

ISSN 2087-3336 (Print) | 2721-4729 (Online)

TEKNOSAINS: Jurnal Sains, Teknologi dan Informatika

Vol. 13, No. 1, 2026, page. 34-48

<http://jurnal.sttmcileungsi.ac.id/index.php/tekno>

<https://doi.org/10.37373/tekno.v13i1>

A comparative study of tree-based machine learning algorithms for artificial lift optimization

Geovanny Branchiny Imasuly*, Marcia V. Rikumahu

*Department of Petroleum Engineering, Pattimura University, Maluku, Indonesia, 97233. Jln. Ir. M. Putuhena

*Corresponding Author: gimasuly@gmail.com

Submitted: 23/4/2025

Revised: 15/6/2025

Accepted: 14/7/2025

ABSTRACT

The selection of an appropriate artificial lift method is critical in the oil and gas industry to ensure production continuity as reservoir pressure declines. However, current selection processes still largely rely on technical expertise and conventional heuristic approaches, which are often insufficient for handling the complexity of reservoir characteristics and dynamic operational conditions. This study aims to evaluate the performance of three tree-based machine learning algorithms—Decision Tree, Random Forest, and Gradient Boosting—in predicting the optimal artificial lift method. Historical field data, including fluid flow rate, temperature, API gravity, and artificial lift method labels, were used to train the models. The data underwent preprocessing steps such as data cleaning, encoding, and splitting into training and testing sets before being modeled using the Scikit-learn library. The performance of the three models was evaluated using standard classification metrics, including accuracy, precision, recall, and F1-score. The results indicate that the Decision Tree algorithm achieved an accuracy of approximately 81%, Random Forest yielded the highest accuracy at around 94% (with a validation accuracy of 93.67%), while Gradient Boosting performed the least effectively with an accuracy of about 64%. Feature importance and SHAP analysis revealed that temperature was the most influential variable in selecting the artificial lift method, followed by API gravity and fluid flow rate. In conclusion, Random Forest was the most effective model, offering the best combination of accuracy and stability in predicting the optimal artificial lift method.

Keyword: Artificial lift; machine learning; decision tree; random forest; gradient boosting

1. Introduction

Artificial lift is one of the essential techniques in the oil and gas industry for maintaining production rates when reservoir pressure declines [1][2][3][4]. The appropriate selection of *artificial lift* methods is crucial to ensuring the continuity of hydrocarbon production and operational efficiency, particularly in fields with complex and heterogeneous reservoir conditions. However, in practice, the selection process still heavily relies on technical expertise and heuristic approaches, which tend to be subjective and not always optimal [5]. With the advancement of digital technologies, data-driven approaches have been increasingly adopted in decision-making processes within the energy sector, including in oil and gas production management. *Machine learning (ML)*, particularly decision tree-based algorithms, has shown great potential in uncovering hidden patterns and making accurate predictions based on historical well operation data [6]. Algorithms such as *Decision Tree*, *Random*



TEKNOSAINS: Jurnal Sains, Teknologi & Informatika is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License. ISSN 2087-3336 (Print) | 2721-4729 (Online)

Forest, and *Gradient Boosting* offer advantages in terms of interpretability and the ability to handle high-dimensional data and nonlinear relationships among input parameters [7][8].

Previous studies have demonstrated that machine learning approaches can significantly enhance prediction accuracy in various petroleum engineering applications, ranging from reservoir classification to production rate forecasting [9][10]. However, comprehensive evaluations comparing the effectiveness of different decision tree algorithms in the context of *artificial lift* selection remain limited. Yet, choosing the appropriate model can have a direct impact on production performance, cost efficiency, and well longevity.

This study aims to address this gap by comparing the performance of three widely used ML algorithms, *Decision Tree*, *Random Forest*, and *Gradient Boosting*, in predicting the optimal *artificial lift* method. The dataset includes critical parameters such as *flow rate*, *temperature*, *API gravity*, and the *artificial lift* method label applied in the field. Model performance is evaluated using standard metrics, including the *confusion matrix*, *precision*, *recall*, and *F1-score*, to gain a comprehensive understanding of model accuracy. In addition, a feature importance analysis is conducted to identify the most influential parameters in the model's decision-making process. The findings of this study are expected to provide practical insights for data-driven decision-making in the oil and gas industry and to highlight the potential of machine learning as a reliable, adaptive, and efficient tool for well production optimization.

2. Method

This study employs a quantitative experimental approach aimed at comparing the performance of three decision tree-based machine learning algorithms: Decision Tree, Random Forest, and Gradient Boosting in predicting the optimal artificial lift method. The research process is conducted sequentially through the stages of data collection, preprocessing, model training, performance evaluation, and feature analysis. Figure 1 illustrates the research method's flow.

