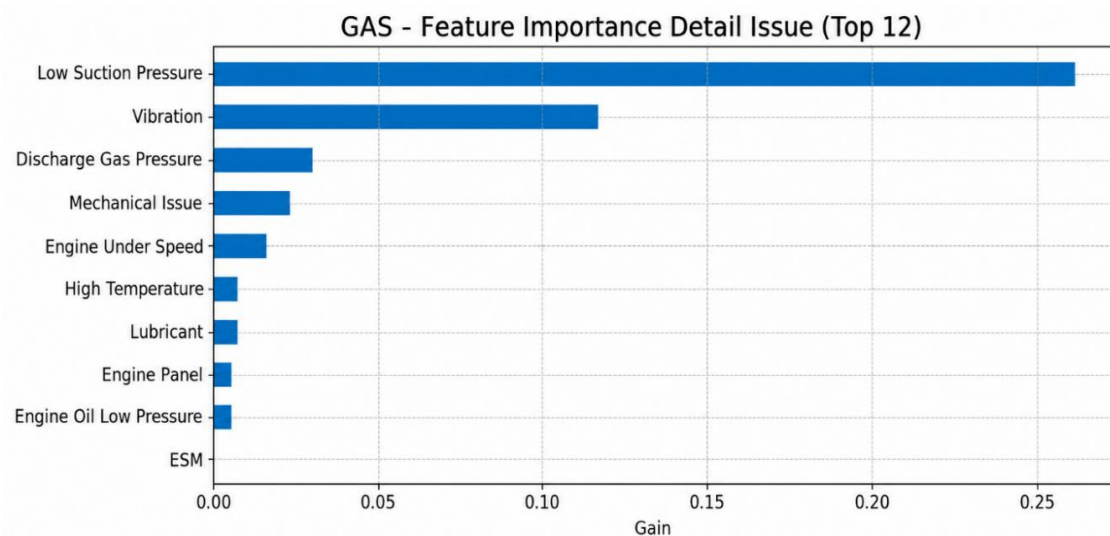


Figure 6.

Ranking of operational-issue features by gain-based importance in the XGBoost model for gas LPO UPDT prediction.

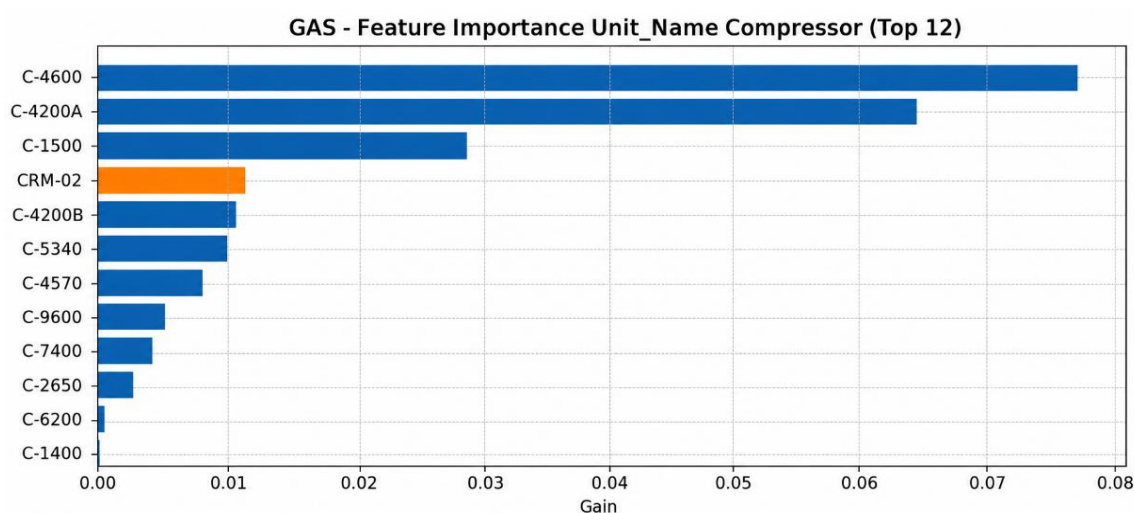


Figure 7.

Gain-based importance of compressor-related features in predicting gas LPO UPDT.

The ranking identifies C-4600, C-4200A, and C-1500 as the compressor identifiers most influential in the model's predictions.

When the compressor name is used as a feature, C-4600, C-4200A, and C-1500 receive the highest importance scores for gas LPO (Figure 7). The ranking does not by itself prove that these units caused every high-LPO month, but it shows that their event histories were repeatedly useful to the model. They should therefore be reviewed alongside maintenance and operating records when the forecast indicates elevated gas risk. Oil LPO shows a similar facility signal, although the ranking is distributed across a different group of variables. C-7400, C-2650, and C-1500 are the leading compressor identifiers, followed by Vibration, Low Suction Pressure, and Engine Oil Low Pressure (Figures 9 and 10). The result suggests that restrictions in shared surface equipment can affect oil realization as well as gas delivery. Figure 8 compares the actual and predicted monthly oil UPDT-LPO values. During the testing period, the model generally follows the observed trend, although larger deviations occur in December 2024, April 2025, and November 2025. The presence of abrupt peaks and near-zero values contributes to the higher RMSE and increases the sensitivity of the percentage-based error. Overall, the oil model achieved an MAE of 4.02, an RMSE of 6.19, and an sMAPE of 19.06%, which remains within the adopted criterion for good monthly predictive accuracy.

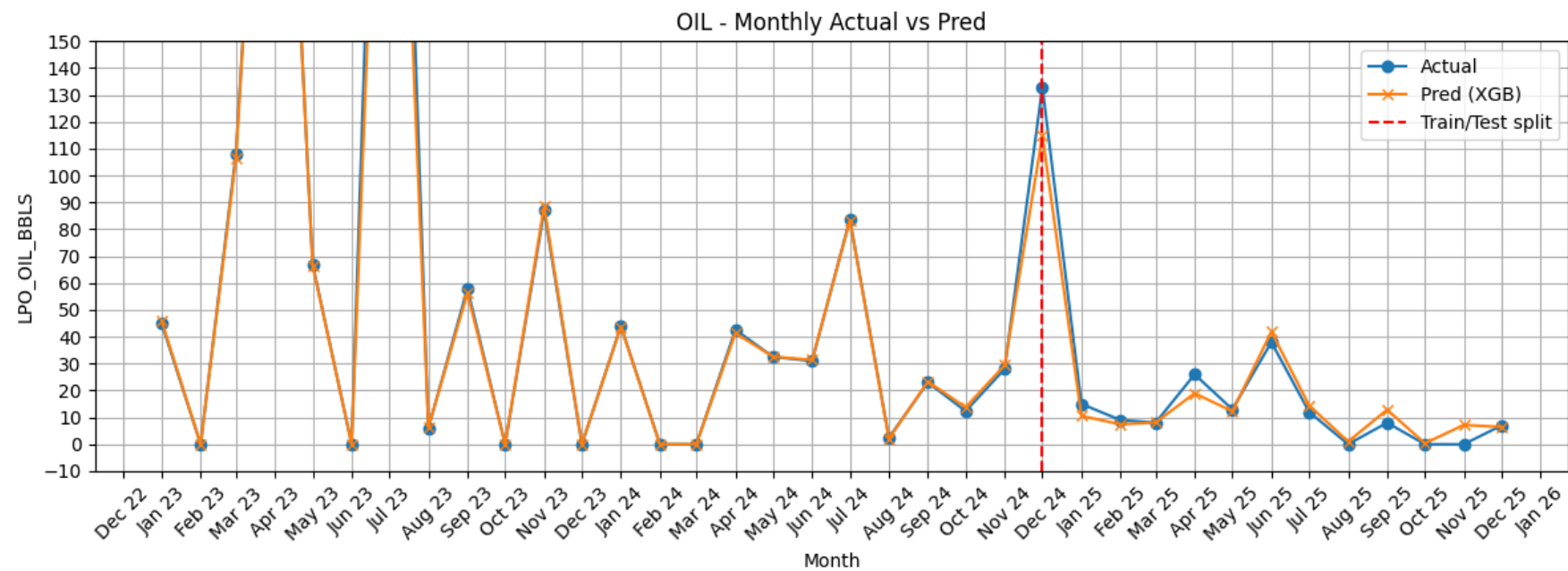


Figure 8.
Comparison of actual and XGBoost-predicted monthly oil LPO associated with unplanned downtime (UPDT).

