

Application of decision tree algorithm for accurate prediction of liquid loading in gas-condensate wells

Zastosowanie algorytmu drzewa decyzyjnego do dokładnego przewidywania obciążenia otworu cieczą w odwiertach gazowo-kondensatowych

Kanan Aliyev

Azerbaijan State Oil and Industry University

ABSTRACT: Liquid loading occurs when gas fails to lift co-produced condensates to the surface, causing backpressure, reduced production, and possibly resulting in well shutdown. This occurs when gas velocity falls below the critical level required to carry liquids, leading to their accumulation in the wellbore. Accumulation can occur in both vertical and horizontal wells, reducing efficiency, particularly in wet or retrograde gas wells. Accurate prediction and monitoring are crucial but often challenging due to the complexities of multiphase flow and estimating bottom-hole pressure. This study assesses the effectiveness of the decision tree algorithm for predicting the loading status of these wells, aiming to improve predictive accuracy and operational decision-making. Two decision tree models were developed using wellhead pressure and gas production rate as input features. The first model, with a maximum tree depth of 3, was designed to prevent overfitting by limiting the complexity of the decision tree. This constraint helped maintain model simplicity while still achieving an accuracy of 80%. The depth limitation ensured that the tree did not grow excessively, which can sometimes lead to overfitting, and instead focuses on capturing the most significant data patterns with a limited number of decision nodes. The second model, with constraints on node splits and leaf samples but no depth limitation, reached an accuracy of 78%. The results revealed that gas production rate is a more influential factor than wellhead pressure in determining well loading status, with the second model indicating that wellhead pressure becomes less relevant when the gas rate exceeds 75,365.2 m³/day. Both models performed well overall but showed potential for improvement. Future work should focus on enhancing model accuracy through advanced techniques such as ensemble methods and by increasing the dataset size through the inclusion of additional well data. Despite the limitations of a relatively small sample size, the findings underscore the potential of decision tree models in optimizing well productivity and operational efficiency in gas-condensate reservoir management.

Key words: gas-condensate reservoirs, liquid loading, machine learning, decision tree algorithm, supervised classification.

STRESZCZENIE: Zjawisko obciążenia otworu cieczą występuje, gdy gaz nie jest w stanie wynieść współprodukowanego kondensatu na powierzchnię, co powoduje wzrost ciśnienia zwrotnego, spadek wydajności oraz potencjalnie prowadzi do likwidacji odwiertu. Zjawisko to zachodzi wówczas, gdy prędkość gazu spada poniżej poziomu krytycznego niezbędnego do transportu cieczy, co prowadzi do gromadzenia się jej w odwiercie. Do akumulacji może dochodzić zarówno w odwiertach pionowych, jak i horyzontalnych, co zmniejsza ich wydajność, szczególnie w przypadku złóż gazu mokrego, lub w których występuje kondensacja wsteczna. Tym samym niezbędne jest dokładne prognozowanie i monitorowanie tego zjawiska, co jednak często jest utrudnione ze względu na złożoność przepływu wielofazowego oraz trudności w oszacowaniu ciśnienia przy dnie odwiertu. W niniejszym artykule przeprowadzono ocenę skuteczności zastosowania algorytmu drzewa decyzyjnego do przewidywania statusu zalegania cieczy w odwiertach, w celu zwiększenia dokładności prognoz oraz usprawnienia procesu podejmowania decyzji operacyjnych. Opracowano dwa modele drzewa decyzyjnego, w których jako dane wejściowe zastosowano ciśnienie na głowicy odwiertu oraz wydajność wydobywania gazu. Pierwszy model, z maksymalną głębokością drzewa ograniczoną do 3, został zaprojektowany tak, aby uniknąć przeuczenia poprzez ograniczenie złożoności struktury decyzyjnej. To ograniczenie pozwoliło zachować prostotę modelu przy jednoczesnym osiągnięciu dokładności na poziomie 80%. Ograniczenie głębokości zapobiegło nadmiernemu rozrostowi drzewa, który mógłby prowadzić do przeuczenia modelu, jednocześnie umożliwiając uchwycenie najistotniejszych wzorców w analizowanych danych przy użyciu niewielkiej liczby węzłów decyzyjnych. Drugi model, z ograniczeniami dotyczącymi liczby podziałów węzłów i liczby próbek w liściach, lecz bez limitu głębokości drzewa, osiągnął dokładność 78%. Wyniki wykazały, że wydajność wydobywania jest czynnikiem bardziej znaczącym niż ciśnienie na głowicy odwiertu w określaniu statusu zalegania cieczy, a w przypadku drugiego modelu zauważono, że ciśnienie na głowicy odwiertu traci na znaczeniu, gdy wydajność gazu przekracza 75 365.2 m³/dobę. Oba modele wykazały dobrą ogólną skuteczność, ale również potencjał do dalszego udoskonalania. Przyszłe prace powinny skupić

się na poprawie dokładności predykcyjnej poprzez zastosowanie zaawansowanych metod, takich jak techniki zespołowe, jak również zwiększenie liczby danych poprzez uwzględnienie informacji z dodatkowych odwiertów. Pomimo ograniczeń związanych ze stosunkowo niewielką liczbą próbek, wyniki podkreślają potencjał modeli drzew decyzyjnych w optymalizacji wydajności odwiertów oraz efektywności operacyjnej w zarządzaniu złożami gazowo-kondensatowymi.

Słowa kluczowe: złoża gazowo-kondensatowe, obciążenie otworu cieczą, uczenie maszynowe, algorytm drzewa decyzyjnego, klasyfikacja nadzorowana.