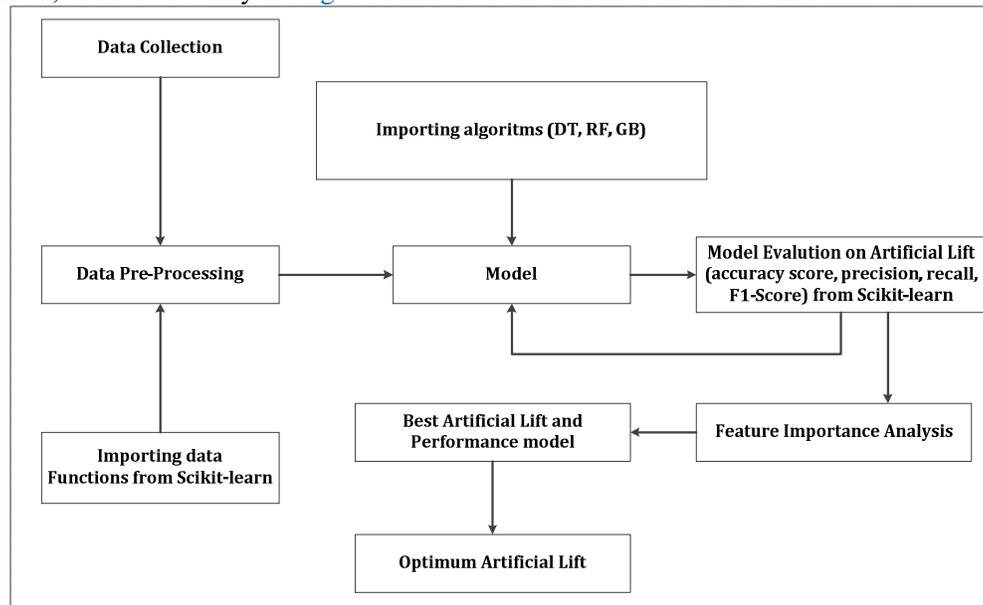


Figure 1. Flow of Research

Data Collection

The dataset analyzed in this study consists of historical operational data from oil and gas wells, encompassing key variables such as fluid flow rate (barrels per day), reservoir temperature (°F), API gravity, and the corresponding artificial lift method applied (ESP, Gas Lift, Hydraulic-Piston Pump, Jet Pump, Plunger Cavity Pump, and Rod Pump). The data were compiled from an internal production information system and underwent anonymization to ensure the confidentiality of operational records. A total of 2,000 data entries were utilized in this research, which were subsequently divided into two subsets: 70% for model training and 30% for model testing.

Data Preprocessing

The data collected may contain errors, missing values, redundant information, and other inconsistencies, which can hinder the performance of machine learning models. Therefore, the data must undergo a cleansing procedure known as preprocessing [11]. Preprocessing in machine learning typically involves five key steps:

- a. Importing the dataset
- b. Importing the libraries
- c. Encoding categorical data
- d. Data visualization
- e. Splitting the dataset into training set and test set

This entire preprocessing workflow can be automated using machine learning algorithms, mathematical modeling, and statistical techniques.

Algorithm and Model Training

This study employs supervised learning algorithms to classify data into specific categories [12]. Three tree-based algorithms from the Scikit-learn library were used in the model training process, as outlined below:

Decision Tree Classifier

The model was constructed using `DecisionTreeClassifier()` from `sklearn.tree`. This algorithm builds a decision tree by splitting the data based on the most informative features [13].

Random Forest Classifier

To improve accuracy, the `RandomForestClassifier()` from `sklearn.ensemble` was employed. This algorithm constructs multiple decision trees and combines their outputs through a majority voting mechanism [13].

Gradient Boosting Classifier

The `GradientBoostingClassifier()` from `sklearn.ensemble` was used to build the model iteratively, focusing on correcting the errors made by previous iterations [13].

Model Evaluation Metrics

Model evaluation is a crucial component in the model development process. The following are commonly used evaluation metrics:

Confusion Matrix

The confusion matrix is one of the most widely used metrics for evaluating the performance of classifiers, particularly in multiclass classification tasks. An illustration of the confusion matrix for binary classification (two classes) [11] is presented in Figure 2.

		Predicted class	
		P	N
Actual class	P	TP	FN
	N	FP	TN

Figure 2. An illustration of confusion matrix

The confusion matrix is recommended to obtain more detailed information for evaluating a model's accuracy in correctly and incorrectly predicting each class. Additionally, the terms used in the confusion matrix are defined in Figure 2, based on a 2×2 matrix for classes A and B:

- True positive (TP): the number of instances where the machine learning model correctly classifies samples as class A.
- False negative (FN): the number of instances where the machine learning model incorrectly classifies samples as not belonging to class A, when they actually do.
- False positive (FP): the number of instances where the machine learning model incorrectly classifies samples as class B, when they actually do not belong to class B.
- True negative (TN): the number of instances where the machine learning model correctly classifies samples as class B.

Based on this matrix, several accuracy metrics can be calculated, including precision, recall, and F1-score.

Precision

Precision is defined as the ratio of true positive samples correctly predicted for a particular class to the total number of samples predicted as belonging to that class. A high precision value indicates that the majority of samples predicted by the model to belong to a specific class indeed truly originate from that class [11].

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive} \quad (1)$$

Recall

Recall refers to the proportion of correctly identified true positive instances compared to the total actual positive instances for a particular class. A high recall value means that most of the actual positive samples are correctly recognized, with only a small number being mistakenly classified into other categories [14].

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Positive} \quad (2)$$

F1-score

F1-score memberikan rata-rata harmonis dari precision dan recall. Pada saat pemilihan model, model dengan nilai F1-score tertinggi dipilih di antara beberapa algoritma klasifikasi [15].

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (3)$$

Accuracy

Evaluation was conducted on the test data to measure the generalization capability of each model [16].

$$Accuracy = \frac{Number\ of\ Corret\ Predictions}{Number\ of\ all\ predictions} \quad (4)$$

Feature Importance Analysis

After training the models, a feature importance analysis was conducted to identify the relative contribution of each input feature to the prediction outcomes [17]. This technique was applied to the Random Forest and Gradient Boosting models, both of which inherently support feature importance evaluation. The insights gained from this analysis help to understand the key factors influencing the selection of artificial lift methods in a data-driven manner.

3. Results and Discussion

Preliminary Data Exploration

The first step involved presenting the frequency distribution of various *Artificial Lift* (AL) methods used in the oil industry. *Artificial Lift* is a technique employed to enhance oil production from wells that lack sufficient reservoir pressure to naturally flow fluids to the surface. The six types of AL methods shown in the diagram are *Jet Pump*, *Sucker Rod Pump* (SRP), *Gas Lift*, *Electric Submersible Pump* (ESP), *Progressive Cavity Pump* (PCP), and *Hydraulic Pumping System* (HPS), as illustrated in Figure 3.