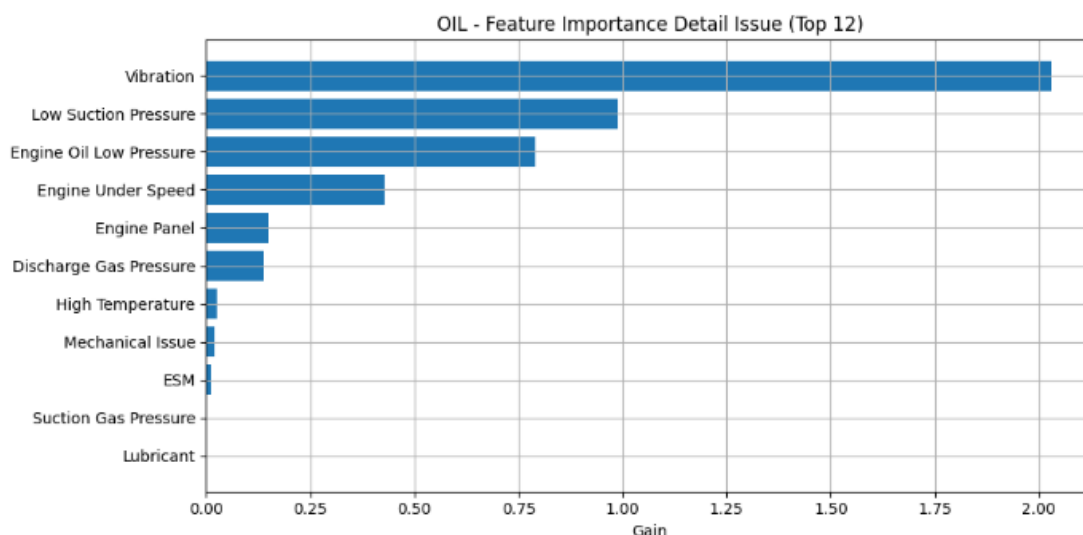


Figure 9.

Ranking of operational-issue features by gain-based importance in the XGBoost model for oil LPO UPDT prediction. Vibration, low suction pressure, and engine oil low pressure are the leading issue categories.

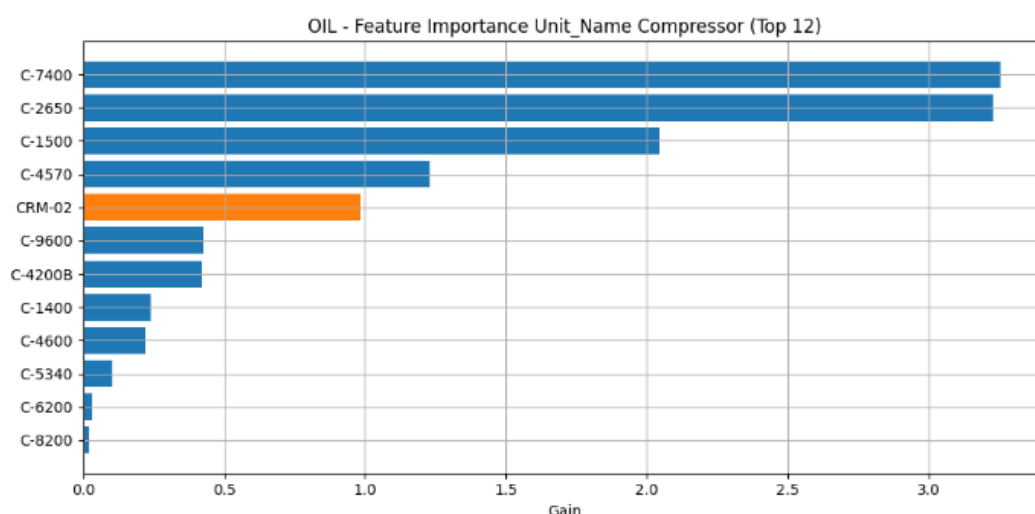


Figure 10.

Top 12 Most Influential Compressors – Oil.

Taken together, the rankings show that the forecast is not being driven by the production series alone. The type of disruption and the unit on which it occurred add useful information, supporting the inclusion of event-level features in the LPO model.

4.4 2026 forecast results

- 2026 gas UPDT forecast

The Base gas-UPDT forecast varies from 6.64 to 35.67 MMSCF per month, with a monthly average of 23.36 MMSCF (Figure 11). July is the highest month, followed by November and December, whereas September has the lowest estimate. The concentration of higher values in the second half of 2026 identifies periods that require closer review before additional production downtime is scheduled. Changing Event Load also moves the forecast in several months, confirming that disturbance intensity contributes to the predicted profile.

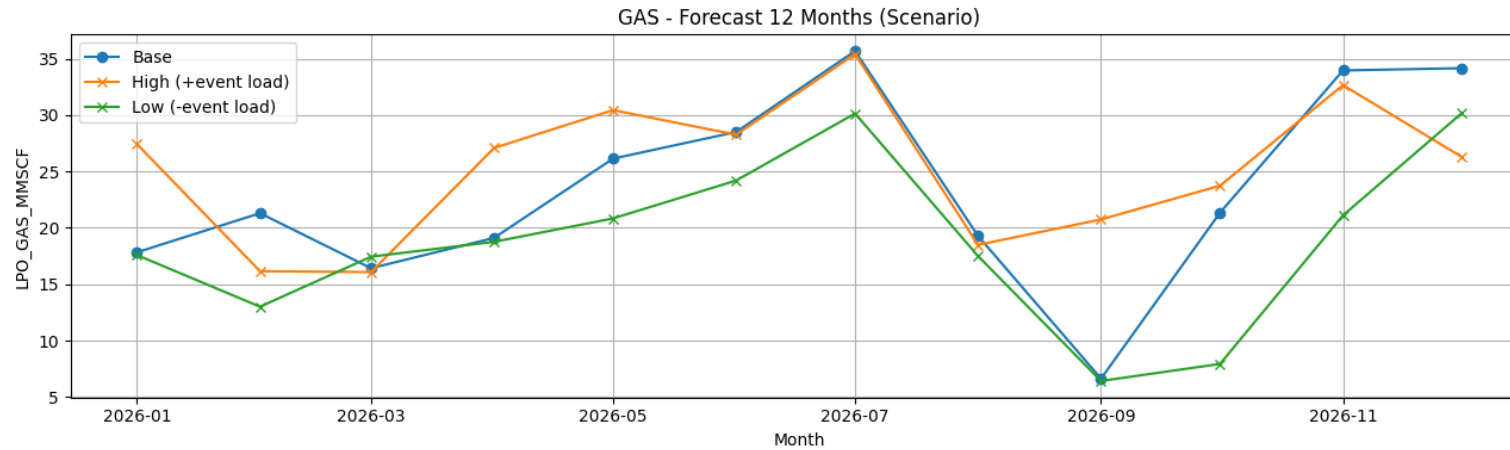


Figure 11.

