

Web-based geothermal drilling stuck pipe prediction using decision tree algorithm

Rosyihan Muhtadlor¹, Nur Rohman Rosyid¹, Anni Karimatul Fauziyyah¹, Lalu Hendra Permana Setiawan², Irfan Saputra², Pavel Stasa³, Filip Benes³, Muhammad Syafrudin⁴, Ganjar Alfian¹

¹Department of Electrical Engineering and Informatics, Vocational College, Universitas Gadjah Mada, Sleman, Indonesia

²PT. Parama Data Unit, Jakarta, Indonesia

³Department of Economics and Control Systems, Faculty of Mining and Geology, VSB-Technical University of Ostrava, Ostrava, Czech Republic

⁴Department of Artificial Intelligence and Data Science, Sejong University, Seoul, Republic of Korea

Article Info

Article history:

Received Dec 30, 2024

Revised Dec 28, 2025

Accepted Jan 10, 2026

Keywords:

Decision trees

Drilling

Geothermals

Machine learning

Stuck pipe

ABSTRACT

In geothermal drilling operations, data from rig-mounted sensors play a crucial role in maintaining operational efficiency and preventing drilling failures. However, sensor uncertainties and complex subsurface conditions can lead to stuck pipe incidents, causing significant non-productive time and financial losses. This study proposes web-based drilling monitoring system integrated with machine learning (ML) to predict stuck pipe occurrences in geothermal drilling. Several ML algorithms—decision tree (DT), random forest (RF), naïve Bayes (NB), multilayer perceptron (MLP), and support vector machine (SVM)—were evaluated using geothermal drilling data from an Indonesian geothermal project conducted in 2023. To address class imbalance, the synthetic minority oversampling technique (SMOTE) was applied to the training dataset. Feature selection was performed using the correlation coefficient method, and predictions were generated using a 5-minute sliding window. Among the evaluated models, the DT consistently demonstrated superior performance across multiple prediction horizons (PH), achieving an accuracy of 97.4%, precision of 98.6%, recall of 72.9%, and a ROC-AUC of 0.729 using the top five selected features. The trained model was integrated into web-based monitoring platform that provides visualization and predictive alerts. This system enables early detection and better decision-making, helping improve drilling efficiency, reduce stuck pipe risks, and enhance operational safety.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Ganjar Alfian

Department of Electrical Engineering and Informatics, Vocational College, Universitas Gadjah Mada

St. Yacaranda, Depok, Sleman, Yogyakarta 55281, Indonesia

Email: ganjar.alfian@ugm.ac.id

1. INTRODUCTION

Geothermal energy is derived from the earth's internal heat, which is continually generated through the radioactive decay of isotopes within the crust and mantle. This heat can be transferred to the surface through thermal conduction and subsurface fluid circulation, making it accessible for energy applications depending on geological conditions [1]. To harness this energy, wellbores must be drilled to establish flow pathways connecting the surface with subsurface geothermal reservoirs [2]. Geothermal drilling aims to access subsurface thermal reservoirs but faces numerous operational challenges, including hard and abrasive rock formations, high temperatures, and severe fluid losses, which often result in stuck pipes and significant cost overruns [3]. Given the uncertain and complex nature of subsurface conditions, historical data from

previous drilling becomes highly valuable, even though it may not cover all possible scenarios. Therefore, simulation and experimental approaches can complement and enrich the information provided by historical data to support more accurate decision-making in drilling operations [4]. These data can be utilized with proven effective approaches such as an offshoot of artificial intelligence (AI), namely machine learning (ML), to prevent stuck pipes in drilling. This approach enables prediction of stuck pipe so that preventive action can be taken, resulting in more economical drilling costs [5].

Previous study proposed classification ML models, namely decision tree (DT) which was compared with logistic regression (LR), and naïve Bayes (NB). These models were chosen for their simplicity, suitability to large datasets, fast training time, and good predictive ability [6]–[10]. In addition, DT model has several advantages compared to random forest (RF) even though RF is a combination of many DT, namely faster data training and decision making. This is suitable for the case of clogged pipe prediction, which must be done in real-time, requiring speed in modeling and prediction [11]. In addition, sliding window and feature selection methods were also added to improve the performance of prediction model [12], [13].

Previous studies have explored the use of ML models to predict stuck pipe incidents, and techniques such as sliding windows and feature selection have also produced encouraging outcomes. Nevertheless, no prior study has combined these three elements—ML models, sliding window techniques, and feature selection—within a web-based monitoring platform that delivers predictive insights to drilling operators. This study seeks to develop such a system, which includes alert notifications generated from ML predictions and line-graph visualizations of drilling parameters to support clearer interpretation and analysis. The proposed approach is anticipated to enhance the efficiency of geothermal drilling operations by enabling accurate detection of stuck pipe risks and reducing potential operational losses.

2. LITERATURE REVIEW

Khan *et al.* [14] examined the utilization of ML for stuck pipes in oil drilling. This research uses artificial neural network (ANN) and support vector machine (SVM) methods. Based on the research, it was concluded that ANN is better than SVM with 88.89% accuracy, 91.89% recall, and 86.36% specificity. Qodirov and Shestakov [15] used ANN and sliding window methods. This study concluded that the accuracy of the ANN model with a sliding window reached 86%. The prediction model built can be used in the well drilling process to minimize the risk of stuck pipes. Study conducted by Elmousalami and Elaskary [16] used k-nearest neighbors (KNN), DT, RF, extremely randomized trees (extra trees), NB, SVM, LR, ANN, extreme gradient boosting (XGBoost), adaptive boosting (AdaBoost), and stochastic gradient boosting (SGB) methods. Some algorithms add hyperparameters so that the results become more optimal. From the study, it was concluded that extra trees provided the best classification accuracy of 100%.