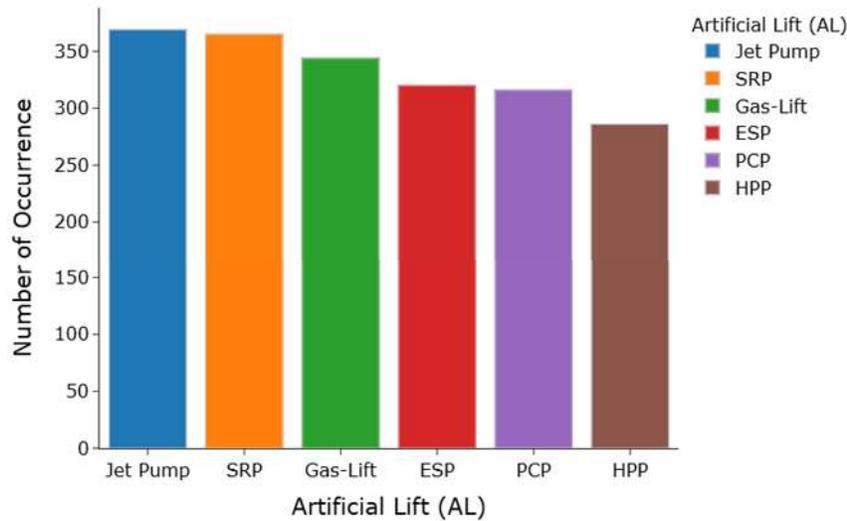


Figure 3. Frequency Distribution of AL

Based on Figure 3, the *Jet Pump* is the most widely used artificial lift method, with approximately 369 instances, highlighting its popularity across various operating conditions, particularly for handling challenging fluids. The *Sucker Rod Pump* (SRP) follows closely with a nearly equal number of occurrences and is commonly employed for heavy oil and medium-depth wells. *Gas Lift* is used in over 344 cases, indicating its effectiveness in wells with high liquid production. Both *Electrical Submersible Pumps* (ESP) and *Progressive Cavity Pumps* (PCP) show similar usage frequencies, around 316 instances each, demonstrating their continued relevance depending on well conditions. Meanwhile, *Hydraulic Pumping Power* (HPP) is the least utilized method, with 286 cases, likely due to the associated costs and system complexity.

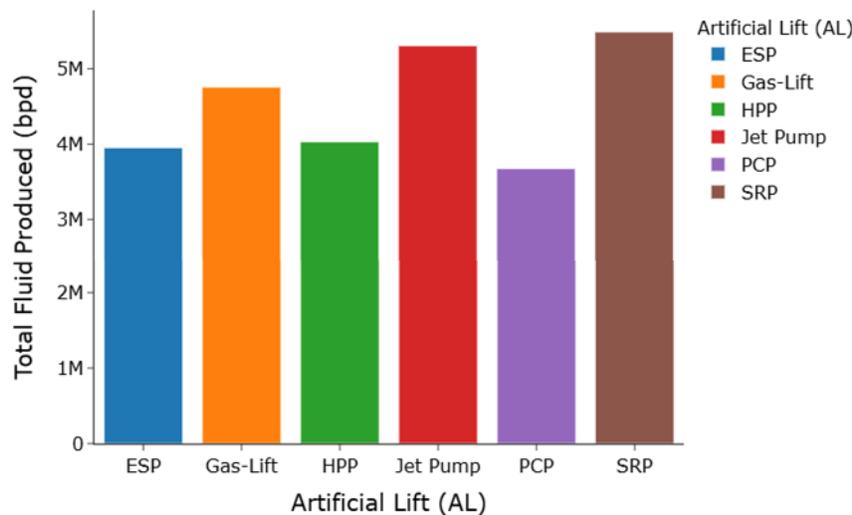


Figure 4. Cumulative fluid production by AL

Figure 4 presents the total *fluid production (bpd)* for each commonly used *Artificial Lift* method in the oil and gas industry. SRP and *Jet Pump* recorded the highest production volumes, each exceeding 5 million bpd, demonstrating their effectiveness in handling low-pressure wells or heavy fluids. *Gas Lift* follows closely with nearly 4.8 million bpd, noted for its flexibility and minimal mechanical components. ESP and *Jet Pump* show similar performance at around 4 million bpd, while PCP records the lowest production at approximately 3.7 million bpd, likely due to technical limitations under certain fluid conditions. This comparison indicates that the selection of an AL method depends not only on its usage frequency (Figure 3) but also on its production performance. Therefore, integrating both parameters is essential in developing a data mining-based decision support system.

Descriptive Statistics

A summary of descriptive statistics is presented for three key variables in the oil and gas production system, namely *Flow Rate (bpd)*, *Temperature (°F)*, and *API Gravity*, each based on 2,000 data samples. These statistics provide an initial overview of the distribution, spread, and characteristics of the values for each parameter. As shown in [Table 1](#).

Table 1. Numerical Summary

	Flow-Rate (bpd)	Temperature (F)	API Gravity
count	2000.000000	2000.000000	2000.000000
mean	13603.820797	217.884695	27.968547
std	7675.177612	71.128887	10.326875
min	186.122857	96.896840	9.123698
25%	7214.747476	151.946890	18.738353
50%	13945.247747	213.704325	29.102568
75%	20242.976139	278.655900	36.846436
max	26696.453963	354.122960	45.719381

LabelEncoder

LabelEncoder is used to perform *label encoding* on the categorical variable 'LABEL' in a dataset, which represents the types of *Artificial Lift* methods, as follows:

- 'ESP' → 0
- 'Gas-Lift' → 1
- 'HPP' (Hydraulic Piston Pump) → 2
- 'Jet Pump' → 3
- 'PCP' (Progressive Cavity Pump) → 4
- 'SRP' (Sucker Rod Pump) → 5

Table 2. Label endcoder data

	Flow-Rate (bpd)	Temperature (F)	API Gravity	Label
0	1817.883890	142.233766	17.276019	4
1	14049.543105	169.647585	10.928627	3
2	14286.195601	144.252621	16.350240	4
3	16754.462488	122.175633	30.406455	2
4	19015.329010	130.158968	24.749859	5
...
1995	4905.466869	324.587214	44.406807	0
1996	10978.587286	96.896840	41.951731	3
1997	23611.653251	330.720778	23.927918	2
1998	23548.017880	312.008711	10.614995	4
1999	15064.847740	218.505406	218.505406	5

[Table 2](#) presents the results of *label encoding* performed on the categorical variable "Label," which represents the type of *Artificial Lift* method. This process is a continuation of the data transformation step, where each string category is converted into discrete numerical values ranging from 0 to 5. The *encoded* data enables compatibility with *machine learning* algorithms that require numerical inputs. This *encoding* step forms the foundation of an intelligent decision system architecture that is accurate, fast, and customizable.

Classification Report and Confusion Matrix

Following the encoding process, the dataset was partitioned into training and testing sets, with 70% of the data allocated for training and the remaining 30% for testing. The classification models were developed using three tree-based machine learning algorithms: *Decision Tree*, *Random Forest*, and *Gradient Boosting*.

To evaluate the performance of these multiclass classification models, several metrics were utilized, including precision, recall, F1-score, and support for each class, along with the overall accuracy. These metrics are comprehensively presented in the classification report, which provides a

detailed view of how well each model performs across the six classes: *ESP*, *PCP*, *SRP*, *Jet Pump*, *Gas Lift*, and *Hydraulic Pump*.