Introduction

Liquid loading poses a significant challenge in gas-condensate wells, particularly as declining reservoir pressures affect production efficiency. In the early stages, high gas flow rates can effectively transport liquid through the tubing within the gas core. However, as reservoir pressure declines, liquid accumulation begins in the wellbore, leading to an increase in “liquid hold-up” within the tubing. This buildup raises the tubing gradient, which adds backpressure on the formation and lowers surface tubing pressure (Hamidov and Fataliyev, 2016; Lea and Rowlan, 2019). In wells without a packer, a sudden drop in wellhead pressure (P_{wh}), combined with an increase in tubing-casing annulus pressure (P_{csg}), is a strong indicator of liquid loading (Hearn, 2010; Lea and Rowlan, 2019; Tugan, 2020).

Decline curve analysis (DCA) is a method for detecting liquid loading in gas-condensate wells by analyzing the relationship between oil or gas flow rates (q) and time (t) in producing wells (Belyadi et al., 2019; Franchi, 2010). Under normal conditions, where gas is produced without interference, the DCA exhibits smoothly declining, exponential curves due to reservoir depletion. However, liquid loading is often signaled by abrupt deviations from this pattern, leading to steeper declines in the curve (Lea and Rowlan, 2019). Additionally, factors such as the reservoir's stress sensitivity and a significant drop in bottom-hole formation pressure can further accelerate the reduction in gas production (Zhu, 2009). Other issues, including corrosion, inadequate cementation, and technological malfunctions, can also cause leaks in the production string, thereby reducing gas flow rates. Therefore, engineers must carefully identify the underlying cause of rapid gas production decline in gas-condensate wells, considering all these potential factors.

Another approach is nodal analysis, a graphical technique commonly employed to illustrate how gas flow rates from the reservoir change in response to pressure variations within the wellbore. This method is particularly beneficial for gas-condensate wells, where it helps predict stable flow conditions. Stability is indicated when the tubing curve intersects the inflow performance relationship (IPR) curve to the right of its minimum point. Conversely, if the operating point falls to the left of the minimum pressure point, liquid loading is anticipated (Waltrich et al., 2015; Lea and Rowlan, 2019). Pagan et al.

(2016) enhanced this technique to more precisely predict the onset of liquid loading, showing strong alignment with both experimental and field data. However, nodal analysis faces challenges in low permeability reservoirs, where the IPR curves intersect the tubing performance relationship (TPR) curve to the left of the minimum pressure point. In such cases, wells might continue to produce without exhibiting liquid loading symptoms (Lea and Rowlan, 2019). This suggests that minimum pressure alone may not always be a reliable predictor of liquid loading, as factors like permeability can also play a significant role.

Liquid loading in gas wells occurs when the gas velocity is insufficient to transport liquid to the surface, requiring production at or above a certain minimum flow rate, known as the critical gas rate. This rate can be estimated using various models, including the well-established Turner correlation (Turner et al. 1969). In this research, the Authors examined both continuous film and entrained droplet movement models, ultimately concluding that the droplet model was the most suitable for their data. This model calculates critical loading velocity with a 20% adjustment based on the forces acting on the droplets, but it is limited to wells with high wellhead pressure and assumes that the droplets are spherical. Coleman et al. (1991) demonstrated that Turner's model accurately predicts the critical flow rate for low-pressure gas wells, particularly those with wellhead pressures below 3.448 MPa (500 psi), without necessitating the 20% adjustment. Building upon this foundation Nosseir et al. (2000), Li et al. (2002), Guo et al. (2006), Zhou and Yuan (2010), Luan and He (2012), Wang et al. (2015), and El Fadili and Shah (2017), have refined Turner's model by incorporating additional factors such as flow patterns, drag coefficients, droplet shapes, kinetic energy, Weber number, and inclination. These enhancements aim to improve the accuracy of predicting critical gas velocity. Despite these advancements, many models are based on assumptions that may not fully capture real conditions, thereby limiting their practical application. Moreover, the complexity of multiphase flow in gas-condensate wells, exacerbated by variations in pressure and temperature throughout the well, presents significant challenges in developing a universally applicable model for predicting liquid loading.

In recent years, advancements in technology, particularly in machine learning (ML), have significantly improved efficiency, automation, and scalability. Various authors have applied dif-

ferent ML algorithms to problems in petroleum engineering, yielding promising results. Studies indicate that the decision tree algorithm, a popular classification method in ML, is particularly effective for predicting liquid loading status in gas-condensate wells. According to the results of Almashan et al. (2020), the boosted decision tree regression (BDTR) model demonstrated its effectiveness in predicting liquid loading (liquid holdup, HL) in multiphase flows within oil and gas wells. Trained on experimental datasets, the BDTR model outperformed traditional empirical correlations, offering superior accuracy without the limitations of specific flow conditions. This highlights the decision tree model as a powerful tool for accurately predicting liquid loading. In the study conducted by Joseph and Bassey (2022), the decision tree model was successfully utilized to classify the liquid loading status of gas wells. This model demonstrated a good degree of accuracy, predicting about 81% of the status of the wells under investigation. Validated with actual data, the decision tree effectively identified liquid loading phenomena, highlighting its usefulness compared to empirical models. In the study by Chemmakh et al. (2023), the extreme gradient boosting (XGBoost) model proved to be a highly effective tool for predicting the status of wells as either “loaded” or “unloaded” based on various completion and fluid properties, including diameter, liquid density, gas density, liquid viscosity, gas viscosity, angle of inclination, superficial liquid velocity, and interfacial tension. By learning from previous data, XGBoost outperformed other models and correlations, achieving an impressive F-1 score of 0.947 and correctly classifying 46 out of 50 test cases. This demonstrates its superior performance in accurately determining the liquid loading status of wells.