Elahifar and Hosseini [17] conducted research on the prediction of stuck pipe drilling in case studies of directional and vertical drilling in the Middle East oil fields. The method used in the research is ANN combined with particle swarm optimization (PSO). The results of this study resulted in an accuracy of 84.61% for directional wells and 80% for vertical wells. Shadizadeh [18] proposed solution to address the stuck pipe issue in oil field drilling in Iran involved using an ANN. The study demonstrated that the ANN model achieved 95% accuracy in testing and validating data for dynamic drilling, and over 93% accuracy for static drilling. Xian and Yang [19] discussed a model for early warning when a drilling jam will occur. The methods used in the study were SVM, PSO, traditional cross-validation (CV), and artificial fish swarm algorithm (AFSA). The results of the study concluded that the SVM model combined with AFSA achieved the highest accuracy of 93.47%. Based on the reviewed literature, no prior research has integrated these four components—ML models, sliding window techniques, feature selection, and handling potentially imbalanced datasets—into a unified web-based monitoring platform that provides predictive insights for drilling operations. Therefore, this study introduces several strategies, including evaluating feature selection to determine the optimal set of features, addressing potential data imbalance to improve model performance, and implementing the trained predictive model within a web-based information system to support decision-making and enhance drilling efficiency.

3. METHODOLOGY

Figure 1 shows the stages of the proposed study that involved developing a geothermal drilling stuck pipe prediction system, starting with problem and goal identification, data collection, data preprocessing, model selection, model evaluation, and model deployment. In the data preprocessing stage, the sliding window and feature selection methods are refined. This refinement is achieved by comparing several measures.

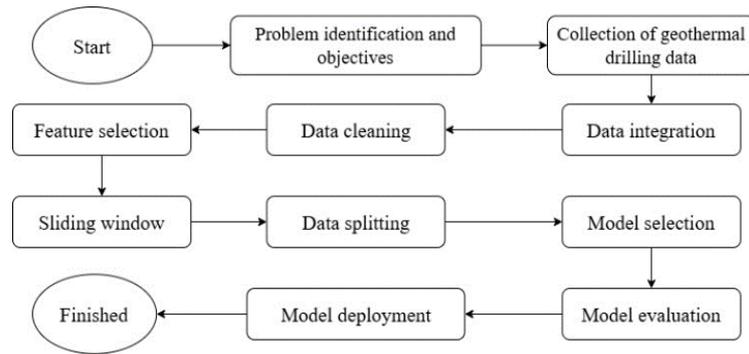


Figure 1. The flowchart of proposed study for drilling prediction

3.1. Data collection

Data collection can be interpreted as the activity of collecting accurate and relevant information for research [20]. This study utilizes a time series dataset on geothermal drilling from a geothermal drilling company in Indonesia. The company specializes in mud logging, directional drilling, and drilling fluids. The dataset employed in this research is derived from recordings captured by internet of things (IoT) sensors. The sensor records drilling data in 10-second intervals which are then sent to the server side. This study uses the drilling dataset from well “A” which has 23 features and 339,065 rows of data taken on June 02, 2023, to July 11, 2023. The label on the dataset is known to consist of two classes, namely normal and stuck drilling. The dataset is imbalanced, with the normal drilling class accounting for 92.52% and the stuck drilling class representing only 7.48%. Such an imbalance can lead ML models to become biased toward the majority class, resulting in poor predictive performance for the minority class, which in this case represents the critical stuck drilling condition. To address this issue, the synthetic minority oversampling technique (SMOTE) will be applied to the training set in this study to generate synthetic samples for the minority class. This approach helps balance the class distribution and expected to improve the model’s ability to accurately detect stuck pipe incidents. Table 1 presents the features of the dataset and their descriptions.

Table 1. Features of the dataset

No	Features	Descriptions
1	Date-time	Sensor-recorded date and time
2	BitDepth	The bit or casing position, indicating the lowest position on the drill string
3	Scfm	Air rate that pumping from aerated unit into the wellbore and combined with water to produce low density fluid less than water when entering the loss zone
4	MudCondIn	Conductivity of drilling mud as it enters the wellbore
5	BlockPos	Shows the position of the traveling block
6	WOB	Force exerted on the drill bit
7	ROPi	Drill bit speed drilling underground formations
8	BVDDepth	Speed in penetrating underground geological formations
9	MudCondOut	Conductivity of drilling mud as it exits the wellbore
10	Torque	Rotating force generated
11	RPM	Rotation speed for one minute
12	Hkld	Total of the weight borne by the hook on the drilling equipment
13	LogDepth	Depth of the borehole at the time of log data measurement
14	H2S_1	Refers to the concentration level of hydrogen sulfide (toxic and corrosive gas)
15	MudFlowOutp	Mud flow discharge that rises to the surface
16	TotSPM	Repetitive motion performed by drilling mud pump in one minute
17	SpPress	Pressure in the support pipe system of drilling equipment
18	MudFlowIn	Mud flow rate to be pumped to the well
19	CO2_1	Carbon dioxide (CO2) gas concentration
20	Gas	Concentration of discovered gas
21	MudTempIn	Mud temperature entering the borehole
22	MudTempOut	Temperature of mud exiting the borehole and returning to the surface
23	TankVolTot	Total volume in the tank

