PROCEEDINGS, 49th Workshop on Geothermal Reservoir Engineering
Stanford University, Stanford, California, February 12-14, 2024
SGP-TR-227

Identification of Undesirable Events in Geothermal Fluid/Steam Production using Machine Learning

Orkhan Khankishiyev, Saeed Salehi, Hamidreza Karami, Vagif Mammadzada

Well Construction Technology Center, The University of Oklahoma

1106 Lexington Avenue, Norman, OK 73069-8427

orkhan@ou.edu

Keywords: geothermal, fluid/steam production, undesirable events, flow instability, scaling, machine learning, data analysis

ABSTRACT

Geothermal energy production confronts persistent challenges tied to undesirable events occurring during geothermal fluid/steam production, including flow instability, scaling and corrosion issues. Conventional monitoring and preventive methodologies have proven insufficient, necessitating innovative approaches for the prediction and identification of these undesirable events. This paper introduces a novel application of machine learning techniques to address this issue. Specifically, the utilization of supervised classification methods, such as K-Nearest Neighbor (KNN), Random Forest, Artificial Neural Networks (ANNs), and Support Vector Machines (SVMs), is proposed for the identification and prediction of undesirable events in geothermal fluid/steam production processes.

To demonstrate the effectiveness of this approach, a substantial and comprehensive dataset, namely the 3W Petrobras Oil Production Dataset, was leveraged. This dataset encompasses an extensive range of operational data with corresponding undesirable events that occurred, enabling the training and evaluation of machine learning models across diverse real-world scenarios. The results obtained underscore the significant potential of machine learning in the identification and prediction of undesirable events. The core of the analysis centered on the utilization of temperature and pressure data recorded by sensors located both at the bottom of the well and at the wellhead. By employing these vital variables as model inputs, remarkable performance in terms of F1 scores, precision, and recall was achieved. These results highlight the importance of advanced data analytics techniques in geothermal energy production for the identification and prediction of undesirable events.

This research contributes to the field of geothermal energy by introducing a data-driven approach for the identification and prediction of undesirable events during geothermal fluid/steam production. The successful application of machine learning algorithms, as demonstrated through the case study employing the 3W Petrobras Dataset, represents a significant advancement in ensuring the sustainability and efficiency of geothermal energy production processes. Furthermore, the insights gained from this study lay the foundation for proactive measures to identify and predict these undesirable events, ultimately enhancing the reliability and economic viability of geothermal power generation.

1. INTRODUCTION

Geothermal energy stands as a promising renewable resource, offering a sustainable alternative to traditional energy sources . The efficient and uninterrupted extraction of geothermal fluids and steam is paramount for ensuring the viability and productivity of geothermal power plants (Tester et al., 2006). However, the occurrence of undesirable events during geothermal fluid and steam production, such as corrosion, scaling, and reservoir decline, poses significant challenges to operational stability and overall efficiency. Predicting and preventing these events is crucial for maintaining the longevity and effectiveness of geothermal energy systems. In this context, the integration of machine learning techniques presents a promising avenue to forecast and prevent such occurrences. This paper introduces a comprehensive analysis leveraging the 3W Petrobras Oil Production Dataset (Vargas et al., 2019), aiming to utilize machine learning models to predict and mitigate these undesirable events, thereby contributing to the stability and efficiency of geothermal energy production.

Anticipating and preventing such events is crucial for the longevity and optimal functioning of geothermal systems. Traditionally, these issues have been mitigated through reactive approaches, where monitoring and corrective measures were implemented after the occurrence of the events. However, this approach is not only costly but also may lead to downtimes and inefficiencies in the energy production process. Advancements in data analytics, particularly in the field of machine learning, offer a proactive and potentially more effective solution (Marins et al., 2021). By analyzing historical data from geothermal production, these techniques could forecast potential issues before they occur, allowing for preventive measures to be taken, thus maintaining the operational stability and efficiency of geothermal power plants.