Almashan et al. (2020), Joseph and Bassey (2022), and Chemmakh et al. (2023) suggested that classical machine learning techniques provide significant advantages and promising opportunities for investigating liquid loading in gas wells, offering a more robust alternative to empirical models that depend on numerous simplifications. This research introduces a simplified decision tree approach that utilizes only wellhead pressure and gas production rate to effectively predict the liquid loading status in gas-condensate wells. This minimal-input model distinguishes itself from prior work by emphasizing interpretability and practicality without compromising predictive performance.

The novelty of this work lies in its focus on operational efficiency, relying solely on basic surface data to predict liquid loading. Furthermore, the model is trained on a comprehensive dataset compiled from multiple sources, enhancing its robustness and generalizability across various field conditions. This approach offers a cost-effective and easily deployable alternative to more data-intensive machine learning models.

Theoretical background

The decision tree algorithm is a widely used tool in machine learning, known for its straightforward approach to data classification. It is easy to understand, interpret, and implement, making it a more favorable choice for studies (Aggarwal et al., 2020; James et al., 2023). Decision trees are applied in various areas, including business, science, and healthcare, where they help simplify complex decision-making tasks (Sharma and Kumar, 2017; Aggarwal et al., 2020).

One of the strengths of decision trees is their versatility; they work well with both categorical and continuous data and can handle missing values effectively (James et al., 2023). The structure of a decision tree is hierarchical, with each node representing a decision based on input features, and each branch representing a decision rule. The tree ends in leaf nodes, which represent the final class label or outcome (Yang, 2019). Figure 1 illustrates the structure of a simple decision tree.

The algorithm works by splitting the dataset into subsets, starting at the root node. At each step, the feature that best separates the classes is chosen, often using metrics like Gini impurity or information gain. Equation 1 shows calculation of the Gini impurity:

$$\text{Gini Impurity}(t) = 1 - \sum_{i=1}^n p_i^2 \quad (1)$$

where:

p_i – proportion of samples belonging to class i at node t_i ,

n – total number of classes.

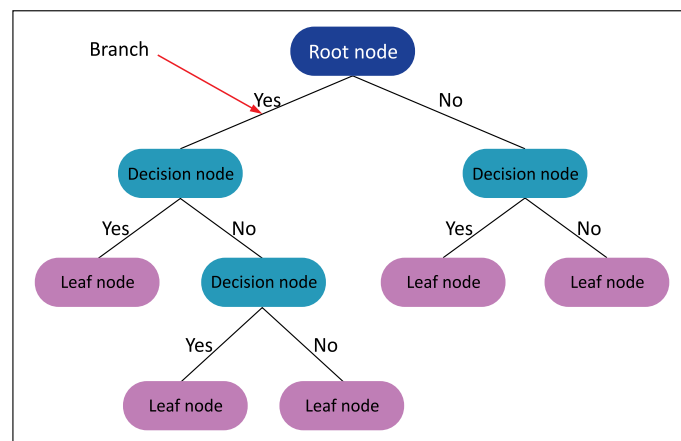


Figure 1. Structure of a simple decision tree

Rysunek 1. Struktura prostego drzewa decyzyjnego

The Gini impurity ranges from 0 to 0.5, where a lower value indicates a better split, with 0 denoting a pure node and 0.5 representing the most impure node. The decision tree aims to minimize this impurity by splitting the dataset at each node, choosing the split that results in the lowest weighted average Gini impurity across the child nodes. This process continues until a stopping criterion is reached, such as a maximum tree

depth or a minimum number of samples per leaf (Hassan et al., 2023). The resulting tree-like structure provides a clear visual representation of the decision-making process, from the root to the leaf nodes.

However, decision trees can sometimes overfit, especially with complex datasets. This means they might not perform well on new, unseen data. To address this, techniques such as pruning are used to simplify the tree by halting its growth early and removing branches that do not add significant predictive value, based on criteria such as Gini impurity or information gain. This pre-pruning technique typically sets thresholds, such as a maximum tree depth or a minimum number of samples per leaf, to determine when to stop splitting. Additionally, ensemble methods like random forests have been developed to improve the accuracy and reliability of decision trees by combining the results of multiple trees (James et al., 2023).

Materials and Methods

To build a robust dataset for training and testing the classification model, data from various sources were combined, yielding 228 records. The dataset includes not only Turner's experimental data but also field data used in various studies, covering reservoirs from the Gulf of Mexico, Yakela-Dalaoba, Xinjiang, and other Chinese condensate gas fields (Turner et al. 1969, Luo et al., 2014; Zhou et al., 2016; Ikpeka and Okolo, 2019; Ming and He, 2020; Xiao et al., 2021). Despite the heterogeneity of the source datasets, certain key parameters such as wellhead pressure, gas rate, and loading status were consistent across them. Leveraging this commonality, the

datasets were synthesized to create a cohesive dataset, thereby augmenting the sample size while preserving diversity in well characteristics. The new dataset comprises data from 109 loaded wells and 119 not-loaded wells. The liquid loading status of wells is represented by categorical values (such as “not loaded” and “loaded”). To facilitate computational analysis, these categorical values were converted into numerical integers, with “not loaded” wells assigned a value of 0 and “loaded” wells assigned a value of 1. This transformation allows decision tree models to process the data more effectively. The analysis was performed using the Scikit-Learn library (version 1.5, Pedregosa et al. 2011), an open-source ML library, in Python (version 3.12.5). In this study, two models were developed. The first model was constrained by setting the maximum depth of the tree to 3. In the second model, the depth restriction was removed, instead specified a minimum of 35 samples required to split a node and 15 samples per leaf. These hyperparameter adjustments were made to control model complexity and reduce overfitting. Both models were trained using 70% of the dataset, while the remaining 30% was reserved for testing. The primary difference between the models lies in their structural constraints.