3.2. Feature selection with correlation coefficient

Correlation coefficient (r) is a method to measure the relationship between two variables. The resulting value of this method ranges from -1 to +1, with -1 indicating a perfect negative correlation, +1 indicating a perfect positive correlation, and 0 meaning there is no correlation between the variables [21]. In this study, the correlation coefficient r is employed for feature selection to identify the most influential

features affecting the output class. We chose the top 5, 10, and 15 features with the highest correlation coefficients from Table 1 to serve as input features for the ML models. The formula for calculating the Pearson correlation coefficient is shown in (1).

$$r = \frac{n \sum_{i=1}^n X_i Y_i - \sum_{i=1}^n X_i \sum_{i=1}^n Y_i}{\sqrt{n \sum_{i=1}^n X_i^2 - (\sum_{i=1}^n X_i)^2} \sqrt{n \sum_{i=1}^n Y_i^2 - (\sum_{i=1}^n Y_i)^2}} \quad (1)$$

3.3. Data transformation with sliding window

To process a time series dataset where future predictions are based on values from previous time periods, the sliding window method is selected as an appropriate approach [22]. Sliding window is a method of analyzing a subset of data that “rolls” over time. This algorithm is highly adaptable and thus robust to variations in data that change suddenly [23]. Sliding windows can be used in various computing fields such as data analysis and pattern matching algorithms. Sliding windows use the concept of a constant-sized “window” applied to each input data to handle successive portions of the data. Basically, this technique uses a pre-existing data set to predict future values [24]. Figure 2 illustrates a sliding window method, using data from the previous 5 minutes as input to forecast the next 5-minute prediction horizon (PH).

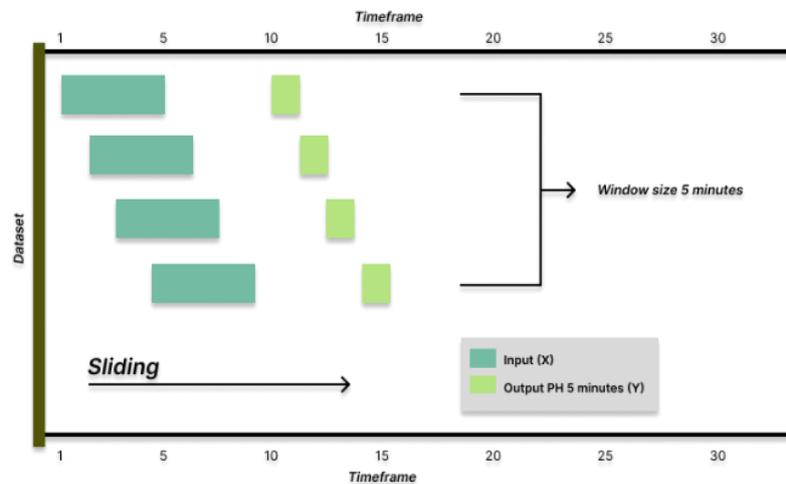


Figure 2. Sliding window illustration

3.4. Decision tree model

DT is a prediction model that is very suitable for classification [25]. This model is a directed acyclic graph that begins at a root node, where each internal (non-terminal) node tests a single feature. Each branch corresponds to the outcome of that test, guiding the instance down a specific path. This process continues through successive feature evaluations until a terminal (leaf) node is reached, which provides the final class prediction [26].

The DT model was chosen because the model is simple, suitable for large datasets, fast training time, and good prediction ability so that it is suitable for use in the case of stuck pipe prediction in drilling [6], [7], [25]. In addition, in this case the DT model has several advantages compared to RF, namely faster data training and decision making which is suitable for stuck pipe prediction cases that require speed in modeling and prediction [11]. Figure 3 shows the DT formation flowchart. Hyperparameter optimization was conducted for the DT model, resulting in the optimal configuration of `max_depth=5`, `min_samples_split=2`, `min_samples_leaf=2`, and `criterion=entropy`. To ensure the reliability of the model, each predictive model was executed repeatedly under identical conditions to validate its robustness and consistency.

3.5. Evaluation metrics for the models

Evaluation metrics measure model performance by predicting testing set that the model has never seen before. Thus, it can be known how well the model performs in solving the given task [27], [28]. This study employs a hold-out validation method, which is more appropriate for time series data to preserve temporal dependencies. In this approach, the dataset is divided into two portions, with the first 47% used for training and the remaining 53% reserved for testing. We ensured that the minority class was represented in both the training and testing sets. Therefore, this ratio was chosen to maintain a stratified distribution, ensuring that the minority class appears in both subsets, as the stuck condition occurs approximately in the

middle of the time series dataset. Table 2 shows an explanation of the confusion matrix components, where true positive (TP) is a true positive prediction, false positive (FP) is a false positive prediction, true negative (TN) is a true negative prediction, and false negative (FN) is a false negative prediction [29]. The positive label indicates a stuck pipe condition, while the negative label corresponds to normal drilling conditions. Model performance was evaluated using several metrics, including accuracy, precision, recall, F1-score, and receiver operating characteristic area under the curve (ROC-AUC), all derived from values in the confusion matrix. Accuracy measures the proportion of correctly classified instances among all samples. Precision reflects the proportion of correctly identified positive cases out of all predicted positive cases, while recall (or sensitivity) indicates the model's ability to detect actual positive cases. The F1-score, representing the harmonic mean of precision and recall, provides a balanced assessment of both metrics. Additionally, the ROC-AUC evaluates the model's ability to distinguish between positive and negative classes, offering an overall measure of classification effectiveness. To address class imbalance, the SMOTE was applied to the training set, ensuring a more balanced class distribution during model training.