Monthly gas LPO UPDT forecasts for 2026 under Base, High, and Low operational-event-load scenarios. The comparison illustrates how changes in assumed event load affect the predicted timing and magnitude of unplanned production losses.

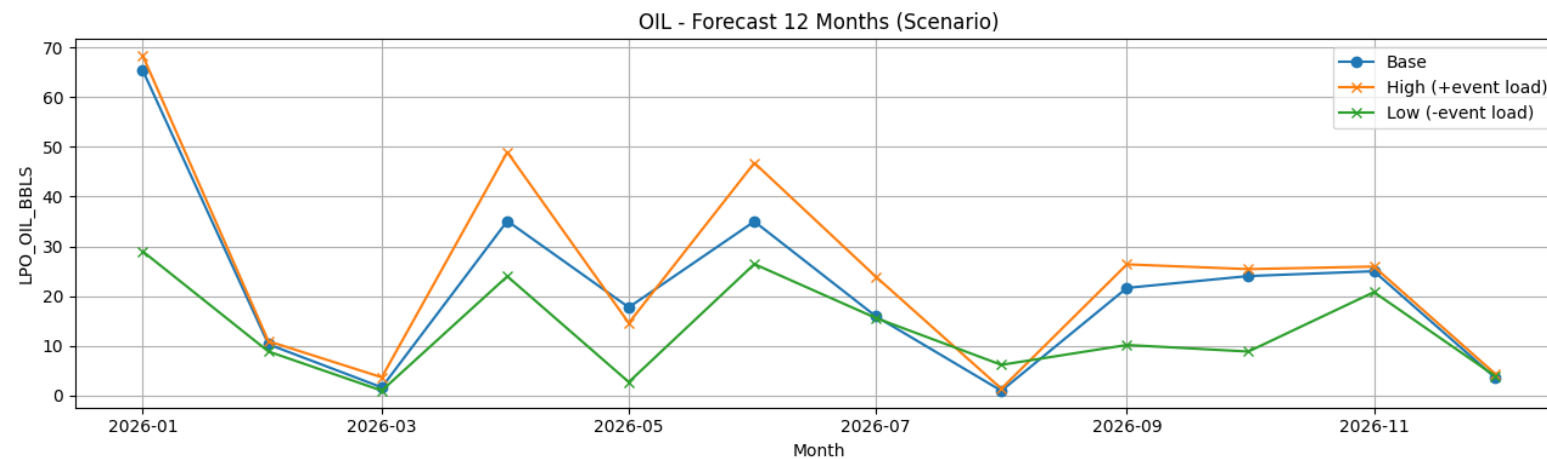


Figure 12.

Monthly gas LPO UPDT forecasts for 2026 under Base, High, and Low operational-event-load scenarios. The comparison illustrates how changes in assumed event load affect the predicted timing and magnitude of unplanned production losses.

- 2026 oil UPDT forecast

Oil UPDT is more variable than gas in the 2026 Base forecast, ranging from 0.99 to 65.32 per month and averaging 21.37 (Figure 12). The maximum occurs in January, with additional high values in April and June. Under the High Event Load case, the largest increases occur in April, June, July, and September. This irregular response is consistent with the more intermittent oil pattern observed during model validation.

- Interpretation of forecast scenario crossing

The High, Base, and Low curves occasionally cross. This behavior follows from the structure of the scenario rather than from a plotting error. XGBoost represents the data with a set of nonlinear tree splits, so changing Event Load can move an observation into a different combination of terminal leaves. At the same time, the scenario changes only selected disturbance variables; deferment history, earlier monthly behavior, affected units, and issue combinations continue to influence the prediction. A High Event Load can therefore produce a lower value than the Base case in a month where the remaining features point to lower loss. No monotonic constraint was imposed, so the model was not forced to increase LPO whenever Event Load increased.

- Total LPO forecast

Total monthly LPO was calculated by adding the forecasted PDT and UPDT components. For gas, the combined forecast totals approximately 1,377.75 MMSCF in 2026 (Figure 13). December has the largest monthly LPO at 224.95 MMSCF, followed by May at 212.88 MMSCF, March at 189.60 MMSCF, and July at 177.41 MMSCF. The ranking differs from the UPDT-only profile because planned downtime contributes materially to the total in several months.

The oil total is dominated by forecasted PDT in several months (Figure 14). May records the highest value at approximately 29,600.76, followed by April at 17,015.10 and December at 7,875.57. The scale and timing of these peaks show that scheduled activity, rather than UPDT alone, determines the principal oil-LPO windows in the 2026 plan.

This distinction matters for workover scheduling. A month with moderate UPDT can still be unsuitable for additional well activity if the existing plan already contains substantial downtime. The scheduling rule therefore uses the combined LPO profile while retaining PDT and UPDT as separate diagnostic inputs.

4.5 Risk Index and WO Scheduling Results

Table 2 links the principal gas and oil LPO drivers with the monthly Risk_Index, the proposed Plan_WO_Perf and Plan_WO_PCTGL allocations, and the resulting action category. It is the point at which the forecast is converted from a production-loss estimate into a planning recommendation. As shown in Figure 15, the number of proposed workover jobs generally falls as the monthly Risk Index increases. Low-risk months absorb more activity, whereas high-risk months retain capacity but avoid adding downtime to an already exposed production period. The allocation is therefore intentionally uneven across the year.

A High-Risk Index denotes a month with greater forecast production-loss exposure. In those months, the DSS recommends holding or reducing workover activity unless a candidate is sufficiently urgent. Lower values create a more favorable window for additional jobs. The decision is not made from the index alone: the driver notes identify the relevant issue type, affected unit, and contribution to gas or oil LPO. The resulting actions are reported as Boost WO, Normal/Selective WO, or Hold/Reduce WO.

- Risk index development results

For gas, December has the highest monthly Risk Index, followed by May, March, and July; these are also the months with the largest combined gas LPO. The oil index peaks in May and then April, reflecting the strong PDT contribution in those months. The agreement between the index ranking and the forecast profile confirms that the scoring step retained the main production-loss signal.

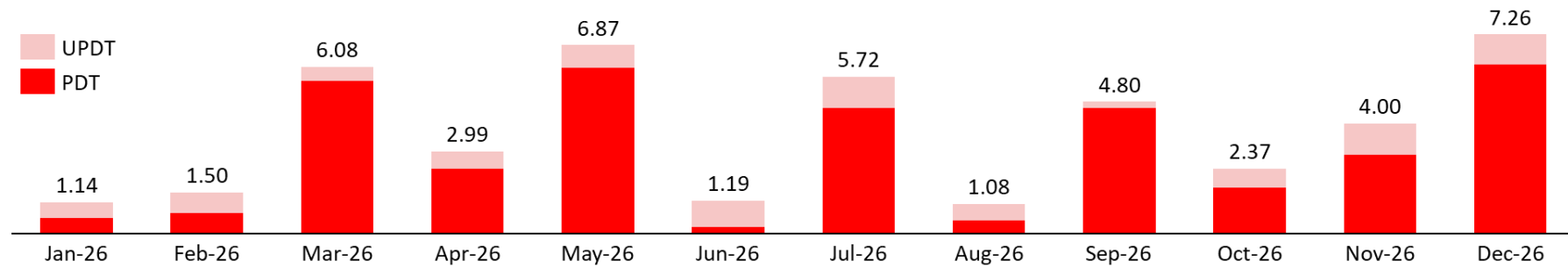


Figure 13.

Monthly total gas LPO forecast for 2026, showing the contributions of planned downtime (PDT) and unplanned downtime (UPDT). The stacked bars identify periods of elevated production-loss exposure relevant to workover planning.