Results and discussion

The distribution of the two predictors – wellhead pressure and gas production rate – is depicted in Figure 2a and 2b, respectively. The distribution patterns of both wellhead pressure and gas rate are characterized by notable left-skewness, which is unsurprising in the context of gas production operations due to factors such as reservoir depletion.

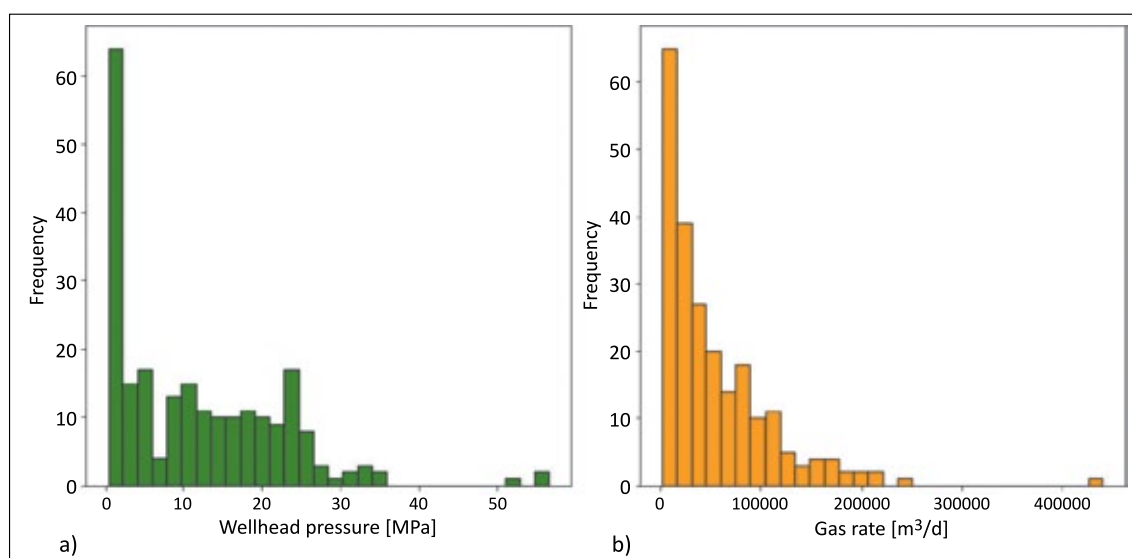


Figure 2. Histogram of input parameters: a) wellhead pressure, b) gas production rate

Rysunek 2. Histogram pokazujący rozkład parametrów wejściowych: a) ciśnienia na głowicy, b) wydajności wydobywania gazu

Table 1. Statistical description of the input parameters**Tabela 1.** Statystyczny opis parametrów wejściowych

	Wellhead pressure [MPa]	Gas production rate [m ³ /d]
Count	228	228
Mean	11.67	55,013.66
Std	10.61	55,701.66
Min	0.34	2,054.02
25%	1.40	15,319.87
50%	10.30	36,542.83
75%	19.24	79,040.50
Max	56.64	440,580.01

In most scenarios, wellhead pressure remains below 27.58 MPa, with approximately 50% of records falling below 10.3 MPa. Similarly, half of the gas rate values are below 36,542.83 m³/day. All statistical information regarding these two parameters is presented in Table 1.

Following the exploratory data analysis, the dataset was partitioned into training and test sets, with 70% of the data allocated for model training and the remaining 30% reserved for evaluating its performance. Initially, a decision tree model was constructed without constraints using the decision tree classi-

fier from the Scikit-Learn library. This resulted in a complex and lengthy tree due to the high variability in the data and the complex relationships between the predictor and target features. This complexity led to overfitting, where the model learned too much from the training data, including its noise and exceptions. To address overfitting, adjustments were made, and two different models were developed.

The first model involved setting a maximum depth of 3 for the decision tree. Figure 3 illustrates the structure of this first decision tree. The tree's initial split is based on the gas rate: if the gas rate is below a certain threshold, the model moves to the left branch to further evaluate whether the well is loaded. The subsequent splits on the left side are determined by gas rate and wellhead pressure values. According to this model, a well is predicted to be loaded with 90% probability if the gas production rate is below 19,761.396 m³/day. Conversely, the well is classified as not loaded in two scenarios: (1) when the gas rate ranges between 19,761.396 m³/day and 75,365.289 m³/day and the wellhead pressure is below 19.651 MPa, and (2) when the gas rate exceeds 75,365.289 m³/day and the wellhead pressure is above 5.878 MPa.