In addition, the confusion matrix was employed to visualize the distribution of correct and incorrect predictions across all classes, allowing for the identification of common misclassification patterns. This analysis enables a thorough comparison of the three models based on both quantitative performance metrics and classification behavior. The following section discusses in detail the performance of each model based on the results of the classification report and the corresponding confusion matrices.

Decision tree classifier

The classification results of the *Decision tree classifier* model are presented in Table 3, which summarizes the evaluation metrics including *precision*, *recall*, *F1-score*, and support (number of samples).

Table 3. Decision tree classification report

	Precision	Recall	F1- Score	Support
0	0.88	0.79	0.83	92
1	0.86	0.86	0.86	99
2	0.77	0.74	0.76	88
3	0.82	0.83	0.83	129
4	0.77	0.88	0.82	86
5	0.85	0.84	0.84	106
accuracy			0.82	600
macro avg	0.83	0.82	0.82	600
weighted	0.83	0.82	0.82	600
avg				

The performance evaluation of the *Decision Tree* model is presented in Table 3. The evaluation metrics include precision, recall, and F1-score for each of the six classes, along with an overall accuracy of 82% based on a test set of 600 samples. The macro and weighted average scores are both 0.82, indicating that the model performs relatively consistently across all classes.

Class 0 achieved the highest precision (0.88), but with a recall of only 0.79, suggesting that while the model is selective, it does not capture all instances of this class effectively. In contrast, Class 4 shows the highest recall (0.88) but with a lower precision (0.77), indicating a tendency toward over-prediction of this class.

The lowest performance was observed in Class 2, with an F1-score of just 0.76, possibly due to imbalanced data distribution or less representative features. Meanwhile, Classes 1, 3, and 5 demonstrated stable performance, with F1-scores ranging between 0.83 and 0.86.

Random forest classifier

The classification results of the *Random Forest classifier* are presented in Table 4, which summarizes the performance evaluation metrics based on *precision*, *recall*, and *F1-score*.

Table 4. Random forest classification report

	Precision	Recall	F1-Score	Support
0	0.95	0.91	0.93	92
1	0.92	0.96	0.94	99
2	0.93	0.92	0.93	88
3	0.93	0.94	0.93	129
4	0.94	0.98	0.96	86
5	0.94	0.92	0.93	106
accuracy			0.94	600
macro avg	0.94	0.94	0.94	600
weighted avg	0.94	0.94	0.94	600

The *Random Forest* model demonstrated excellent performance, achieving an overall accuracy of 94%, a significant improvement over the *Decision Tree* model's 81%. The average precision, recall,

and F1-score (both macro and weighted) were consistently 0.94, indicating balanced performance across all classes. Compared to the Decision Tree, the Random Forest significantly reduced misclassification errors—particularly in classes 2 and 5—and exhibited greater predictive stability, attributed to its ensemble approach that effectively mitigates overfitting.

Gradient boosting classifier

The classification results of the *Gradient Boosting Classifier* are presented in [Table 5](#), which summarizes the performance evaluation metrics based on *precision*, *recall*, and *F1-score*.

Table 5. Gradient boosting classification report

	Precision	Recall	F1-Score	Support
0	0.64	0.61	0.62	92
1	0.57	0.66	0.61	99
2	0.63	0.57	0.60	88
3	0.75	0.70	0.72	129
4	0.72	0.59	0.65	86
5	0.57	0.68	0.62	106
accuracy			0.64	600
macro avg	0.64	0.64	0.64	600
weighted avg	0.65	0.64	0.64	600

The Gradient Boosting Classifier model reached an overall accuracy of 64%, which was lower than the accuracies of the *Decision Tree* (81%) and *Random Forest* (94%) models. Both the macro and weighted average F1-scores were 0.94, indicating that the model maintained high accuracy and consistency across all categories. Nonetheless, the macro and weighted averages for precision, recall, and F1-score hovered around 0.64, highlighting issues related to class imbalance and misclassification among certain classes. The highest F1-score appeared in class 3 (*Jet Pump*), though the difference compared to other classes with lower performance was relatively minor. Additionally, the confusion matrix is illustrated in [Figure 5 - Figure 7](#).

Confusion Matrix Decision Tree

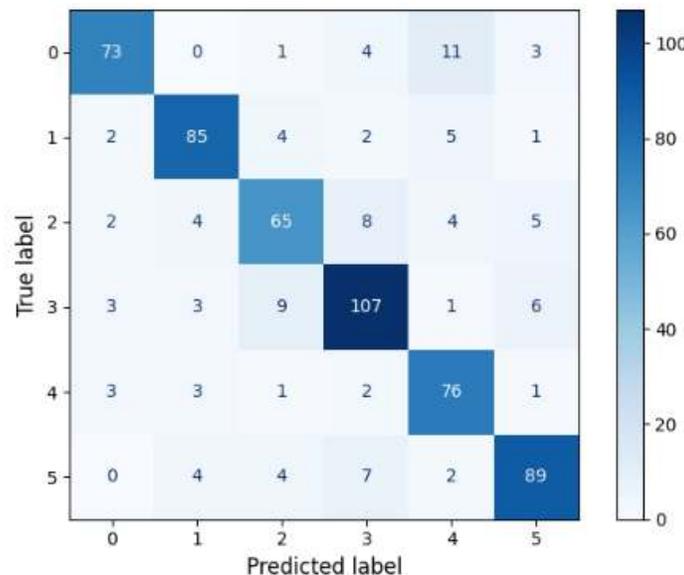


Figure 5. Confusion matrix decision tree

[Figure 5](#) presents the confusion matrix of the Decision Tree model used for six-class classification, demonstrating a reasonably good overall accuracy of 82%. The model performs relatively well in identifying *Jet Pump* (class 3) and *ESP* (class 0). However, it frequently

misclassifies samples from *Gas-Lift* (class 1), *PCP* (class 4), and *SRP* (class 5). For instance, many *SRP* samples are incorrectly predicted as *HPP* or *PCP*, while *Gas-Lift* is often confused with *PCP*. This suggests that the model struggles to differentiate between artificial lift types with overlapping operational characteristics, indicating limitations in its ability to generalize across classes with similar features.