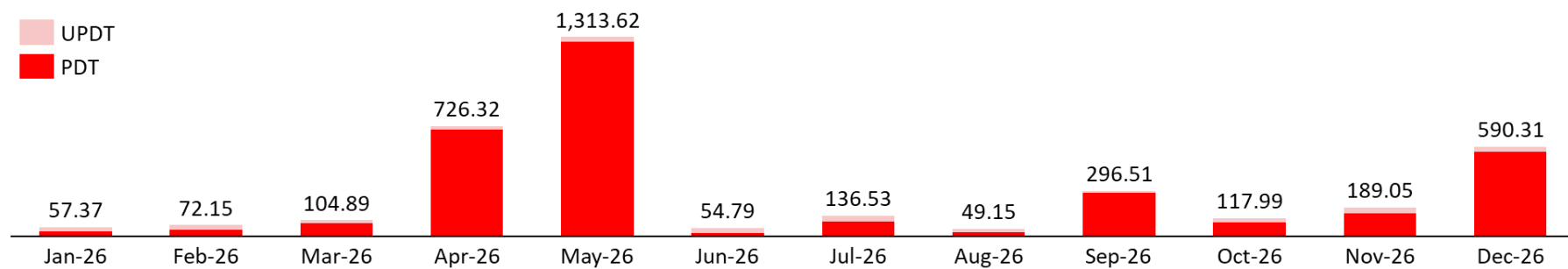


Figure 14.

Monthly total oil LPO forecast for 2026, separated into planned downtime (PDT) and unplanned downtime (UPDT) components.

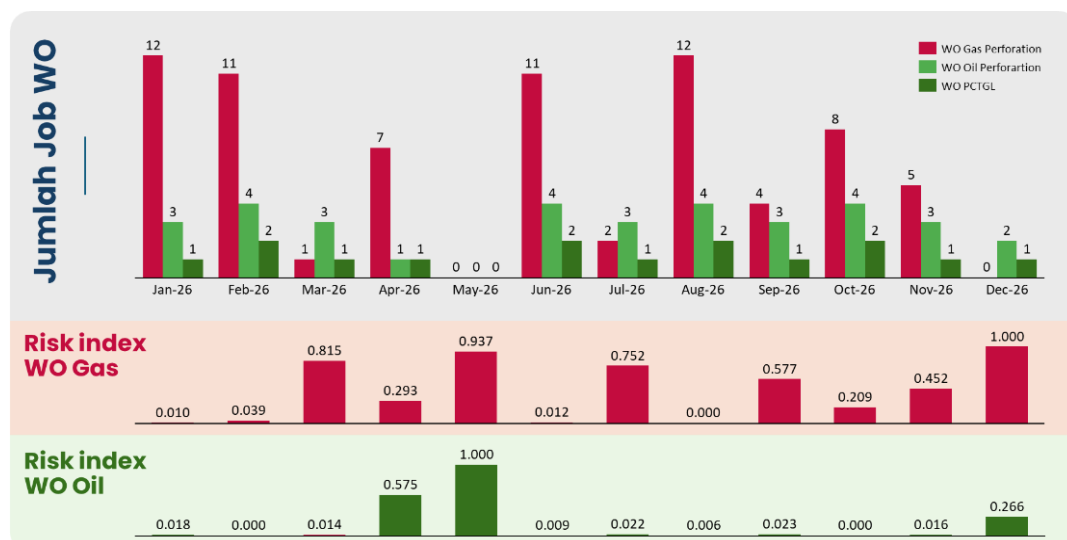


Figure 15.
Proposed monthly allocation of gas perforation, oil perforation, and PCTGL jobs for 2026, alongside the corresponding gas and oil Risk Indices

- Workover scheduling allocation results

The proposed 2026 program contains 73 gas perforation/workover jobs, 34 oil perforation/workover jobs, and 15 oil PCTGL jobs. All remain within the annual capacity assumptions. The busiest month contains 16 performance jobs against a limit of 20, while PCTGL reaches two jobs against a limit of five. Gas workover activity is concentrated in January, February, June, and August; none is assigned in May or December because of the higher monthly risk.

- Planning action interpretation

Boost WO is assigned to the lower-risk windows in January, February, June, August, and October. April, September, and November fall into Normal/Selective WO, which permits execution but requires tighter candidate screening. March, May, July, and December are classified as Hold/Reduce WO because their forecast LPO, particularly for gas, is relatively high. These categories redistribute the same limited capacity; they do not imply that every low-risk candidate should be executed or that urgent integrity work should be deferred.

4.6 Impact of workover scheduling on production performance

The recommended schedule was next evaluated in a production simulation. Two cases were compared: the DSS allocation described above and a conventional schedule that does not adjust monthly job placement using the predicted risk.

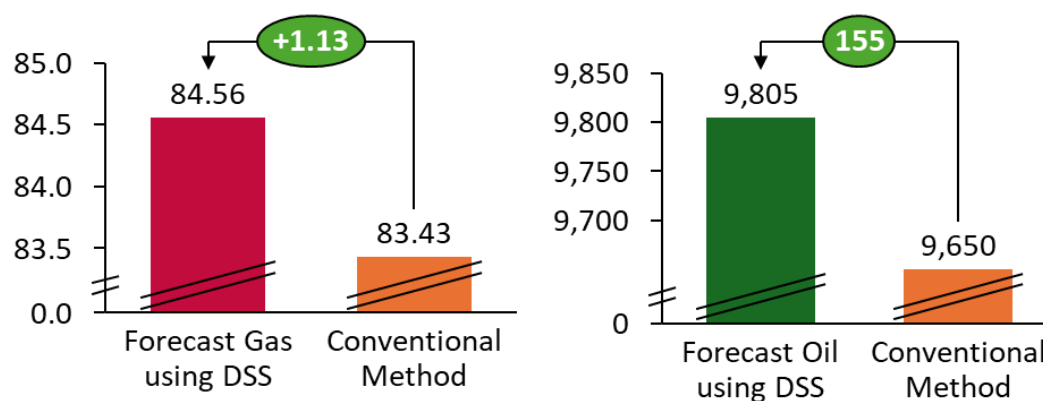


Figure 16.
Forecast Comparison Using DSS and Conventional Method.

Table 2.
Risk Index and Workover Scheduling Results.