The integration of machine learning into the geothermal energy domain holds substantial promise (Okoroafor et al., 2022). Leveraging the vast amount of data collected from various sensors, production logs, and operational parameters during geothermal fluid and steam extraction, machine learning techniques can decipher patterns, detect anomalies, and predict potential undesirable events. By employing predictive models on historical data, these techniques offer the potential to enhance operational efficiency, reduce maintenance costs, and prolong the productive life of geothermal systems. The utilization of machine learning for geothermal energy represents a paradigm shift from reactive approaches to a more proactive and predictive framework, facilitating the early identification and mitigation of issues that might affect the stability and efficiency of geothermal power production.

The paper unfolds into several sections to encapsulate the methodology and findings of this study. A comprehensive review of existing literature on geothermal energy production, emphasizing undesirable events and the significance of machine learning in this domain is provided in the next section. This review provides a foundation for understanding the context and significance of the study in the broader spectrum of geothermal energy research. The subsequent sections delve into the methodology employed, outlining the data collection, preprocessing steps, feature selection, and engineering strategies, along with a detailed exploration of various machine learning models employed.

The analysis and results section presents a detailed examination of the dataset, offering a descriptive analysis of its components, and subsequently, an evaluation of the performance of the machine learning models employed. The efficacy of these models in predicting and preempting undesirable events is highlighted, providing insights into the potential implications for the geothermal energy sector. Additionally, the discussion section delves deeper into the interpretation of results, exploring the practical implications for geothermal fluid and steam production, while also addressing the limitations of the study and outlining potential avenues for future research.

2. UNDESIRABLE EVENTS IN GEOTHERMAL PRODUCTION

For over a century, geothermal energy has played a crucial role in sustainable power generation. However, the development of geothermal resources has not been without its substantial cost (Khankishiyev et al., 2023) and challenges (Vivas & Salehi, 2021). Operational difficulties manifest in various forms, extending beyond technical complexities to encompass political, cultural, and environmental concerns. The chemistry of geothermal fluids, which can occasionally contain significant concentrations of minerals and gases, poses a considerable risk to the integrity of wells and surface installations through which the geothermal fluids flow. This is primarily due to the potential for scaling and corrosion, emerging as the most frequent technical issues in geothermal utilization (Khankishiyev & Salehi, 2023). Addressing these challenges is imperative for the successful and sustainable utilization of geothermal energy.

Geothermal wells can face various production problems that may affect their efficiency and output. Some common issues include:

Scaling and Mineral Deposition: Minerals present in geothermal fluids can precipitate and form scales on the wellbore and production equipment, reducing the flow of geothermal fluids. Geothermal fluids often carry dissolved minerals that can precipitate and deposit on wellbore surfaces and within the reservoir, resulting in scaling and reduced permeability (Klapper et al., 2019). Scaling can impede fluid flow, decrease heat transfer efficiency, and lead to equipment failure (Stahl et al., 2000).

Corrosion: Well casing and production equipment made from steel corrodes due to the chemical nature of geothermal fluids. Super-hot geothermal environments often involve exposure to corrosive and abrasive fluids. The presence of aggressive chemicals and high-velocity fluid flows can lead to corrosion and erosion of wellbore materials, including casing, drill bits, and downhole equipment (Karlsdóttir et al., 2019). Previously, a lot of corrosion and scaling related problems in geothermal wells have been reported in several literature. (Ocampo et al., 2005; Zhao et al., 2023)

Casing and Tubing Erosion: The high-velocity flow of geothermal fluids can cause erosion of the wellbore, leading to reduced well integrity and increased maintenance needs. Erosion-corrosion occurs during the transport of two-phase flow at high velocities. This phenomenon happens when abrasive particles, moving at an angle to the substrate surface, result in wear. The extent of erosion-corrosion tends to increase with higher flow rates and presence of solid particles in the fluid (Nogara & Zarrouk, 2018). Severe erosion corrosion may occur if the protective coating, primarily composed of corrosion products, is incapable of continuous and rapid reformation. According to Kurata et al. (1992), erosion-corrosion accounts for 25.4% of reported damages in geothermal power systems in Japan. Table 1 provides a summary of the corrosion, cavitation and erosion findings from KJ-39, IDDP-1, and IDDP-2 wells in Iceland, highlighting the severity of the corrosion problem in geothermal environments.