Confusion Matrix Random Forest

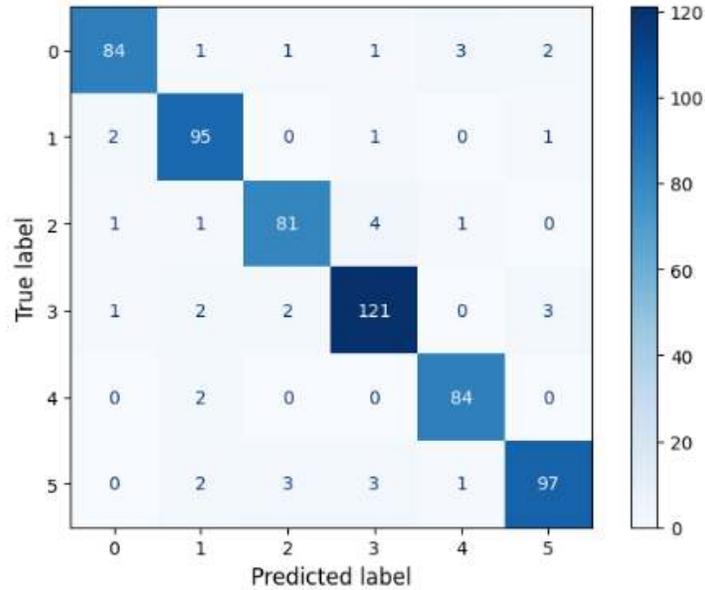


Figure 6. Confusion matrix random forest

Figure 6 presents the confusion matrix of the *Random Forest* model in classifying six types of *artificial lift*. The results demonstrate excellent classification performance, with the majority of predictions concentrated along the main diagonal, indicating correct classifications. The model accurately recognizes all classes, particularly *Jet Pump* (class 3), *Gas-Lift* (class 1), and *PCP* (class 4). Misclassification errors are minimal and evenly distributed, suggesting that the *Random Forest* model effectively distinguishes between class features, outperforming the previous model.

Confusion Matrix Gradient Boosting

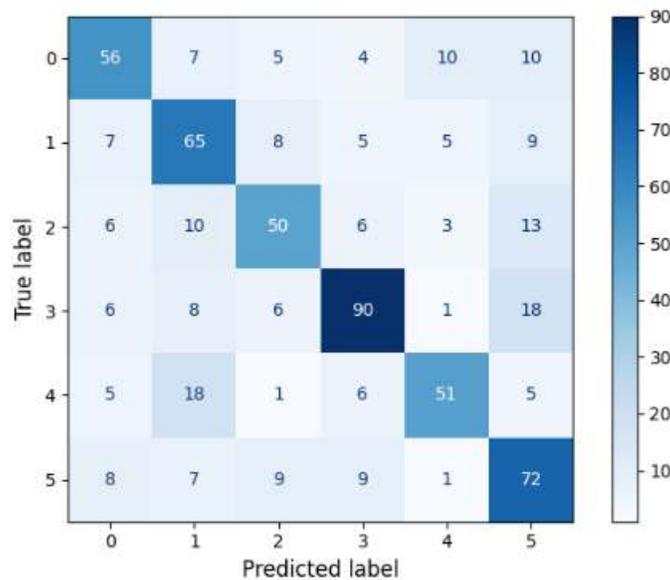


Figure 7. Confusion matrix gradient boosting

Figure 7 presents the confusion matrix of the *Gradient Boosting* model for classifying six types of *artificial lift*. The results indicate reasonably good performance, with most predictions concentrated along the main diagonal, particularly for class 3 (*Jet Pump*) and class 5 (*SRP*). However, notable misclassifications remain, especially for class 4 (*PCP*), which is frequently confused with class 1 (*Gas-Lift*), and class 2 (*HPP*), which is often misclassified as other classes. These findings suggest that while the model achieves a decent level of accuracy, there is still room for improvement in distinguishing classes with similar characteristics.

Model Accuracy Comparison

Compare the accuracy of the *Decision Tree*, *Random Forest*, and *Gradient Boosting* models on both the *training and validation datasets*. This evaluation aims to assess the generalization capability of each model.

Table 6. Comparison of model accuracy on training and validation data

Model	Traning Accuracy	Validation Accuracy
Decision Tree	1.000000	0.821667
Random Forest	1.000000	0.936667
Gradient Boosting	0.904286	0.640000

Table 6 shows that the *Decision Tree* achieved 100% accuracy on the training data but dropped to 82.17% on the validation data, indicating overfitting. In contrast, *Random Forest* maintained stable performance with a validation accuracy of 93.67%, demonstrating good generalization ability due to its ensemble approach. *Gradient Boosting* recorded a training accuracy of 90.43% and a validation accuracy of 64.00%, indicating suboptimal performance with the default configuration. Additionally, Figure 8 compares the accuracy of the three models in classifying *Artificial Lift* methods based on operational features.

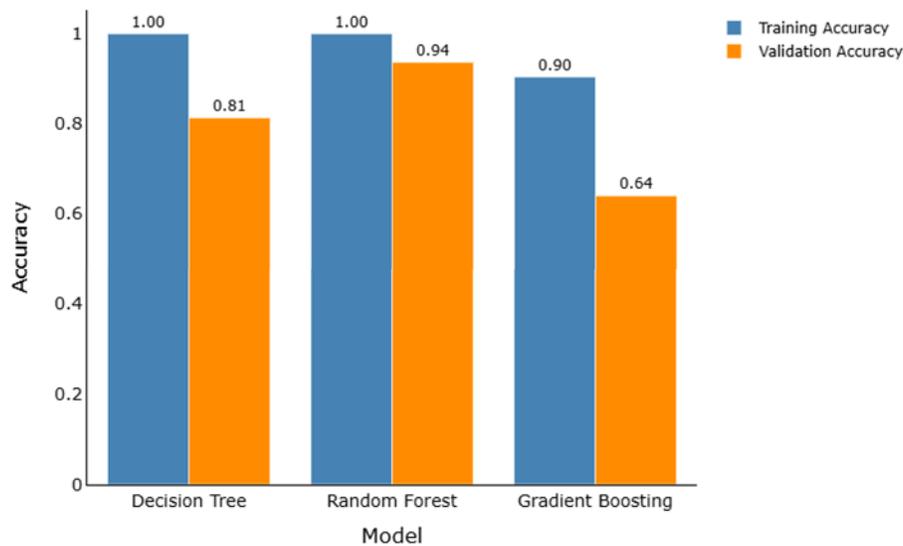


Figure 8. Comparison of training and validation accuracy among different models

Figure 8 also illustrates the comparison between training and validation accuracy, indicating that the *Decision Tree* and *Gradient Boosting* models exhibit signs of overfitting, as evidenced by a substantial gap between training and validation performance. The *Decision Tree* achieved perfect training accuracy (1.00) but only reached 0.82 on the validation set, while *Gradient Boosting* showed a more pronounced drop from 0.90 to 0.64. In contrast, the *Random Forest* model demonstrated the most balanced performance, with a training accuracy of 1.00 and a high validation accuracy of 0.94, indicating strong generalization capability and superior model stability.