Month	Gas driver note total	Oil driver note total	Risk index GAS	Risk index OIL	Plan WO gas Perf	Plan WO oil Perf	Plan WO oil PCTGL	Planning action
Jan-2026	UPDT — Issue: Engine Oil Low Pressure, Vibration, Lubricant Unit: C-1500 (PAM), C-4200A (SEM), C-2650 (SEM) PDT — Issue: C-3600C Engine Exchange, C-1400 PM 4K, C-12000 PM 2K Unit: C-3600C (NLM), C-1400 (SEM), C-12000 (MUT)	UPDT — Issue: Engine Oil Low Pressure, Engine Under Speed, Engine Panel Unit: C-1500 (PAM), C-7400 (PAM), C-2650 (SEM) PDT — Issue: C-12000 PM 2K, BCPHP PM serv 2k, NCPHP PM serv 2K Unit: C-12000 (MUT), BCPHP (BDK), NCPHP (NLM)	0.010	0.018	12	3	1	Boost WO
Feb-2026	UPDT — Issue: Engine Oil Low Pressure, High Temperature, Lubricant Unit: C-4600 (SEM), C-1500 (PAM), C-4200A (SEM) PDT — Issue: C-4570 Booster GOH, C-2800C PM 3K, C-2650 PM 2K Unit: C-4570 (MUT), C-2800C (MUT), C-2650 (SEM)	UPDT — Issue: Engine Under Speed, Engine Oil Low Pressure, ESM Unit: C-1500 (PAM), C-2650 (SEM), C-8200 (MUT) PDT — Issue: C-1280 PM 2K, C-2650 PM 2K, C-2800C PM 3K Unit: C-1280 (MUT), C-2650 (SEM), C-2800C (MUT)	0.039	0.000	11	4	2	Boost WO
Mar-2026	UPDT — Issue: Engine Panel, High Temperature, Low Suction Pressure Unit: C-4200A (SEM), C-1500 (PAM), C-2650 (SEM) PDT — Issue: Retrofit Panel Semberah Plant 14 (C-4500), C-7400 PM 2K, C-1580 PM 2K Unit: C-7400 (MUT), C-1580 (SEM)	UPDT — Issue: Engine Under Speed, ESM, Discharge Gas Pressure Unit: C-7400 (PAM), C-2650 (SEM), C-5340 (NLM) PDT — Issue: C-8300 PM 2K, C-1500 PM 2K, NCPHP PM serv 2K Unit: C-8300 (MUT), C-1500 (PAM), NCPHP (NLM)	0.815	0.014	1	3	1	Hold / reduce WO
Apr-2026	UPDT — Issue: Engine Oil Low Pressure, Mechanical Issue, High Temperature Unit: C-4600 (SEM), C-4200A (SEM), C-4200B (MUT) PDT — Issue: C-7400 GOH, C-12000 PM 2K, C-6700B PM 3K Unit: C-7400 (MUT), C-12000 (MUT), C-6700B (NLM)	UPDT — Issue: Engine Under Speed, Engine Oil Low Pressure, Vibration Unit: C-1500 (PAM), C-2650 (SEM), C-1400 (SEM) PDT — Issue: C-7400 GOH, C-12000 PM 2K, BCPHP PM serv 2k Unit: C-7400 (MUT), C-12000 (MUT), BCPHP (BDK)	0.293	0.575	7	1	1	Normal / selective WO
May-2026	UPDT — Issue: High Temperature, Discharge Gas Pressure, Low Suction Pressure Unit: C-4200A (SEM), C-1500 (PAM), C-2650 (SEM) PDT — Issue: C-12000 TOH, C-2030B PM 8K, C-2700 PM 8K Unit: C-12000 (MUT), C-2030B (NLM), C-2700 (MUT)	UPDT — Issue: Engine Under Speed, ESM, Discharge Gas Pressure Unit: C-7400 (PAM), C-2650 (SEM), C-5340 (NLM) PDT — Issue: C-12000 TOH, C-1500 PM 2K, C-8100 PM serv 2K Unit: C-12000 (MUT), C-1500 (PAM), C-8100 (SEM)	0.937	1.000	0	0	0	Hold / reduce WO
Jun-2026	UPDT — Issue: Engine Under Speed, Engine Oil Low Pressure, High Temperature Unit: C-4600 (SEM), C-4200A (SEM), C-1500 (PAM) PDT — Issue: C-2800C PM 3K, C-8200 PM 3K, C-4800 PM 2K Unit: C-2800C (MUT), C-8200 (MUT), C-4800 (MUT)	UPDT — Issue: Engine Oil Low Pressure, Vibration, ESM Unit: C-1500 (PAM), C-2650 (SEM), C-4570 (MUT) PDT — Issue: C-8300 PM 2K, NCPHP PM serv 2K, C-2800C PM 3K Unit: C-8300 (MUT), NCPHP (NLM), C-2800C (MUT)	0.012	0.009	11	4	2	Boost WO
Jul-2026	UPDT — Issue: Engine Oil Low Pressure, Lubricant, High Temperature Unit: C-4200A (SEM), C-1500 (PAM), C-4200B (MUT) PDT — Issue: C-1400 Exchange Engine, C-7040 PM 8K, C-12000 PM 2K Unit: C-1400 (SEM), C-7040 (MUT), C-12000 (MUT)	UPDT — Issue: Engine Oil Low Pressure, Discharge Gas Pressure, Low Suction Pressure Unit: C-1500 (PAM), C-4200B (MUT), C-9600 (MUT) PDT — Issue: C-12000 PM 2K, C-7400 PM 2K, BCPHP PM serv 2k Unit: C-12000 (MUT), C-7400 (MUT), BCPHP (BDK)	0.752	0.022	2	3	1	Hold / reduce WO
Aug-2026	UPDT — Issue: Lubricant, Low Suction Pressure, Engine Panel Unit: C-4200A (SEM), C-1500 (PAM), C-6200 (MUT) PDT — Issue: C-2500 PM 8K, C-7080 PM 8K, C-6700B PM 3K Unit: C-2500 (MUT), C-7080 (BDK), C-6700B (NLM)	UPDT — Issue: Engine Under Speed, ESM, Lubricant Unit: C-2650 (SEM), C-4200B (MUT), C-5340 (NLM) PDT — Issue: C-1500 PM 2K, C-8100 PM serv 2K, C-2500 PM 8K Unit: C-1500 (PAM), C-8100 (SEM), C-2500 (MUT)	0.000	0.006	12	4	2	Boost WO

Table 2. Risk Index and Workover Scheduling Results. (continued).

Month	Gas driver note total	Oil driver note total	Risk index GAS	Risk index OIL	Plan WO gas Perf	Plan WO oil Perf	Plan WO oil PCTGL	Planning action
Sep-2026	UPDT — Issue: Mechanical Issue, High Temperature, Lubricant Unit: C-4600 (SEM), C-4200A (SEM), C-1500 (PAM) PDT — Issue: C-2700 Booster GOH, C-12000 PM 2K, C-4600 PM 2K Unit: C-2700 (MUT), C-12000 (MUT), C-4600 (SEM)	UPDT — Issue: Engine Under Speed, Vibration, Mechanical Issue Unit: C-7400 (PAM), C-2650 (SEM), C-4200B (MUT) PDT — Issue: C-12000 PM 2K, C-8300 PM 2K, NCPHP PM serv 2K Unit: C-12000 (MUT), C-8300 (MUT), NCPHP (NLM)	0.577	0.023	4	3	1	Normal / selective WO
Oct-2026	UPDT — Issue: Engine Panel, Low Suction Pressure, Vibration Unit: C-4600 (SEM), C-4200A (SEM), C-9600 (MUT) PDT — Issue: C-4800 Engine Exchange, C-2450 PM 8K, C-8200 PM 3K Unit: C-4800 (MUT), C-2450 (SEM), C-8200 (MUT)	UPDT — Issue: Engine Under Speed, Vibration, High Temperature Unit: C-7400 (PAM), C-2650 (SEM), C-9600 (MUT) PDT — Issue: C-1280 PM 2K, C-2450 PM 8K, C-2650 PM 2K Unit: C-1280 (MUT), C-2450 (SEM), C-2650 (SEM)	0.209	0.000	8	4	2	Boost WO
Nov-2026	UPDT — Issue: Engine Under Speed, High Temperature, Engine Panel Unit: C-4200A (SEM), C-1500 (PAM), C-2650 (SEM) PDT — Issue: C-2800B GOH, C-5500A TOH, C-4800 PM 2K Unit: C-2800B (MUT), C-5500A (NLM), C-4800 (MUT)	UPDT — Issue: Vibration, ESM, Discharge Gas Pressure Unit: C-1400 (SEM), C-7400 (PAM), C-2650 (SEM) PDT — Issue: C-7400 PM 2K, C-1500 PM 2K, Pencegahan LPO per tahun sebesar 26.59 MMSCF & 2933 BBL dengan penggantian jalur pipa flowline ke sumur injeksi di Badak Plant Unit: C-7400 (MUT), C-1500 (PAM)	0.452	0.016	5	3	1	Normal / selective WO
Dec-2026	UPDT — Issue: Low Suction Pressure, Discharge Gas Pressure Unit: C-4600 (SEM), C-4200A (SEM), C-1500 (PAM) PDT — Issue: C-2800C GOH, C-8300 Booster GOH, C-2600 PM 8K Unit: C-2800C (MUT), C-8300 (MUT), C-2600 (MUT)	UPDT — Issue: Engine Under Speed, Low Suction Pressure, High Temperature Unit: C-2650 (SEM), C-5340 (NLM), C-8200 (MUT) PDT — Issue: C-8300 Booster GOH, C-12000 PM 2K, BCPHP PM serv 2k Unit: C-8300 (MUT), C-12000 (MUT), BCPHP (BDK)	1.000	0.266	0	2	1	Hold / reduce WO