The accuracy of this model is 80%. Additionally, Tables 2 and 3 present the confusion matrix and classification report for this model, which provide a detailed view of its classification

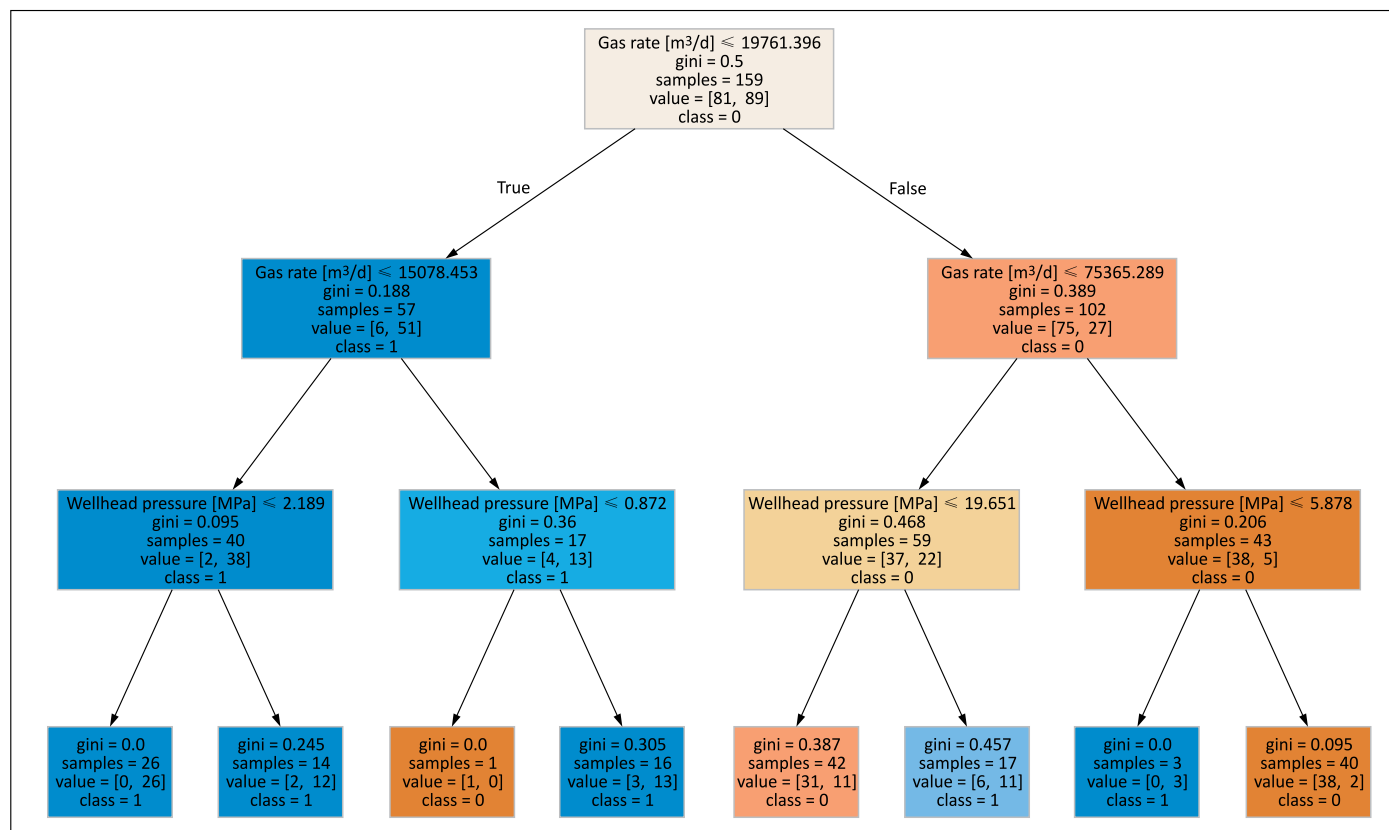
**Figure 3.** Structure of the first decision tree (maximum depth: 3) built using the training set. Class = 1 : “loaded”, class = 0 : “not loaded”**Rysunek 3.** Struktura pierwszego drzewa decyzyjnego (maksymalna głębokość: 3) zbudowanego na danych treningowych. Klasa = 1 „obciążony”, Klasa = 0 : „nieobciążony”

Table 2. Confusion matrix of the first decision tree (maximum depth: 3) evaluated on the test set

Tabela 2. Macierz pomyłek pierwszego modelu (maksymalna głębokość drzewa: 3) wyznaczona na danych testowych

	Actual positive (1, “loaded”)	Actual negative (0, “not loaded”)
Predicted positive (1, “loaded”)	31	7
Predicted negative (0, “not loaded”)	7	24

Table 3. Classification report of the first decision tree (maximum depth: 3)

Tabela 3. Raport klasyfikacji pierwszego modelu (maksymalna głębokość drzewa: 3)

	Precision	Recall	F1 Score	Support
0 (“not loaded”)	0.82	0.82	0.82	38
1 (“loaded”)	0.77	0.77	0.77	31
Accuracy	–	–	0.80	69
Macro avg	0.79	0.79	0.79	69
Weighted avg	0.80	0.80	0.80	69
– not applicable				

performance. The confusion matrix compares the predicted and actual labels of the classification model. The model correctly predicted 31 positive instances (true positives, or loaded cases) and 24 negative instances (true negatives, or unloaded cases). However, it made 7 false positive predictions, where unloaded cases were incorrectly classified as loaded, and 7 false negative predictions, where loaded cases were incorrectly classified as unloaded. The matrix indicates that the model performs reasonably well for both classes.