Wellbore Leaks: Leaks in the wellbore, due to severe casing corrosion/erosion or faulty connections, can result in the escape of geothermal fluids into surrounding formations. Reduction of casing wall thickness from severe corrosion may lead to casing buckling and leaks. Such leaks, often caused by corrosion, poor welding, inadequate cementing, thermal cycling, wear from drill pipes, or erosion, can have severe and costly impacts (Phi et al., 2019). In the upper well section, a leak could lead to the production of steam and fluid into the annulus between casing strings. If not promptly addressed, a leakage path to the surface could, in extreme cases, cause a dangerous steam eruption, throwing mud and rocks and leaving a large crater (Kalvenes, 2017).

Reservoir Decline: Over time, geothermal reservoirs may undergo a decline in both temperature and pressure, resulting in diminished energy production. This decline in reservoir conditions poses a significant challenge to sustained geothermal power generation. As the reservoir's thermal energy decreases, the efficiency of energy extraction diminishes, leading to lower overall power output (Zais & Bodvarsson, 1980). Additionally, decreasing pressure levels contribute to reduced fluid flow, further impacting the system's performance. Effective reservoir management becomes crucial to counteracting this natural decline. Reinjection of produced fluids is the main method to maintain reservoir pressure and temperature can be employed to mitigate the effects of reservoir decline (Beckers et al., 2017).

To identify and address these issues effectively, a combination of proper design, regular maintenance, and advanced monitoring techniques is essential in the operation of geothermal wells. The majority of the undesirable events listed above impact the temperature, pressure, and flow rate of the steam/fluid during production from the well. Identifying those issues using physical models is incredibly challenging due to the complexity of the model and the number of physical factors that need to be considered. However, when the pressure and temperature at the bottomhole and at the wellhead, and flow rate at the wellhead is recorded and labeled accordingly, the machine learning algorithms can be utilized to train the model and predict the issues, even during the transition stage.

Table 1. Summary of corrosion findings from KJ-39, IDDP-1 and IDDP-2 wells in Iceland

Well Name / Location	IDDP-2, Reykjanes, Iceland	IDDP-1, Krafla, Iceland	KJ-39, Krafla, Iceland
MD, m/ft	4650 / 15256	2102 / 6896	2800 / 9186
Temp., °C/ °F	427 / 800	450 / 858	350 / 662
Casing / Tubing Size, in	Casing: 9 5/8" Perforated liner: 7" Tubing: 3.5 in	Casing: 9 5/8" Slotted liner: 9 5/8"	Casing: 9 5/8" Slotted liner: 7 5/8"
Casing/Tubing Material	Casing and perforated liner: Carbon steel, L80, BTC Tubing: Carbon steel, API 5DP PSL1 grade G-105	Casing: Carbon steel, K55, Hydril 563 Liner: Carbon steel, K55, BTC	Casing: Carbon steel, API 5L K55 Liner: Carbon steel, API 5L K55
Failure mode	Axial cracks in tool joints distributed evenly on the circumferential of the joint boxes, uniform, and pitting corrosion in pipe bodies	Erosion caused by SiO2 precipitation, corrosion of production liner, nozzles in wellhead, hydrogen embrittlement of API K55 casing material. Master valve failure due to corrosion, API T95 were less affected by sulfide corrosion and hydrogen embrittlement.	Cavitation damage and corrosion inside and the outside surfaces of the pipes that started only a few weeks after the opening of the fluid flow from the well
Damage location	A hole was created on casing because of corrosion at 2300 m (7545.932 ft), corrosion damage is observed in injection string from 4000 m (13123.36 ft) to 4659 m (15285.43 ft) that was inside a perforated liner	Slotted liner, production casing, wellhead equipment	Wellbore was blocked at 1600 m (5249.34 ft) due to parts of the carbon steel liner corroding and breaking in the well. The liner was broken in half at 1600 m (5249.34 ft) when trying to retrieve it
Damage cause	Sulfide stress corrosion cracking, thermal stresses, sulfide corrosion	Hydrochloric acid and sulfide stress corrosion cracking, pitting corrosion, horizontal cracks, and fissures parallel to the surface, hydrogen embrittlement, aggressive silicate precipitation, thermal stresses	Uniform and pitting corrosion, hydrochloric acid and sulfide stress corrosion, cavitation corrosion, hydrogen embrittlement and cracking, thermal stresses
Chemical composition	3.0E-07 mg/L H2S and 23.7 mg/L CO2 gas in formation steam, 12 ppm O2 in injection fluid	732 ppm CO2, 339 ppm H2S, 93 ppm HCl, 10 ppm H2 gas in formation steam	4085 ppm CO2, 560 ppm H2S, 330 ppm HCL, 75 ppm N2, 60 ppm H2 gas in formation steam
References	(Friðleifsson et al., 2017), (Karlsdóttir et al., 2019)	(Friðleifsson et al., 2015), (Markússon & Hauksson, 2015), (Hauksson et al., 2014)	(Karlsdóttir & Thorbjornsson, 2012)