Figure 16 compare production forecast result using decision support system (DSS) vs conventional method. Conventional case produces 83.43 MMscfd of gas and 9,650 BOPD of oil in the 2026 simulation. Using the LPO forecast, Risk Index, and monthly workover actions shifts jobs toward lower-risk windows and raises the projected rates to 84.56 MMscfd and 9,805 BOPD. The difference is 1.13 MMscfd of gas and 155 BOPD of oil. Because the number of jobs is still constrained, the improvement is attributed mainly to timing rather than to adding unlimited intervention activity.

The monthly plots show that most of the production improvement occurs when jobs are moved into low-risk periods and removed from high-risk periods. Job count alone is therefore an incomplete measure of schedule quality; the background LPO burden in the execution month also matters. This timing strategy is feasible in Field ‘YUD’ because the rigless contract is charged per executed job. Reducing activity in one month does not create a direct standby cost, so a Hold/Reduce action can avoid incremental deferment without leaving a paid unit idle.

Figure 17 and Table 3 illustrate how the gas LPO profile is translated into monthly workover decisions. The upper bars in Figure 17 compare the gas production rates generated by the conventional and DSS-based schedules, while the lower bars represent the PDT and UPDT contributions to monthly LPO. As summarized in Table 3, March, May, July, and December record relatively high gas Risk Indices of 0.82, 0.94, 0.75, and 1.00, respectively, and are therefore classified as Hold/Reduce WO periods with limited or no perforation jobs. In contrast, January, February, June, August, and October are classified as Boost WO periods and receive higher workover allocations. April, September, and November are categorized as Normal/Selective WO periods. This risk-based redistribution increases the projected average gas production from 83.43 to 84.56 MMscfd, representing an improvement of 1.13 MMscfd over the conventional schedule.

Figure 18 and Table 4 present the corresponding results for oil production. The largest oil LPO values occur in May, April, and December, at approximately 1,314, 726, and 590 bbl, respectively. Accordingly, Table 4 assigns no perforation or PCTGL jobs in May, while workover activity is limited in April and December. Higher numbers of perforation and PCTGL jobs are allocated to months with lower production-loss exposure, particularly February, June, August, and October. The operational recommendations in Table 4 reflect the combined field-level risk assessment; therefore, the final action category considers both gas and oil conditions rather than the oil Risk Index alone. Although the DSS profile is slightly lower than the conventional profile in several months because of job rescheduling, the production benefits are recovered in subsequent lower-risk periods. Overall, the DSS increases the projected average oil production from 9,650 to 9,805 BOPD, corresponding to an improvement of 155 BOPD. These results demonstrate that workover performance depends not only on the total number of jobs but also on their timing relative to the monthly PDT and UPDT profiles.

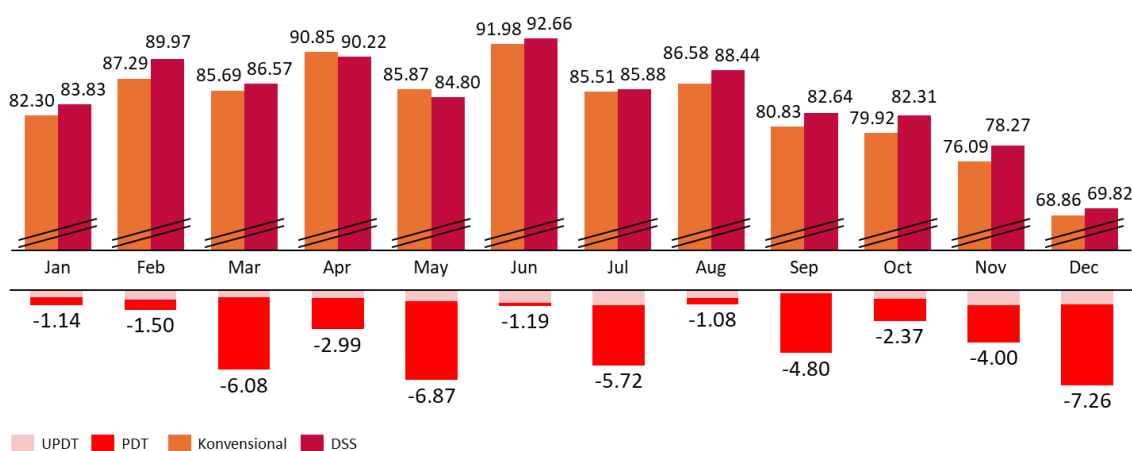


Figure 17.
Monthly Gas Production and LPO Chart.

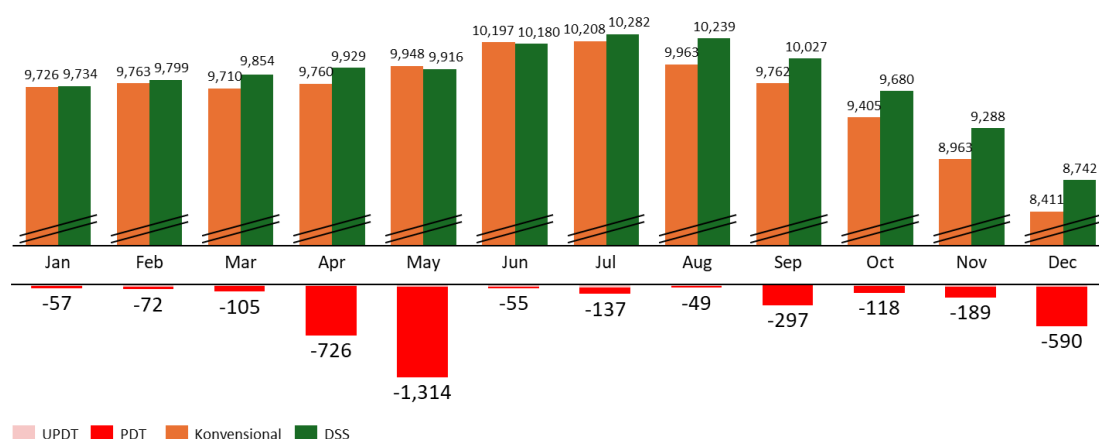


Figure 18.
Monthly Oil Production and LPO Chart.

Table 3.
Risk Index, Workover Schedule, and Gas Perforation Action Plan.

Job Type	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Risk Index	0,01	0,04	0,82	0,29	0,94	0,01	0,75	0,00	0,58	0,21	0,45	1,00
Workover Job (Perf)	12	11	1	7	0	11	2	12	4	8	5	0
Remark	Boost WO	Boost WO	Hold / reduce WO	Normal / selective WO	Hold / reduce WO	Boost WO	Hold / reduce WO	Boost WO	Normal / selective WO	Boost WO	Normal / selective WO	Hold / reduce WO

Table 4.
Risk Index, Workover Schedule, and Oil Perforation and PCTGL Action Plan.