F1 score, recall, and precision are important evaluation metrics used in classification tasks to assess model performance beyond overall accuracy. In classification, precision measures the proportion of true positive predictions among all predicted positives, indicating the accuracy of positive predictions. High precision means fewer false positives, whereas low precision suggests frequent misclassification of negatives as positives. Recall measures the proportion of true positives among all actual positives, reflecting the model’s ability to identify relevant positive cases. High recall indicates fewer false negatives, while low recall suggests missed positives. The F1 score is the harmonic mean of precision and recall, providing a balanced measure of both metrics. A high F1 score indicates good overall

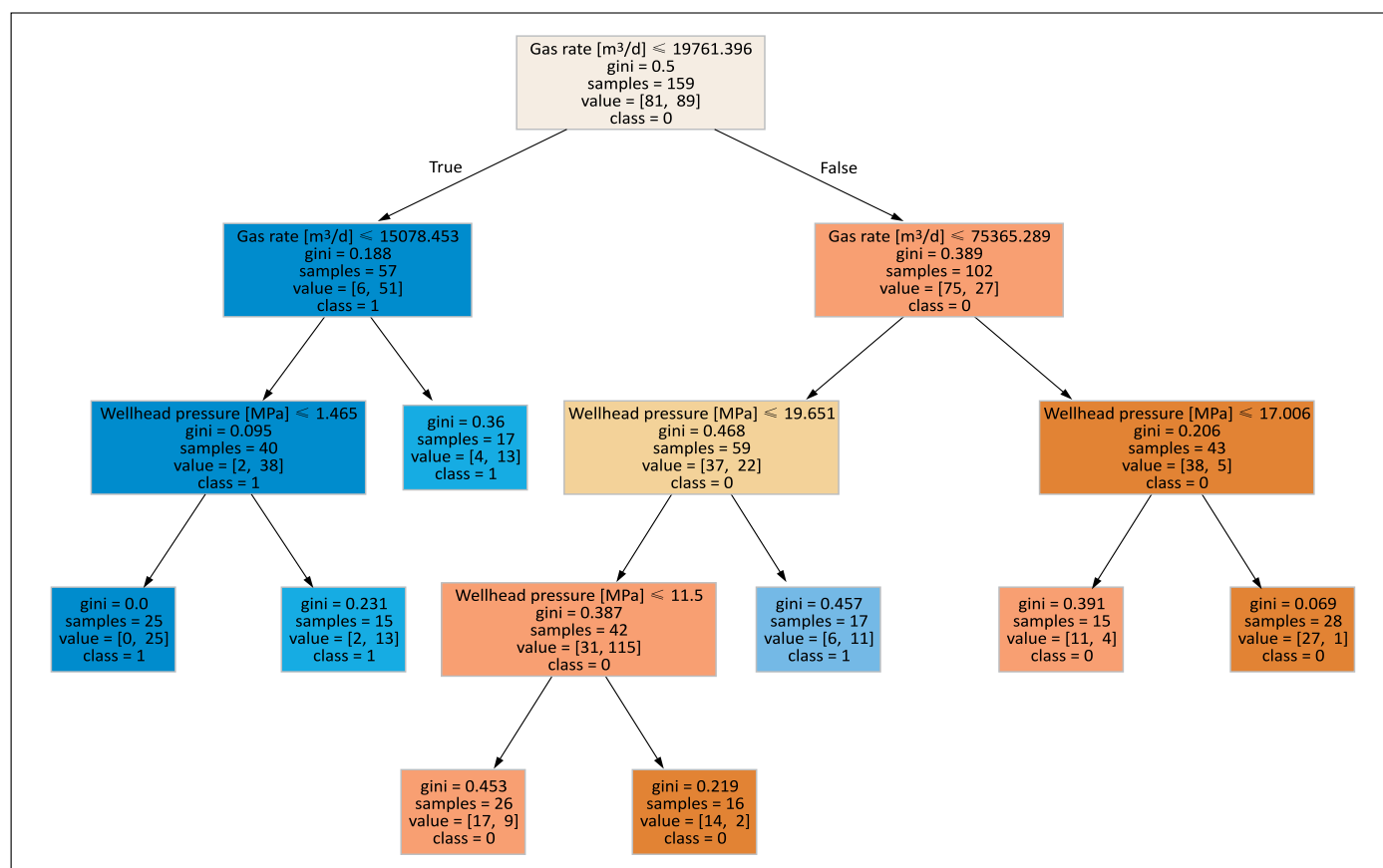


Figure 4. Structure of the second decision tree built using the training set (minimum number of samples to split a node: 35, minimum number of samples per leaf: 15). Class = 1 : “loaded”, class = 0 : “not loaded”

Rysunek 4. Struktura drugiego drzewa decyzyjnego zbudowanego na danych treningowych (minimalna liczba próbek do podziału węzła: 35, minimalna liczba próbek na liściu: 15). Klasa = 1 „obciążony”, Klasa = 0 : „nieobciążony”

performance, while a low F1 score suggests a trade-off between precision and recall. Despite good precision and recall in the first model, there is room for improvement, particularly for class 1 (“loaded”), where the F1 score is somewhat lower at 0.77. The “Support” column in Table 3 represents the number of instances for each class in the dataset. This indicates that, while the model is effective, there are opportunities to enhance its performance, especially for distinguishing class 1 (“loaded”) more accurately.

In developing the second model, several adjustments were implemented to refine its performance. Specifically, the minimum number of samples required to split a node was set to 35, and the minimum number of samples per leaf was set to 15. Unlike the first model, there was no restriction placed on the maximum depth of the decision tree for the second model. Figure 4 illustrates the structure of the second decision tree.

Although the second model produced results similar to the first, its accuracy was slightly lower, at 78%. The criteria for determining the loading condition of the well remained consistent between both models. However, the second model revealed that when the gas rate exceeds 75,365.289 m³/day, the wellhead pressure no longer affects the well’s loading status. In this case, the well is classified as not loaded regardless of the wellhead pressure.