3.6. Model deployment

Once the optimal model is obtained, it is deployed into a web-based system using the Streamlit framework. This web application serves as a prediction tool that processes historical drilling data to identify potential stuck pipe incidents. In practical implementation, the data is collected from sensors installed on the drilling rig and stored in a central database for analysis. When valid input data is provided, the system generates predictions indicating either normal drilling conditions or potential stuck pipe events. Additionally, the platform features interactive visualization tools that allow users to monitor and analyze historical drilling parameters over time, supporting early detection and informed decision-making during drilling operations.

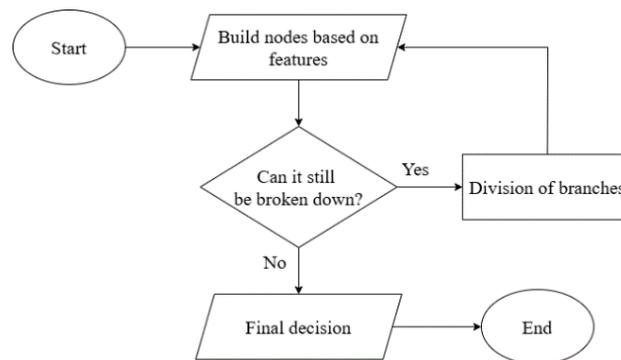


Figure 3. Decision tree flowchart

Table 2. Confusion matrix

Actual	Predicted	
	Positive	Negative
Positive	True positive (TP)	False negative (FN)
Negative	False positive (FP)	True negative (TN)

4. RESULT AND DISCUSSION

4.1. Model performance evaluation

In this section, we compared the performance of the proposed DT model with other ML algorithms, namely NB, RF, multilayer perceptron (MLP), and SVM. In addition, we also investigate the impact of applying the SMOTE on model performance. Tables 3 to 5 summarized the results of each model at three PH: 5 minutes (PH5), 10 minutes (PH10), and 15 minutes (PH15). The models utilize the last five minutes of drilling data to predict whether a stuck pipe incident will occur in the next 5, 10, or 15 minutes, respectively.

Based on the results presented in Tables 3 to 5, the DT consistently achieves strong and stable performance across all PH. The proposed DT model, developed using the top five selected features, achieved an accuracy of 0.973-0.974, precision of 0.986, and recall of 0.728-0.729. The F1-score, ranging between 0.624-0.629, indicates a balanced performance between precision and recall, while the ROC-AUC values (around 0.728-0.729) further confirm the model's good discriminative capability in identifying the stuck pipe class. When compared to other algorithms, the DT model outperforms NB and MLP in terms of both precision and overall stability. Although the RF model shows comparable results, DT offers a simpler

structure with lower computational complexity, making it more suitable for integration into the proposed web-based monitoring system. The SVM, in contrast, demonstrates the weakest performance across all metrics, particularly in F1-score and ROC-AUC, suggesting limited capability in handling the non-linear characteristics of the dataset.

Table 3. Performance of prediction model for PH 5

Model	Accuracy	Recall	Precision	F1-score	ROC AUC
NB	0.933	0.715	0.667	0.410	0.736
RF	0.974	0.729	0.986	0.629	0.730
MLP	0.901	0.687	0.602	0.308	0.514
SVM	0.911	0.489	0.487	0.022	0.500
DT	0.973	0.728	0.986	0.626	0.728
NB+SMOTE	0.937	0.717	0.677	0.423	0.715
RF+SMOTE	0.973	0.729	0.986	0.628	0.731
MLP+SMOTE	0.966	0.711	0.868	0.549	0.500
SVM+SMOTE	0.908	0.505	0.505	0.058	0.500
DT+SMOTE	0.973	0.727	0.986	0.624	0.727

Table 4. Performance of prediction model for PH10

Model	Accuracy	Recall	Precision	F1-score	ROC AUC
NB	0.933	0.717	0.667	0.411	0.738
RF	0.974	0.730	0.987	0.631	0.732
MLP	0.902	0.695	0.606	0.317	0.506
SVM	0.949	0.511	0.617	0.048	0.500
DT	0.974	0.729	0.986	0.628	0.729
NB+SMOTE	0.937	0.717	0.677	0.423	0.715
RF+SMOTE	0.973	0.729	0.986	0.628	0.731
MLP+SMOTE	0.966	0.711	0.868	0.549	0.500
SVM+SMOTE	0.908	0.505	0.505	0.058	0.500
DT+SMOTE	0.973	0.727	0.986	0.624	0.727

Table 5. Performance of prediction model for PH15

Model	Accuracy	Recall	Precision	F1-score	ROC AUC
NB	0.933	0.718	0.667	0.412	0.739
RF	0.974	0.731	0.987	0.632	0.734
MLP	0.903	0.688	0.604	0.311	0.456
SVM	0.959	0.584	0.980	0.289	0.500
DT	0.974	0.728	0.986	0.627	0.728
NB+SMOTE	0.937	0.719	0.678	0.426	0.717
RF+SMOTE	0.974	0.731	0.987	0.632	0.734
MLP+SMOTE	0.965	0.712	0.862	0.549	0.500
SVM+SMOTE	0.931	0.523	0.541	0.092	0.500
DT+SMOTE	0.974	0.729	0.986	0.629	0.729

4.2. Most influential features

The prediction models presented in Tables 3 to 5 were developed using the five most relevant features identified through correlation coefficient-based feature selection. Table 6 lists these features in descending order of their correlation with the stuck pipe label, with Scfm, Hkld, BitDepth, BVDepth, and LogDepth emerging as the top contributors. These features reflect key operational parameters that influence drilling performance, particularly the mechanical and hydraulic conditions associated with potential stuck pipe incidents. By focusing on these five features, the proposed DT model effectively reduces data dimensionality and computational complexity while maintaining robust predictive capability. This selective approach helps minimize the risk of overfitting and enhances model interpretability—important factors for deployment in a practical, web-based monitoring environment. The strong correlation values of Scfm and Hkld indicate that air flow and hook load play a significant role in identifying early signs of stuck pipe conditions, reinforcing their importance in predictive analysis for geothermal drilling operations.