Model Performance

Table 7 additional performance indicators used to assess the three primary classification algorithms, Decision Tree, Random Forest, and Gradient Boosting, include ROC AUC (Receiver Operating Characteristic Area Under the Curve) and PR AUC (Precision-Recall Area Under the Curve). These metrics offer a more comprehensive evaluation of model effectiveness, especially when dealing with multiclass scenarios and imbalanced data distributions.

Table 7. Model performance evaluated using ROC AUC and PR AUC Metrics

Model	Mean ROC AUC	Mean PR AUC
Decision Tree	0.891	0.702
Random Forest	0.994	0.970
Gradient Boosting	0.901	0.702

ROC AUC (Receiver Operating Characteristic AUC)

Random Forest demonstrated the best classification performance with an average *ROC AUC* of 0.994, approaching a perfect score. *Gradient Boosting* recorded a *ROC AUC* of 0.901, indicating a reliable performance. Meanwhile, the *Decision Tree* achieved an *ROC AUC* of 0.891, reflecting moderate performance with a tendency toward overfitting.

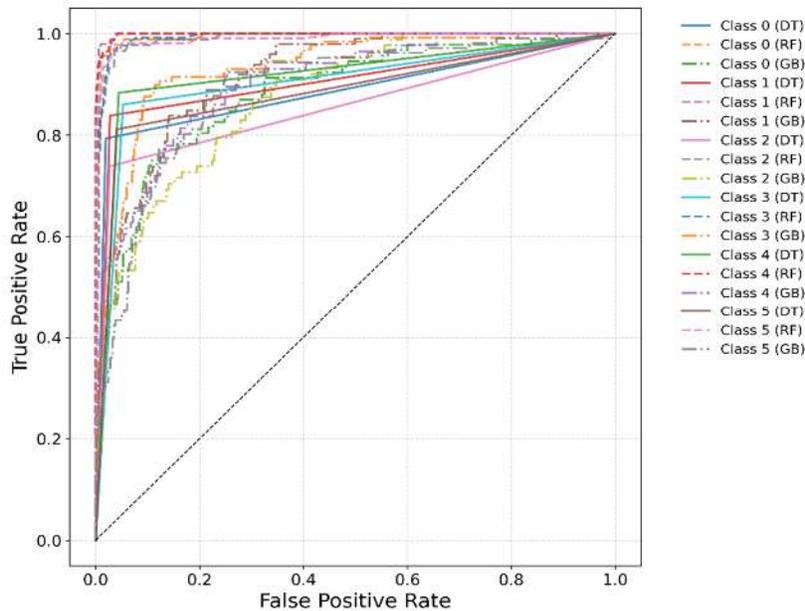


Figure 9. Receiver operating characteristic

Based on Figure 9, the ROC illustrates the model's ability to distinguish between classes, with a curve approaching the top-left corner indicating strong classification performance. *Random Forest* demonstrated the best and most consistent ROC performance across all classes (0–5), reflecting superior classification accuracy and generalization capability. In contrast, the *Decision Tree* and *Gradient Boosting* models exhibited more variable performance, particularly for Class 2 and Class 5, where a noticeable decline in class separability was observed.

PR AUC (Precision-Recall AUC)

Random Forest demonstrated the best performance with a *PR AUC* value of 0.970, indicating high efficiency in minimizing false positives and false negatives, critical aspects in classification systems for operational decision-making. Meanwhile, both *Decision Tree* and *Gradient Boosting*

recorded the same *PR AUC* of 0.702, reflecting limited *precision* and *sensitivity*, which is consistent with the low *F1-scores* observed in these two models.

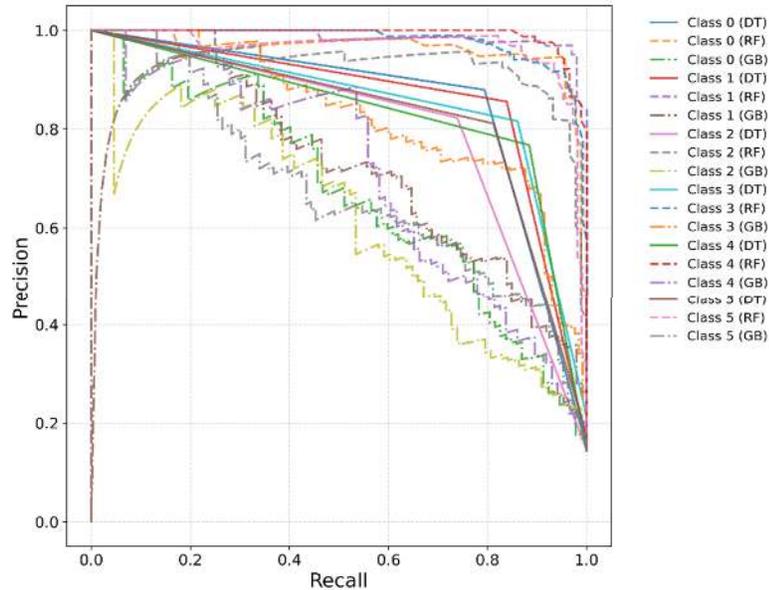


Figure 10. Precision-Recall AUC.

Figure 10 the Precision-Recall curve, illustrates the balance between the model’s *precision* and *recall*. The *Random Forest* algorithm demonstrated the best performance, maintaining a stable curve even at high *recall* values, indicating that *precision* was consistently preserved. In contrast, *Decision Tree* and *Gradient Boosting* exhibited significant declines in *precision* at high *recall* levels. Notably, for classes 4 and 5, *Gradient Boosting* showed the steepest decrease, confirming its limitations on this dataset.