Table 4. Confusion matrix of the second decision tree for test set (minimum number of samples to split a node: 35, minimum number of samples per leaf: 15)

Tabela 4. Macierz pomyłek drugiego drzewa decyzyjnego dla danych testowych (minimalna liczba próbek do podziału węzła: 35, minimalna liczba próbek na liściu: 15)

	Actual positive (1, “loaded”)	Actual negative (0, “not loaded”)
Predicted positive (1, “loaded”)	30	8
Predicted negative (0, “not loaded”)	7	24

Table 5. Classification report of the second decision tree (minimum number of samples to split a node: 35, minimum number of samples per leaf: 15)

Tabela 5. Raport klasyfikacji drugiego modelu (minimalna liczba próbek do podziału węzła: 35, minimalna liczba próbek na liściu: 15)

	Precision	Recall	F1 Score	Support
0 (“not loaded”)	0.81	0.79	0.80	38
1 (“loaded”)	0.75	0.77	0.76	31
Accuracy	—	—	0.78	69
Macro avg	0.78	0.78	0.78	69
Weighted avg	0.78	0.78	0.78	69
— not applicable				

Tables 4 and 5 present the confusion matrix and classification report for the second model, providing a comprehensive view of its classification performance. The model demonstrates robust performance in distinguishing between negative and positive instances, with precision, recall, and F1 scores indicating generally good results. Nonetheless, there is room for improvement, particularly in reducing false positives for class 1 (“loaded”) and false negatives for class 0 (“not loaded”). Addressing these issues could enhance the model’s overall accuracy and reliability.

Conclusion

This study advances the prediction of liquid loading status in gas-condensate wells using decision tree models. The models were trained and evaluated on a dataset consisting of wellhead pressure and gas production rate, with the goal of improving predictive accuracy and understanding the factors influencing loading status. The main findings and contributions of this research are as follows:

1. Two decision tree models were developed. The first model, with a maximum depth of 3, achieved an accuracy of 80%, successfully classifying wells based on both gas rate and wellhead pressure. The second model, incorporating constraints on node splits and leaf samples but with no depth restriction, achieved a slightly lower accuracy of 78%. This model revealed that when the gas rate exceeds 75,365.289 m³/day, wellhead pressure no longer influences loading status.
2. The analysis highlighted that gas rate plays a more significant role in determining well loading status compared to wellhead pressure. This insight underscores the importance of gas rate in the predictive model, as it was a critical factor in both decision trees.
3. Both models demonstrated strong performance in classifying well loading status, with reasonable precision and recall. However, the results also identified areas for improvement, particularly in reducing false positives for class 1 (“loaded”) and false negatives for class 0 (“not loaded”). This suggests that while decision trees are effective, there is potential for further refinement.
4. To enhance predictive accuracy, incorporating advanced techniques such as ensemble methods (e.g., random forests or boosting) could address current limitations. Additionally, expanding the dataset to include more diverse and larger samples could improve model generalization and accuracy, reducing overfitting and capturing more complex patterns in the data.

In summary, decision tree models offer valuable insights into the prediction of liquid loading status in gas-condensate wells.

Despite their effectiveness, there is room for improvement through advanced modeling techniques and larger datasets. Future research should focus on these aspects to further refine the models and enhance their predictive capabilities.