Job Type	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Risk Index	0.02	0.00	0.01	0.57	1.00	0.01	0.02	0.01	0.02	0.00	0.02	0.27
Workover Job (Perf)	3	4	3	1	0	4	3	4	3	4	3	2
WO PCTGL	1	2	1	1	0	2	1	2	1	2	1	1
Remark	Boost WO	Boost WO	Hold / reduce WO	Normal / selective WO	Hold / reduce WO	Boost WO	Hold / reduce WO	Boost WO	Normal / selective WO	Boost WO	Normal / selective WO	Hold / reduce WO

4.7 Discussion

The error metrics and feature rankings indicate that XGBoost captured a meaningful relationship between operating events and LPO. The gas forecast is the more stable of the two, and its leading variables - Low Suction Pressure, Vibration, Discharge Gas Pressure, and Engine Oil Low Pressure - point consistently to compressor performance. This is consistent with the field configuration: when compression is restricted, deliverable gas cannot be fully realized even if reservoir and well capacity remain available.

Oil LPO is less regular. Its larger RMSE and sMAPE reflect abrupt peaks and several near-zero observations rather than a uniform bias across the test period. Compressor and issue-level variables still appear in the oil ranking, suggesting that shared facility conditions influence oil realization, but they do not explain the full variation. Additional well and process-level variables may be needed to resolve the remaining oil error.

A second result is the importance of separating PDT from UPDT. In April and May, for example, the oil profile is driven mainly by planned downtime. A scheduler using only the UPDT forecast could interpret those months as available and add intervention-related deferment to an already heavy plan. Combining the components prevents that specific planning error while still showing whether the exposure is planned or unexpected.

The Risk Index provides the link between that forecast and the workover decision. It does not merely rank months by predicted loss; it also introduces uplift, urgency, job duration, resources, and integrity information. Consequently, a high-loss month is treated cautiously rather than automatically closed, and a low-loss month still requires technically ready candidates.

The production simulation supports this use of the forecast. Relative to the conventional plan, the DSS case gains 1.13 MMscfd of gas and 155 BOPD of oil without exceeding the assumed annual or monthly job capacity. The result should be interpreted as a scenario comparison, not as a guaranteed field gain, because actual execution will depend on candidate success, facility availability, and operating conditions in 2026.

4.8 Main research findings

Optimization in this case means choosing a better execution window, not keeping the rigless unit continuously occupied. Costs are incurred for completed jobs, so shifting or withholding a non-urgent job does not create the standby exposure associated with a continuously contracted rig. The economic value of the DSS therefore comes from two sources: production gained when work is placed in a favorable month and production loss avoided when unnecessary activity is removed from a high-risk month.

Three findings are most relevant to field planning. First, both models met the adopted sMAPE threshold, although the gas forecast was more stable. Second, recurring compressor events, downtime counts, low suction pressure, and vibration were among the most informative LPO features, highlighting facility reliability as a central production constraint. Third, the 2026 Risk Index translated the forecast into distinct workover windows: Boost WO in low-risk months, Normal/Selective WO in intermediate months, and Hold/Reduce WO in months with high combined LPO. The simulated production response indicates that this timing-based allocation can outperform the conventional schedule within the same capacity limits.

5. Conclusion

The analysis leads to the following conclusions:

XGBoost provided a quantitative monthly lost production opportunity (LPO) forecast from the available historical field data. The gas and oil models achieved sMAPE values of 16.79% and 19.06%, respectively, and the gas model reproduced the test-period pattern more consistently. Model interpretation also identified plausible operational drivers, particularly compressor-related events, low suction pressure, vibration, and recurring equipment units.

Forecasting planned downtime (PDT) and unplanned downtime (UPDT) separately gave the Risk Index information that an unplanned downtime (UPDT) only estimate would miss. The forecast supported practical

decisions to increase, selectively execute, or postpone workover activities based on operational risk and candidate readiness.

Embedding the forecast and Risk Index in a DSS improved the simulated 2026 production result from 83.43 to 84.56 MMscfd for gas and from 9,650 to 9,805 BOPD for oil. The improvement came from changing job timing within the assumed capacity, rather than from increasing activity without limit. For the rigless arrangement used in Field ‘YUD’, holding a non-urgent job during a high-risk month is also economically credible because costs are based on executed work and no direct idle-unit charge is incurred.

The principal limitation is the short monthly history: only 36 aggregated observations were available after pre-processing. Future work should extend the observation period and add real-time facility measurements, maintenance history, well-level conditions, and clearer event taxonomy. Those additions would allow temporal validation across more than one operating cycle and may improve the model's response to rare oil-LPO peaks.

The scheduling component could also be developed into a formal economic optimization that includes job-specific cost, hydrocarbon price, probability of success, and the revenue exposed during execution. Before field use, every recommendation should continue to pass through engineering review, candidate-readiness checks, and the current operating constraints of Field ‘YUD’.

6. Acknowledgement

The authors would like to thank PT Pertamina Hulu Indonesia, Regional 3 Zona 9, particularly the PHSS Subsurface Development and Planning and operations teams, for their support, technical discussions, and valuable field perspectives throughout this study. We are also grateful to Universitas Pembangunan Nasional “Veteran” Yogyakarta for the academic guidance and research environment provided during the development of this work. Feedback from colleagues who reviewed the preliminary analysis helped improve the interpretation and ensure that the proposed Decision Support System remained relevant to actual production-planning practices. Certain field information has been anonymized to maintain confidentiality. The interpretations and conclusions presented in this paper are solely those of the authors.

7. Declarations

7.1 Author Contributions

Yudha. A: Writing - Original Draft, Writing - Review & Editing, Conceptualization, Formal analysis, Investigation, Visualization, Development of machine learning. Suranto: Supervision, Formal analysis, Validation. Kodong. F. R.: Supervision, Formal analysis, Development of machine learning. Soeparmono. M. R.: Conceptualization, Formal analysis, Data Curation.

7.2 Funding statement

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

7.3 Competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

7.4 The use of AI or AI-assisted technologies

During the preparation of this work, the authors used ChatGPT to assist in developing Python scripts and supporting the data analysis. All outputs generated by the tool were carefully reviewed, verified, and edited by the authors, who take full responsibility for the accuracy and content of this publication.

8. Glossary of terms and symbols

Terms & Symbol	Definition	Unit
LPO	Lost Production Opportunity; production volume that cannot be realized because of well, operational, or facility constraints.	MMSCF (gas); bbl (oil)
PDT	Planned Downtime; production loss associated with scheduled maintenance, shutdowns, or well activities.	MMSCF (gas); bbl (oil)
UPDT	Unplanned Downtime; production loss caused by unexpected equipment failures, well problems, or facility restrictions.	MMSCF (gas); bbl (oil)
DSS	Decision Support System used to support workover scheduling and production-planning decisions.	–
EDA	Exploratory Data Analysis.	–
XGBoost	Extreme Gradient Boosting; the machine-learning algorithm used to predict LPO.	–
WO	Workover; a well intervention performed to restore or improve production performance.	job
Perf	Perforation activity.	job
PCTGL	Permanent Coiled Tubing Gas Lift.	job
MAE	Mean Absolute Error; the average absolute difference between actual and predicted values.	Same as target variable
RMSE	Root Mean Square Error; an error measure that gives greater weight to large prediction errors.	Same as target variable
sMAPE	Symmetric Mean Absolute Percentage Error.	%
R ²	Coefficient of determination; the proportion of variation explained by the model.	–
Risk Index	Composite indicator of monthly production-loss exposure and workover-execution risk.	–
Event Load	Relative intensity of disturbance-related events used in the forecast scenarios.	–
Plan_WO_Perf	Planned monthly allocation of perforation or workover jobs.	job/month
Plan_WO_PCTGL	Planned monthly allocation of PCTGL jobs.	job/month
BOPD	Barrels of oil per day.	bbl/d
MMSCF	Million standard cubic feet.	MMSCF
MMscfd	Million standard cubic feet per day.	MMscf/d
bbl	Barrel of oil.	bbl
L	Overall objective or loss function of the XGBoost model.	–
$l(y_i, \hat{y}_i)$	Training loss between the actual and predicted values for observation i .	–
y_i	Actual value of the target variable for observation i .	MMSCF or bbl
$\hat{y}_i$	Predicted value of the target variable for observation i .	MMSCF or bbl
i	Observation index.	–
k	Tree index in the XGBoost model.	–
f_k	Prediction function produced by tree k .	–
$\Omega(f_k)$	Regularization term for tree k .	–
T	Number of leaves in a decision tree.	count
w	Vector of decision-tree leaf weights.	–
γ	Regularization parameter controlling the penalty for adding tree leaves.	–
λ	L2 regularization coefficient applied to leaf weights.	–