Table 6. Five most relevant features

Feature	Correlation coefficient score
Scfm	0.486942
Hkld	0.442078
BitDepth	0.292394
BVDepth	0.284145
LogDepth	0.274124

The application of SMOTE shows only marginal improvements in model performance, particularly for algorithms such as NB and MLP. For the DT model, SMOTE did not significantly alter the performance metrics, which indicates that the model can effectively handle the imbalanced dataset even without oversampling. Nevertheless, applying SMOTE helped ensure fairer learning conditions across models and provided consistent minority class representation during training. Overall, the results demonstrate that the proposed DT model provides the most reliable and interpretable solution for stuck pipe prediction. Its consistent performance across all PHs, combined with its simplicity and suitability for deployment within a web-based monitoring platform, supports its selection as the best-performing predictive model in this study. Figure 4 presents the confusion matrix of the DT model, while Figure 5 illustrates the confusion matrix of the DT model combined with SMOTE. Meanwhile, Figure 6 displays the ROC curves for all prediction models, providing a comparative visualization of their classification performance.

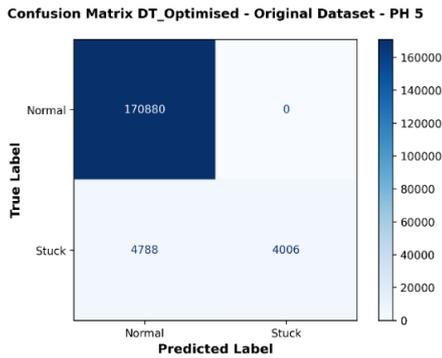


Figure 4. Confusion matrix for DT model

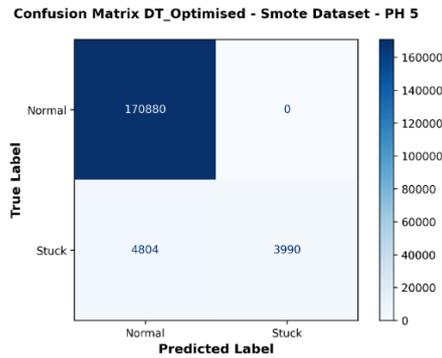


Figure 5. Confusion matrix for DT+SMOTE model

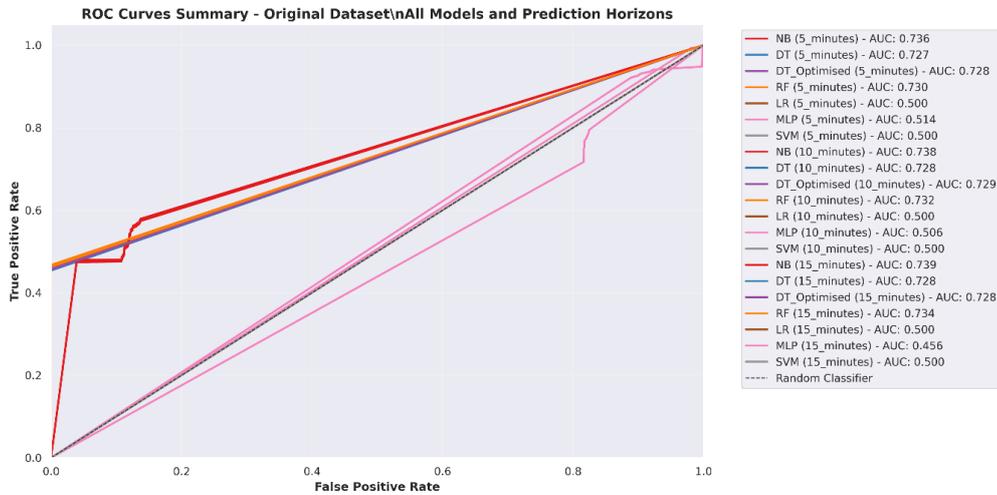


Figure 6. ROC Curves for all prediction models

4.3. Model deployment

After the optimal model was successfully trained, it was integrated into a web-based information system. The system begins with a login page (Figure 7), which serves as the entry point for users to access the application. Users are required to enter their registered username and password to gain access. Upon successful login, the dashboard interface is displayed, providing navigation to several key features, including the prediction page and the data source page for data visualization. A logout option is also available to securely end the session. The prediction page, accessible through the navigation menu, displays the prediction results generated from the input data, as illustrated in Figure 8. Meanwhile, the visualization page allows users to view historical drilling datasets stored in the database, presented as interactive line charts. Users can specify a desired time range through an input form to visualize data within that period. Figure 9

shows an example of the geothermal drilling data visualization interface, demonstrating how the system supports monitoring and analysis of drilling conditions.