Feature Importance

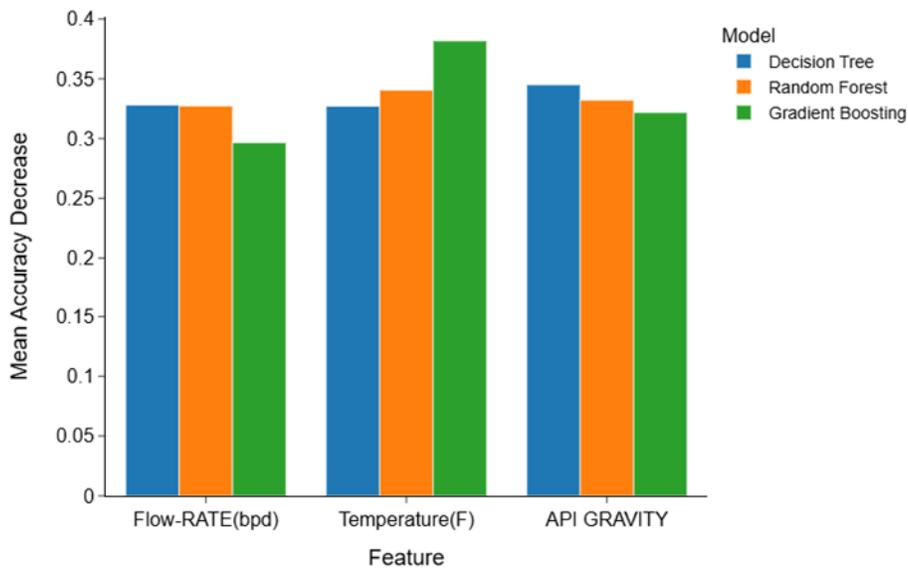


Figure 11. Feature importance comparison across models

Figure 11 presents a comparison of the feature importance scores in the classification models *Decision Tree*, *Random Forest*, and *Gradient Boosting* across three main input parameters: *Flow Rate*

(*bpd*), *Temperature* ($^{\circ}F$), and *API Gravity*. Feature importance reflects the extent to which each input variable contributes to the decision-making process of the model in classifying the type of *Artificial Lift*.

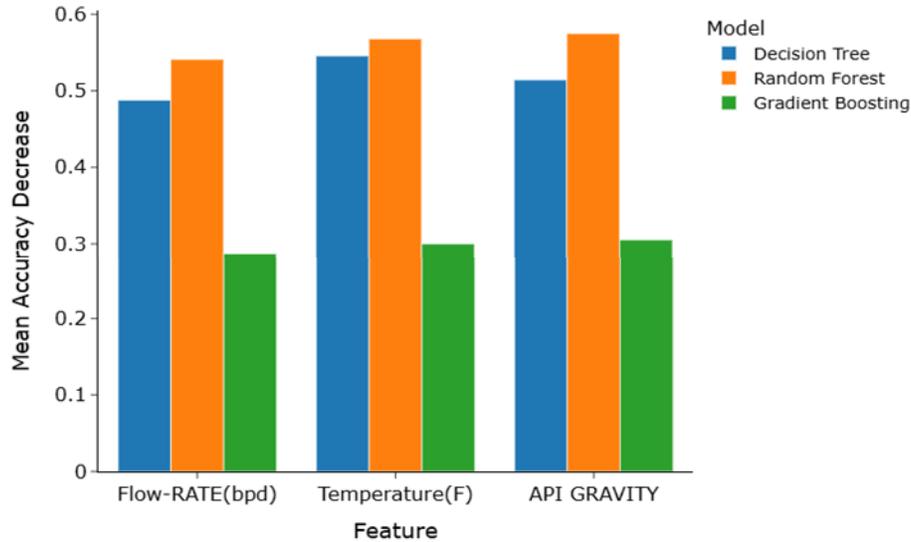


Figure 12. Permutation importance comparison across models

Figure 12 presents the evaluation results of Permutation Feature Importance for three classification models: Decision Tree, Random Forest, and Gradient Boosting, with respect to three key features: Flow Rate (bpd), Temperature ($^{\circ}F$), and API Gravity. Permutation importance measures the average decrease in model accuracy when the values of a specific feature are randomly shuffled, thereby providing insights into the direct contribution of each feature to the model's performance.

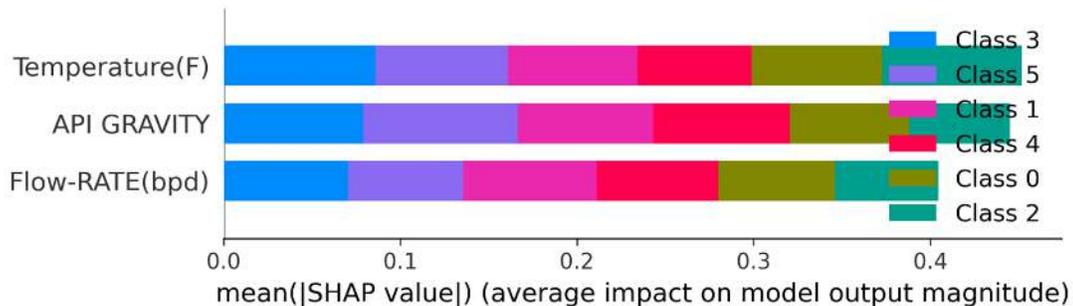


Figure 13. SHAP sensitivity plot for random forest classifier.

Figure 13 presents the *SHAP* (*SHapley Additive exPlanations*) visualization results, illustrating the contribution of each feature to the predictions made by the *Random Forest Classifier* in classifying *Artificial Lift* methods. The plot displays the mean absolute *SHAP* values, reflecting the relative influence of each feature on the model's output. The feature *Temperature* ($^{\circ}F$) demonstrates the highest contribution, particularly for Class 3 (Jet Pump) and Class 5 (SRP), indicating the sensitivity of the Jet Pump method to thermal conditions. *API Gravity* shows a significant impact on Class 1 (Gas Lift) and Class 0 (ESP), aligning with the physical principle of AL selection based on fluid density. Meanwhile, *Flow Rate* (*bpd*) exhibits substantial influence on Class 0 (ESP), 2 (HPP), and 4 (PCP), highlighting the critical role of production volume in the model's decision-making process. The *SHAP* values also indicate that feature contributions are class-specific, reinforcing the findings from the previous feature importance analysis.