3. MACHINE LEARNING IN DETECTION OF UNDESIRABLE EVENTS

Machine learning (ML), a component of artificial intelligence (AI), centers on discerning patterns from historical data to forecast outcomes in novel datasets. In contrast to the broader scope of AI, which includes reasoning, planning, and perception, ML is geared explicitly toward predictive tasks. By leveraging past information, ML excels at anticipating future trends and results, making it an invaluable asset in disciplines like petroleum engineering. Different ML models have been developed and applied in the energy industry (Hu et al., 2023).

Supervised learning is the utilization of machine learning algorithms for addressing problems where a known dependent variable is involved. The detection of flow instability is a classification problem because the dependent variable (undesirable event) is categorical, the class label.

3.1. Machine learning workflow

The machine learning workflow is a systematic process that involves several key stages, from data preparation to model deployment. Once the database is initialized, the first step is pre-processing. It is the procedure of shaping the data set to meet specific needs for further applications. Pre-processing consists of several steps. In the first pre-processing step, data wrangling involves cleaning, structuring data, addressing missing values, and modifying or deleting variables. The second step is transforming the prepared data into a format suitable for machine learning algorithms. Lastly, transforming feature scales ensures the dataset's uniformity by translating each feature's mean to zero and adjusting its variance to the unit scale. Here is an overview of the typical machine learning workflow:

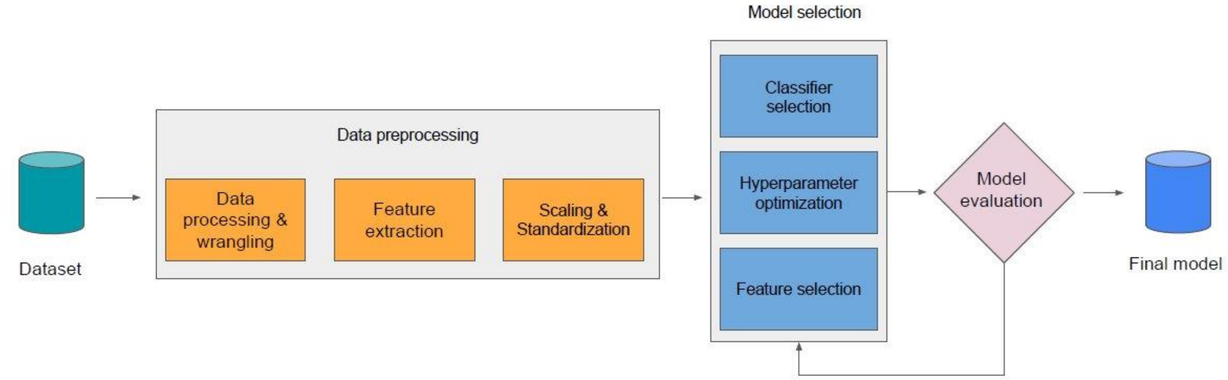

Figure 1: Machine Learning Workflow (Carvalho, 2021)