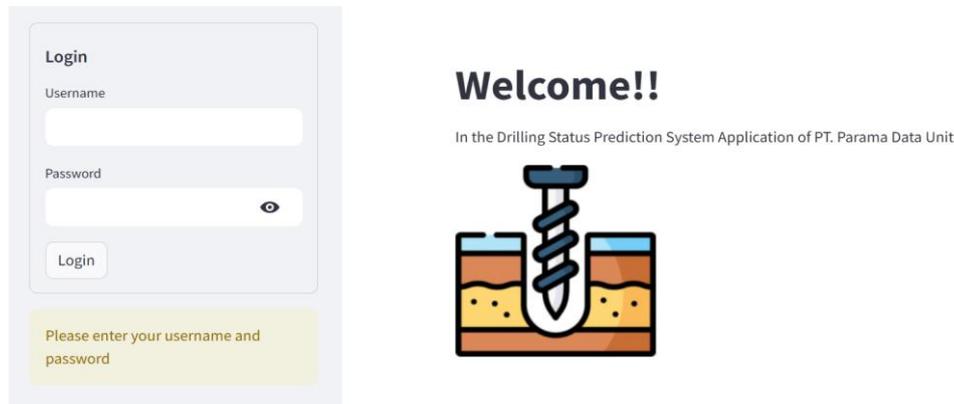


Figure 7. Login page

Previously uploaded data :

	Date-Time	BitDepth	Scfm	MudCondIn	BlockPos	WOB	ROPI	BVDepth	MudCondOut
0	06/30/2022 23:55	126	0	0	12	0	0	128	0
1	06/30/2022 23:55	125	0	0	13	0	0	128	0
2	06/30/2022 23:55	124	0	0	14	0	0	128	0
3	06/30/2022 23:55	123	0	0	15	0	0	128	0
4	06/30/2022 23:55	123	0	0	15	0	0	127	0
5	06/30/2022 23:55	122	0	0	16	0	0	127	0
6	06/30/2022 23:56	121	0	0	17	0	0	127	0
7	06/30/2022 23:56	120	0	0	18	0	0	126	0
8	06/30/2022 23:56	119	0	0	19	0	0	126	0
9	06/30/2022 23:56	118	0	0	20	0	0	125	0

Normal drilling, not stuck

Figure 8. Prediction page



Figure 9. Visualization page

5. CONCLUSION

The proposed DT model, developed using the top five features identified through the correlation coefficient method, achieved the most optimal performance among all evaluated models for predicting stuck pipe incidents in geothermal drilling operations. With an accuracy of 97.3-97.4%, precision of 98.6%, recall of 72.8-72.9%, and ROC-AUC of approximately 0.728-0.729, the DT model demonstrated superior predictive capability and balanced performance across evaluation metrics. The incorporation of a 5-minute sliding window and the five most relevant features contributed to improving both computational efficiency and model interpretability. When compared with other ML algorithms—NB, RF, MLP, and SVM—the proposed DT consistently achieved comparable or better results across all PHs. While the RF model showed similar accuracy, the DT was favored for its simpler structure, lower computational cost, and suitability for deployment in a web-based system. The trained DT model was successfully integrated into a web-based drilling monitoring platform, providing prediction-based alerts and data visualization to assist operators in identifying potential stuck pipe events. This system is expected to improve operational decision-making, enhance drilling efficiency, and reduce the likelihood of costly downtime. For future research, further enhancement of model performance is planned by incorporating deep learning approaches and advanced feature extraction techniques to better capture complex temporal dependencies. Expanding the dataset with larger and more diverse drilling records is also recommended to strengthen the model’s generalization across varying geological and operational conditions.

FUNDING INFORMATION

Authors state no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Rosyihan Muhtadlor		✓	✓	✓	✓	✓		✓	✓	✓	✓			
Nur Rohman Rosyid	✓				✓							✓		✓
Anni Karimatul Fauziyyah		✓				✓			✓		✓	✓		
Lalu Hendra Permana		✓		✓		✓				✓		✓		✓
Setiawan														
Irfan Saputra	✓		✓			✓		✓		✓		✓		
Pavel Stasa	✓			✓	✓			✓		✓	✓			
Filip Benes	✓			✓		✓	✓			✓	✓			
Muhammad Syafrudin		✓	✓		✓	✓				✓			✓	
Ganjar Alfian	✓	✓				✓			✓	✓		✓	✓	

C : Conceptualization

M : Methodology

So : Software

Va : Validation

Fo : Formal analysis

I : Investigation

R : Resources

D : Data Curation

O : Writing - Original Draft

E : Writing - Review & Editing

Vi : Visualization

Su : Supervision

P : Project administration

Fu : Funding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author, [GA], upon reasonable request.