References

- Aggarwal C.C., Lee V.E., Liu L., Jin R., 2020. Data classification: Algorithms and applications. 1st ed. *Springer Cham*, 87–103.
- Almashan P., Narusue Y., Morikawa H., 2020. Decision tree regressions for estimating liquid holdup in two-phase gas-liquid flows. *Abu Dhabi International Petroleum Exhibition & Conference*. DOI: 10.2118/203448-MS.
- Belyadi H., Fathi E., Belyadi F., 2019. Decline curve analysis. [In:] Belyadi H., Fathi E., Belyadi F. (eds.). *Hydraulic Fracturing in Unconventional Reservoirs*. 2nd ed. *Gulf Professional Publishing*, 311–340. DOI: 10.1016/B978-0-12-817665-8.00017-5
- Chemmakha A., Tomomewo O., Ling K., Shammari A., 2023. Evaluation of liquid loading in gas wells using machine learning. *Petroleum & Petrochemical Engineering Journal*, 7(1). DOI: 10.23880/ppej-16000333.
- Coleman S.B., Clay H.B., McCurdy D.G., Norris III L.H., 1991. A new look at predicting gas-well load-up. *Journal of Petroleum Technology*, 43(3): 329–333. DOI: 10.2118/20280-PA.
- El Fadili Y., Shah S., 2017. A new model for predicting critical gas rate in horizontal and deviated wells. *Journal of Petroleum Science and Engineering*, 150: 154–161. DOI: 10.1016/j.petrol.2016.11.038.
- Franchi J.R., 2010. Integrated Reservoir Asset Management. Original ed. *Gulf Professional Publishing*, 145–166. DOI: 10.1016/C2009-0-62240-6.
- Guo B., Ghalambor A., Xu C., 2006. A systematic approach to predicting liquid loading in gas wells. *SPE Production & Operations*, 21(1): 81–88. DOI: 10.2118/94081-PA.
- Hamidov N.N., Fataliyev V.M., 2016. Experimental study into the effectiveness of the partial gas cycling process in the gas-condensate reservoir development. *Petroleum Science and Technology*, 34(7): 677–684. DOI: 10.1080/10916466.2016.1160112.
- Hassan M.A., Ameer H.A., Hossain I., Khan D., 2023. A novel pessimistic decision tree pruning approach for classification. *6th International Conference on Electrical Information and Communication Technology (EICT)*. DOI: 10.1109/EICT61409.2023.10427593.
- Hearn W.J., 2010. Gas well deliquification application overview. *Abu Dhabi International Petroleum Exhibition and Conference*. DOI: 10.2118/138672-MS.
- Ikpeka P.M., Okolo M.O., 2019. Li and Turner modified model for predicting liquid loading in gas wells. *Journal of Petroleum Exploration and Production Technology*, 9: 1971–1993. DOI: 10.1007/s13202-018-0585-6.
- James G., Witten D., Hastie T., Tibshirani R., Taylor J., 2023. An introduction to statistical learning with applications in Python. 1st ed. *Springer Cham*, 331–343. DOI: 10.1007/978-3-031-38747-0.
- Joseph A., Bassey I., 2022. Prediction of liquid accumulation in gas wells to forecast the critical flowrate and the loading status of individual wells. *Journal of Applied Sciences and Environmental Management*, 26(4): 589–594. DOI: 10.4314/jasem.v26i4.5.
- Lea J.F. Jr., Rowlan L., 2019. Gas Well Deliquification. 3rd ed. *Gulf Professional Publishing*, 9–24. DOI: 10.1016/C2017-0-04099-5.
- Li M., Li S.L., Sun L.T., 2002. New view on continuous removal of liquids from gas wells. *SPE Production & Facilities*, 17(1): 42–46. DOI: 10.2118/75455-PA.
- Luan G., He S., 2012. A new model for the accurate prediction of liquid loading in low-pressure gas wells. *Journal of Canadian Petroleum Technology*, 51(6): 493–498. DOI: 10.2118/158385-PA.
- Luo S., Kelkar M., Pereyra E., Sarica C., 2014. A new comprehensive model for predicting liquid loading in gas wells. *SPE Production & Operations*, 29(4): 337–349. DOI: 10.2118/172501-PA.
- Ming R., He H., 2020. A new approach for predicting critical gas rate in condensate gas wells. *International Journal of Oil, Gas and Coal Technology*, 23(1): 76–91. DOI: 10.1504/IJOGCT.2020.104974.
- Nosseir M.A., Darwich T.A., Sayyoub M.H., El Sallaly M., 2000. A new approach for accurate prediction of loading in gas wells under different flow conditions. *SPE Production & Facilities*, 15(4): 241–246. DOI: 10.2118/66540-PA.
- Pagan E.V., Williams W., Waltrich P.J., 2016. A Simplified Transient Model to Predict Liquid Loading in Gas Wells. *SPE Western Regional Meeting*. DOI: 10.2118/180403-MS.
- Pedregosa F., Varoquaux G., Gramfort A., Michel V., Thirion B., Grisel O., Blondel M., Prettenhofer P., Weiss R., Dubourg V., Vanderplas J., Passos A., Cournapeau D., Brucher M., Perrot M., Duchesnay E., 2011. scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12: 2825–2830.
- Sharma D., Kumar N., 2017. A review on machine learning algorithms, tasks, and applications. *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)*, 6(10): 1548–1552.
- Tugan M.F., 2020. Deliquification techniques for conventional and unconventional gas wells: Review, field cases, and lessons learned for mitigation of liquid loading. *Journal of Natural Gas Science and Engineering*, 83: 103568. DOI: 10.1016/j.jngse.2020.103568.
- Turner R.G., Hubbard M.G., Dukler A.E., 1969. Analysis and prediction of minimum flow rate for the continuous removal of liquids from gas wells. *Journal of Petroleum Technology*, 21(11): 1475–1482. DOI: 10.2118/2198-PA.
- Waltrich P.J., Posada C., Martinez J., Falcone G., Barbosa Jr. J.R., 2015. Experimental investigation on the prediction of liquid loading initiation in gas wells using a long vertical tube. *Journal of Natural Gas Science and Engineering*, 26: 1515–1529. DOI: 10.1016/j.jngse.2015.06.023.
- Wang Z., Bai H., Zhu S., Zhong H., Li Y., 2015. An entrained-droplet model for prediction of minimum flow rate for the continuous removal of liquids from gas wells. *SPE Journal*, 20(5): 1135–1144. DOI: 10.2118/174077-PA.
- Xiao H., Li X.-P., Tan X.-H., Li J., Han C., Xiao H., Han Z., Cao L., 2021. A novel model for calculating critical droplet entrainment rate of gas wells considering droplet deformation and multiple parameters. *Energy Science & Engineering*, 9(6): 812–827. DOI: 10.1002/ese3.836.
- Yang F., 2019. An extended idea about decision trees. *International Conference on Computational Science and Computational Intelligence (CSCI)*, 349–354. DOI: 10.1109/CSCI49370.2019.00068.
- Zhou C., Wu X., Li H., Lin H., Liu X., Cao M., 2016. Optimization of methods for liquid loading prediction in deep condensate gas wells. *Journal of Petroleum Science and Engineering*, 146: 71–80. DOI: 10.1016/j.petrol.2016.04.016.
- Zhou D., Yuan H., 2010. A new model for predicting gas-well liquid loading. *SPE Production & Operations*, 25(2): 172–181. DOI: 10.2118/120580-PA.
- Zhu S., 2009. Stress sensitivity effect of low-permeability gas reservoirs and production analysis of gas wells. *Natural Gas Industry*.



Kanan ALIYEV, Ph.D. candidate
Department of Oil and Gas Engineering
Azerbaijan State Oil and Industry University
34 Azadliq Ave., AZ1010 Baku, Azerbaijan
E-mail: kananaliev.kn@gmail.com