9. References

- Allan, D., Bailey, W. J., Sauv  , B., Torrens, R., & Dhabi, A. (2016). The Defining Series: Integrated Asset Modeling. Source: www.slb.com/defining
- Asihono, P. M. S., & Mulyono, N. B. (2023). Maintaining the Reliability of Production Facilities in Operation Management of Upstream Oil and Gas Industries: Case Study at GNK Gas Compression Station PT. PEP ZONA 4. *European Journal of Business and Management Research*, 8(2), 154–162. DOI: <https://doi.org/10.24018/ejbmr.2023.8.2.1858>
- Cio Mario, C., Saputra, R., Saefuddin, S., Triyono, T., Energi Indonesia, S., Metra, S., Gaskari, R., Martinez, H., Songchitruksa, P., Agustinus Kartika Djaya, B., Ranjan Sinha, R., & Vadivel, K. (2019). SPE-196326-MS Well Opportunity Register, Define and Selection: Changing the Game in Production Optimisation Using Automation and Analytics. DOI: <https://doi.org/10.2118/196326-MS>
- Dethlefs, J. C., & Chastain, B. (2012). Assessing Well-Integrity Risk: A Qualitative Model. DOI: <https://doi.org/10.2118/142854-PA>
- Fauzan Amir, M., Nugroho PHW, J., Krisna Permadi, W., Ihsan, A., Nitasari, P. D., & Migas, S. (2022). “Kebijakan, Strategi dan Teknologi Tepat Guna untuk Meningkatkan Pengurusan Lapangan Minyak dan Gas di Indonesia” Data Analytics Applications in Oil and Gas Industry in Indonesia: Some Implementation in SKK Migas and Pertamina Hulu Rokan (PHR). Source link: <https://journal.iatmi.or.id/index.php/ojs/article/download/248/254>
- Hadi, S., Osman, A., Iskandar Bahr  n, A., Thura Khalid, M., Ismaon, M., Rahim, P. A., & Carigali Sdn Bhd, P. (2008). SPE 117954 Prudent and Integrated Production Planning to Maintain Excellent Operations in Angsi Field, Peninsular Malaysia Operations. DOI: <https://doi.org/10.2118/117954-MS>
- Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22(4), 679–688. DOI: <https://doi.org/10.1016/j.ijforecast.2006.03.001>
- Hyoung, J., Lee, Y., & Han, S. (2024). Development of Machine Learning-Based Production Forecasting for Offshore Gas Fields Using a Dynamic Material Balance Equation. *Energies*, 17(21). DOI: <https://doi.org/10.3390/en17215268>
- Mat, M., Zainudin, M., Riyanto, L., Manaf, N H, & Petroleum, B. S. (2025). Unlocking Insights: Data Analytics to Predict Unplanned Production Deferment in Offshore Oil and Gas Wells. Source link: https://cdn.asp.events/CLIENT_SPE_29FE44FF_5056_B733_49EC4D60A02A6A7D/sites/iptc-2025/media/abstract_only/IPTC-24678.pdf
- Maulana, F., Yasutra, A., & Syihab, Z. (2025). Case Study of Heterogeneity Index’s Effect on The Successful Workover Based on The Apriori Algorithm. *Scientific Contributions Oil and Gas*, 48(1), 135–144. DOI: <https://doi.org/10.29017/scog.v48i1.1658>
- Monta  o Moreno, J. J., Palmer Pol, A., Ses   Abad, A., & Cajal Blasco, B. (2013). Using the R-MAPE index as a resistant measure of forecast accuracy. *Psicothema*, 25(4), 500–506. DOI: <https://doi.org/10.7334/psicothema2013.23>
- Natural Gas Market Module of the National Energy Modeling System: Model Documentation 2025. (2025). Source: www.eia.gov
- Nwaochei, F., Adelowotan, A., Liu, T., & Goldman, J. (2019). SPE-198821-MS Prioritizing Non-Rig Well Work Candidates Using Data Science. DOI: <https://doi.org/10.2118/198821-MS>
- Okon, I. (2021). SPE-208246-MS Identification of Potential Candidate’s Wells for Workover Background of the study. DOI: <https://doi.org/10.2118/208246-MS>
- Popa, A., Bloomquist, C., & Marghitoiu, S. (2023). The Investigation of Workover Rig Scheduling Optimization Using Genetic Algorithms - Part I. SPE Western Regional Meeting Proceedings, 2023-May. DOI: <https://doi.org/10.2118/212948-MS>

- Siahaan, E., Mamat, I., Laksana, S. S., Setyowibowo, A., Naufal, A. A., Wulandari, O. E., Metra, S., Nainggolan, A. K., Arinandy, O. I., Ellen, L., Pramita, M. D., Tjong, A., Wilian, Y., & Songchitruksa, P. (2022). Well Portfolio Optimisation: Accelerating Generation of Well Intervention Candidates with Automated Analytics and Machine Learning - A Case Study from Attaka Field, Indonesia. Society of Petroleum Engineers - SPE Asia Pacific Oil and Gas Conference and Exhibition 2022, APOG 2022. DOI: <https://doi.org/10.2118/210614-MS>
- Smithson, T. (2016). The Defining Series: Reservoir Drive Mechanisms. Source: www.slb.com/defining
- Teh, Y. H., Theseira, K., Abdul Karim, A. H., Hashim, N. S., Yakob, A. R., Musa, A. S., Borhan, N. A., Muhamad, S., & Sykahar, M. W. (2008). Preparing a Gas Field Development Plan : Tangga Barat Cluster Gas Project. International Petroleum Technology Conference. DOI: <https://doi.org/10.2523/IPTC-12488-MS>
- Von Flatern, R. (2014). Defining Intervention: Upstream Maintenance and Repair. Oilfield Review, 26(4), 62–63. Source: www.slb.com/defining
- Wijaya, A. A., Aulianagara, R., Guo, W., Naibaho, F. M., Amirudin, U., & Asriwan, F. X. (2020). Multi-detector pulsed neutron tool application in low porosity reservoir - a case study in mutiara field, Indonesia. SPWLA 61st Annual Logging Symposium. DOI: <https://doi.org/10.30632/SPWLA-5082>
- Wisnu Ady Kusmono, B., & Ros Bangun, Y. (2024). Reducing Loss Production Opportunity (LPO) at PT Pertamina EP Donggi Matindok field. Interdisciplinary Social Studies. DOI: <https://doi.org/10.55324/iss.v3i4.730>
- Woodward, W. A., Gray, H. L., & Elliott, A. C. (2017). APPLIED TIME SERIES ANALYSIS WITH R Second Edition (SECOND EDITION). ISBN: 9781498734226