REFERENCES

- [1] J. Zheng *et al.*, "Feasibility of shallow geothermal energy for industrial heating using high-temperature heat pumps in Sichuan Province, China," *Energy for Sustainable Development*, vol. 88, Oct. 2025, doi: 10.1016/j.esd.2025.101785.
- [2] X. Song *et al.*, "Review of high-temperature geothermal drilling and exploitation technologies," *Gondwana Research*, vol. 122, pp. 315–330, Oct. 2023, doi: 10.1016/j.gr.2022.10.013.
- [3] C. Vivas, S. Salehi, J. D. Tuttle, and B. Rickard, "Challenges and opportunities of geothermal drilling for renewable energy generation," *Transactions-Geothermal Resources Council*, vol. 44, pp. 904–918, 2020.
- [4] S. C. H. Geekiyanage, A. Tunkiel, and D. Sui, "Drilling data quality improvement and information extraction with case studies," *Journal of Petroleum Exploration and Production Technology*, vol. 11, no. 2, pp. 819–837, Feb. 2020, doi: 10.1007/s13202-020-01024-x.
- [5] V. S. de M. Siqueira, M. A. S. L. Cuadros, C. J. Munaro, and G. M. de Almeida, "Expert system for early sign stuck pipe detection: feature engineering and fuzzy logic approach," *Engineering Applications of Artificial Intelligence*, vol. 127, Jan. 2024, doi: 10.1016/j.engappai.2023.107229.
- [6] V. F. C. Souza, F. Cicalese, E. S. Laber, and M. Molinaro, "Decision trees with short explainable rules," *Theoretical Computer Science*, vol. 1047, Sep. 2025, doi: 10.1016/j.tcs.2025.115344.
- [7] C. Wang, X. Cui, and S. An, "Neighborhood rough decision tree," *Information Sciences*, vol. 717, Nov. 2025, doi: 10.1016/j.ins.2025.122266.
- [8] A. Bailly *et al.*, "Effects of dataset size and interactions on the prediction performance of logistic regression and deep learning models," *Computer Methods and Programs in Biomedicine*, vol. 213, Jan. 2022, doi: 10.1016/j.cmpb.2021.106504.
- [9] J. K. Harris, "Primer on binary logistic regression," *Family Medicine and Community Health*, vol. 9, Dec. 2021, doi: 10.1136/fmch-2021-001290.
- [10] S. Chen, G. I. Webb, L. Liu, and X. Ma, "A novel selective naïve Bayes algorithm," *Knowledge-Based Systems*, vol. 192, Mar. 2020, doi: 10.1016/j.knsys.2019.105361.
- [11] L. Rokach and O. Maimon, *Data mining with decision trees: theory and applications*, 2nd ed. Hackensack, New Jersey: World Scientific, 2015.
- [12] H. Gong, Y. Li, J. Zhang, B. Zhang, and X. Wang, "A new filter feature selection algorithm for classification task by ensembling pearson correlation coefficient and mutual information," *Engineering Applications of Artificial Intelligence*, vol. 131, May 2024, doi: 10.1016/j.engappai.2024.107865.
- [13] F. Yue, L. Miao, and Z. Zhou, "An equivalent backtracking coarse alignment method with dynamic optimal sliding window integration," *Measurement: Journal of the International Measurement Confederation*, vol. 256, Dec. 2025, doi: 10.1016/j.measurement.2025.118290.
- [14] J. A. Khan *et al.*, "Comparison of machine learning classifiers for accurate prediction of real-time stuck pipe incidents," *Energies*, vol. 13, no. 14, Jul. 2020, doi: 10.3390/en13143683.

- [15] S. Qodirov and A. Shestakov, "Development of artificial neural network for predicting drill pipe sticking in real-time well drilling process," in *2020 Global Smart Industry Conference, GloSIC 2020*, Nov. 2020, pp. 139–144, doi: 10.1109/GloSIC50886.2020.9267873.
- [16] H. H. Elmousalami and M. Elaskary, "Drilling stuck pipe classification and mitigation in the Gulf of Suez oil fields using artificial intelligence," *Journal of Petroleum Exploration and Production Technology*, vol. 10, no. 5, pp. 2055–2068, Jun. 2020, doi: 10.1007/s13202-020-00857-w.
- [17] B. Elahifar and E. Hosseini, "Machine learning algorithm for prediction of stuck pipe incidents using statistical data: case study in middle east oil fields," *Journal of Petroleum Exploration and Production Technology*, vol. 12, no. 7, pp. 2019–2045, Jul. 2022, doi: 10.1007/s13202-021-01436-3.
- [18] S. Shadzadeh, "Drilling stuck pipe prediction in Iranian oil fields: an artificial neural network approach," *Iranian Journal of Chemical Engineering*, vol. 7, no. 4, pp. 29–41, 2010.
- [19] Z. Xian and H. Yang, "An early warning model for the stuck-in medical drilling process based on the artificial fish swarm algorithm and SVM," *Distributed and Parallel Databases*, vol. 40, no. 4, pp. 779–796, Dec. 2022, doi: 10.1007/s10619-021-07344-z.
- [20] H. Taherdoost, "Data collection methods and tools for research; a step-by-step guide to choose data collection technique for academic and business research projects," *International Journal of Academic Research in Management*, vol. 2021, no. 1, pp. 10–38, 2021.
- [21] E. Temizhan, H. Mirtagioglu, and M. Mendes, "Which correlation coefficient should be used for investigating relations between quantitative variables," *American Academic Scientific Research Journal for Engineering*, pp. 2313–4402, 2022.
- [22] C. Ma, W. Li, J. Cao, J. Du, Q. Li, and R. Gravina, "Adaptive sliding window based activity recognition for assisted livings," *Information Fusion*, vol. 53, pp. 55–65, Jan. 2020, doi: 10.1016/j.inffus.2019.06.013.
- [23] J. Wang, W. Jiang, Z. Li, and Y. Lu, "A new multi-scale sliding window LSTM framework (MSSW-LSTM): a case study for GNSS time-series prediction," *Remote Sensing*, vol. 13, no. 16, Aug. 2021, doi: 10.3390/rs13163328.
- [24] C. Chen *et al.*, "Forecast of rainfall distribution based on fixed sliding window long short-term memory," *Engineering Applications of Computational Fluid Mechanics*, vol. 16, no. 1, pp. 248–261, Dec. 2022, doi: 10.1080/19942060.2021.2009374.
- [25] B. Charbuty and A. Abdulazeez, "Classification based on decision tree algorithm for machine learning," *Journal of Applied Science and Technology Trends*, vol. 2, no. 01, pp. 20–28, Mar. 2021, doi: 10.38094/jast20165.
- [26] Y. Izza, A. Ignatiev, and J. M.-Silva, "On explaining decision trees," *arXiv:2010.11034*, Oct. 2020.
- [27] Ž. Vujović, "Classification model evaluation metrics," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 6, pp. 599–606, 2021, doi: 10.14569/IJACSA.2021.0120670.
- [28] D. Müller, I. S.-Rey, and F. Kramer, "Towards a guideline for evaluation metrics in medical image segmentation," *BMC Research Notes*, vol. 15, no. 1, Dec. 2022, doi: 10.1186/s13104-022-06096-y.
- [29] A. A.-Duart, E. Mariotti, D. G.-Gasulla, and J. M. A.-Moral, "A confusion matrix for evaluating feature attribution methods," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, Jun. 2023, pp. 3709–3714, doi: 10.1109/CVPRW59228.2023.00380.