4. Conclusion

Based on the evaluation and comparison of the performance of three primary classification algorithms *Decision Tree*, *Random Forest*, and *Gradient Boosting* in identifying *Artificial Lift* methods using parameters such as *flow rate*, *temperature*, and *API gravity*, the results indicate that *Random Forest* outperforms the others, achieving a validation accuracy of 93.67%, an *ROC AUC* of 0.994, and a *PR AUC* of 0.970. While the *Decision Tree* algorithm is suitable as a baseline model, it is prone to overfitting, and *Gradient Boosting* exhibits unstable performance. Feature importance and *SHAP* analysis confirm that temperature plays a dominant role, particularly in predicting the Jet Pump method, followed by *API gravity* and *flow rate*. Overall, *Random Forest* offers the best combination of accuracy, stability, and interpretability, making it the preferred choice for a decision support system in selecting *Artificial Lift* methods.

References

- [1] A. Sugiharto, "OPTIMASI PRODUKSI LAPANGAN MINYAK MENGGUNAKAN METODE ARTIFICIAL LIFT DENGAN ESP PADA LAPANGAN TERINTEGRASI," vol. 02, no. 1, 2012.
- [2] A. Musnal and Fitrianti, "Optimasi Gas Injeksi Pada Sembur Buatan Gas Lift Untuk Meningkatkan Besarnya Laju Produksi Minyak Maksimum Dan Evaluasi penghentian Kegiatan Gas Lift, Pada Lapangan Libo PT. Chevron Pacific Indonesia Duri.," *J. Earth Energy Eng.*, vol. 6, no. 2, pp. 36–47, 2017, doi: <https://doi.org/10.22549/jeeec.v6i2.993>
- [3] F. M. Ridha and E. Indriani, "PEMILIHAN ARTIFICIAL LIFT UNTUK OPTIMASI PRODUKSI SUMUR X DI PT Y LAPANGAN Z," *J. Eksplor. dan Produksi Migas*, vol. 1, no. 2, 2023. <https://doi.org/10.53026/jepm.v1i2.1052>
- [4] I. I. Ekanem and A. E. Ikpe, *A BROAD PERSPECTIVE OF ARTIFICIAL LIFTING MECHANISMS IN THE OIL AND GAS INDUSTRY: OPERATIONAL PANACEA TO ENHANCED OIL RECOVERY*, no. January. 2024.
- [5] M. A. A. Mahdi, M. Amish, and G. Oluyemi, "Artificial Lift Selection Methods in Conventional and Unconventional Wells: A Summary and Review from Old Techniques to Machine Learning Applications," *Int. J. Innov. Sci. Res. Technol.*, vol. 9, no. 3, 2024, doi: [10.38124/ijisrt/IJISRT24MAR2108](https://doi.org/10.38124/ijisrt/IJISRT24MAR2108). <https://doi.org/10.38124/ijisrt/IJISRT24MAR2108>
- [6] M. Shah, A. Kshirsagar, and J. Panchal, "Applications of Artificial Intelligence (AI) and Machine Learning (ML) in the Petroleum Industry," 2022. <https://doi.org/10.1201/9781003279532>
- [7] M. A. A. Mahdi, M. Amish, and G. Oluyemi, "An Artificial Lift Selection Approach Using Machine Learning: A Case Study in Sudan," *Energies*, vol. 16, no. 6, p. 2853, Mar. 2023, doi: [10.3390/en16062853](https://doi.org/10.3390/en16062853). <https://doi.org/10.3390/en16062853>
- [8] A. Sharma *et al.*, "Predicting Gas Separation Efficiency of a Downhole Separator Using Machine Learning," *Energies*, vol. 17, no. 11, 2024, doi: [10.3390/en17112655](https://doi.org/10.3390/en17112655). <https://doi.org/10.3390/en17112655>
- [9] T. Ounsakul, T. Sirirattanachatchawan, W. Pattarachupong, Y. Yokrat, and P. Ekkawong, "Artificial Lift Selection Using Machine Learning," in *International Petroleum Technology Conference*, International Petroleum Technology Conference, Mar. 2019. doi: [10.2523/19423-MS](https://doi.org/10.2523/19423-MS). <https://doi.org/10.2523/19423-MS>
- [10] F. I. Syed, M. Alshamsi, A. K. Dahaghi, and S. Neghabhan, "Artificial lift system optimization using machine learning applications," *Petroleum*, vol. 8, no. 2, pp. 219–226, Jun. 2022, doi: [10.1016/j.petlm.2020.08.003](https://doi.org/10.1016/j.petlm.2020.08.003). <https://doi.org/10.1016/j.petlm.2020.08.003>
- [11] Y. N. Pandey, A. Rastogi, S. Kainkaryam, S. Bhattacharya, and L. Saputelli, *Machine Learning in the Oil and Gas Industry*. Berkeley, CA: Apress, 2020. doi: [10.1007/978-1-4842-6094-4](https://doi.org/10.1007/978-1-4842-6094-4). <https://doi.org/10.1007/978-1-4842-6094-4>
- [12] C. Molnar and T. Freiesleben, "Supervised Machine Learning for Science," 2024.
- [13] G. L. T. C and M. Shang, *Hands-on Supervised Learning With Python*. 2021.
- [14] D. M. W. Powers, "Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation," pp. 37–63, Oct. 2020, doi: <https://doi.org/10.48550/arXiv.2010.16061>.

- [15] D. D. R. Vora and D. G. S. Bhatia, *Python Machine Learning Projects*. 2023.
- [16] D. Truong, *Data Science and Machine Learning for Non-Programmers*. Chapman and Hall/CRC, 2024. doi: 10.1201/9781003162872. <https://doi.org/10.1201/9781003162872>
- [17] S. P. Sotiroudis, S. K. Goudos, and K. Siakavara, “Feature Importances: A Tool to Explain Radio Propagation and Reduce Model Complexity,” *Telecom*, vol. 1, no. 2, pp. 114–125, 2020, doi: 10.3390/telecom1020009. <https://doi.org/10.3390/telecom1020009>