BIOGRAPHIES OF AUTHORS



Rosyihan Muhtadlor    recently graduated from the Department of Electrical Engineering and Informatics, majoring in Software Engineering Technology at the Vocational College, Universitas Gadjah Mada, Indonesia. He has a strong interest in information technology, with a focus on web development and artificial intelligence. During his time in college, he actively participated in various internship programs and worked on numerous projects, gaining valuable experience in the IT industry. His research interests include machine learning and web development. He can be contacted at email: rosihann14@mail.ugm.ac.id.



Nur Rohman Rosyid    has been an assistant professor in the Department of Electrical Engineering and Informatics at the Vocational College, Universitas Gadjah Mada, Indonesia, since 2007. He holds a master's degree in Electrical Engineering from Universitas Gadjah Mada and a Doctor of Engineering (Dr. Eng.) degree in Electrical Engineering from King Mongkut's Institute of Technology Ladkrabang, Thailand. His research areas include network security, virtualization, network analysis, and artificial intelligence. He can be contacted at email: nrohmanr@ugm.ac.id.



Anni Karimatul Fauziyyah    has been a lecturer in the Department of Electrical Engineering and Informatics at the Vocational College, Universitas Gadjah Mada, Indonesia, since 2021. She holds a master's degree in Information Technology from Universitas Gadjah Mada. Her research interests include network and distributed systems as well as data engineering. She can be contacted at email: anni.karimatul.f@mail.ugm.ac.id.



Lalu Hendra Permana Setiawan    has been serving as the operation manager at PT Parama Data Unit since 2012. He earned his Bachelor of Engineering degree in Geological Engineering from Universitas Pembangunan Nasional Veteran, Yogyakarta in 2005. His research and expertise are centered on mudlogging, formation evaluation, data management, and the development of drilling software. He can be contacted at email: lalu.hendra@pduid.com.



Irfan Saputra    has been serving as an IT Developer at PT Parama Data Unit (PDU) since 2014. He earned his bachelor's degree in Computer Science from STMIK MH Thamrin in 2015. His areas of expertise include software engineering and data analysis. He can be contacted at email: irfan@pduid.com.



Pavel Stasa    is an assistant professor at VSB–Technical University of Ostrava at Department of Economics and Control Systems. He finished his doctoral studies at 2012 and he focuses his attention to modelling and simulation of flow using fluent SW. During his doctoral studies, he spent one semester at Dongguk University in Seoul under the supervising of Prof. Jongtae Rhee, the study was focused on RFID technology. His research area includes modelling and simulation of flow using CFD program fluent, RFID, NFC, IoT, automation of technological processes. Since 2015 he has been the head of RFID laboratory at VSB-TUO. He can be contacted at email: pavel.stasa@vsb.cz.



Filip Benes    received a Ph.D. in 2012; his thesis was focused on RFID in retail. He has 11 years experience with research work on projects at the ILAB RFID. Experience in the fields of AutoID, augmented and virtual reality and IoT. His abilities include, above all, system analysis of a problem, testing, analysis and design of HW and SW equipment, research of virtual and augmented reality potential in perspective fields. He can be contacted at email: filip.benes@vsb.cz.



Muhammad Syafrudin    received a bachelor's degree from UIN Sunan Kalijaga in Yogyakarta, Indonesia, in 2013, and a Ph.D. degree from Dongguk University in Seoul, South Korea, in 2019. He has been an assistant professor at Sejong University in Seoul since March 2022. Previously, he worked as an assistant professor at Dongguk University from 2019 to 2022. He has published numerous research articles in several international peer-reviewed journals, including expert systems with applications, IEEE Access, food control, sensors, and mathematics. His research interests include industrial artificial intelligence (IAI), industrial analytics (IA), industrial informatics (II), and the industrial internet of things (IIoT), encompassing theory, design, and implementation. He has served as a guest editor and reviewer for journals indexed in Scopus and the Science Citation Index (SCI). He can be contacted at email: udin@sejong.ac.kr.



Ganjar Alfian    has been serving as an assistant professor at the Department of Electrical Engineering and Informatics, Vocational College, Universitas Gadjah Mada, Indonesia, since 2022. Before this, he spent five years as an assistant professor at Dongguk University in Seoul, Republic of Korea. In July 2017, he was a short-term visiting researcher at VSB-Technical University of Ostrava in the Czech Republic. He received Dr. Eng. degrees from the Department of Industrial and Systems Engineering, Dongguk University, Seoul, South Korea, in 2016. His research interests include artificial intelligence, RFID, IoT, machine learning, carsharing, and simulation. He is listed in the World's Top 2% Scientist 2022, 2024 and 2025 by Stanford University. He can be contacted at email: ganjar.alfian@ugm.ac.id.