3.2. Machine Learning model selection criteria

The initial phase of model selection involves creating a roster of classification algorithms. This process kicks off by opting for more straightforward options like Decision Tree (DT) and Random Forest (RF). Importantly, these algorithms necessitate minimal hyperparameter tuning and boast scalability. Transitioning beyond simplicity, Adaptive Boosting (AdaBoost) and Support Vector Machines (SVM) present distinct strengths, notably showing reduced vulnerability to overfitting.

In the realm of non-linear challenges, Extreme Learning Machine (ELM) and Multilayer Perceptron (MLP) neural networks emerge as suitable solutions. Additionally, the k-Nearest Neighbors, a non-parametric algorithm with sparse hyperparameters, consistently demonstrates robust performance. Adding to the array of classifiers, the Random Forest algorithm stands out for its adaptability and stability, contributing to a comprehensive and diversified model selection process. Hyperparameters are internal settings of an algorithm unrelated to training data, while parameters are learned values from data used for predictions. The optimization or tuning of hyperparameters is the process of identifying the optimal attributes for each algorithm to achieve maximum classification performance. Grid search is a commonly employed method for this purpose, systematically evaluating the model across all points in a chosen subspace (the grid) to pinpoint the best configuration for a model. Another critical phase in model selection is feature selection, a process proven to enhance classifier performance on a subset of the original features. This step not only improves the algorithm's speed when working with trained models but also addresses challenges associated with feature extraction. While feature extraction often increases problem dimensionality and leads to sparser data, making it more challenging to determine the optimal model parameters, feature selection helps mitigate these side effects.

Feature selection encompasses three main algorithmic groups: i) filter (ranking), ii) wrapper, and iii) embedded (Kumar & Minz, 2014). In the filter approach, the exploration of data occurs independently of the classifier, often involving the analysis of feature variances. Wrapper methods, on the other hand, leverage the relationship between data and classifier, employing the classifier to identify the most suitable features. Finally, embedded methods conduct feature selection during the training phase, seamlessly integrating the process with the model development. Hyperparameters, distinct from parameters derived from training data, are internal algorithm settings. Parameters, learned from data, guide predictions. Hyperparameter optimization, crucial for achieving optimal classification performance, involves fine-tuning algorithm attributes. Grid search, a widely employed method, systematically evaluates the model across various points in a selected subspace (grid), aiding in the identification of the optimal model fit (Lerman, 1980).

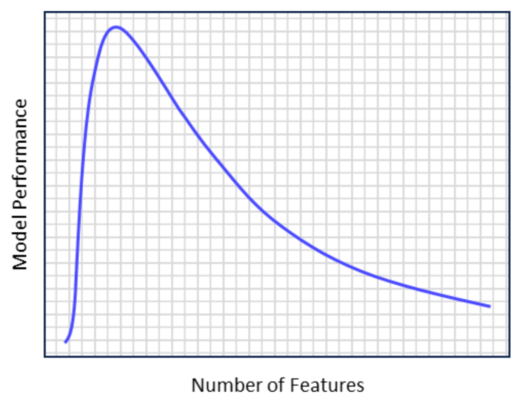

Figure 2. Hughes' phenomenon - curse of dimensionality (Debie & Shafi, 2019)

Within the broader framework of feature engineering, both feature extraction and feature selection face the challenge of the curse of dimensionality, often referred to as Hughes' phenomenon. This phenomenon posits that an algorithm's performance improves with an increasing number of features up to a certain threshold, beyond which it starts to deteriorate (Figure 2). To navigate this issue, sequential feature selection (SFS) acts as a wrapper algorithm, dynamically adjusting the selected feature group by either adding (forward) or removing (backward) one feature at a time until there is no discernible improvement in overall performance. Consequently, the incorporation of feature selection is not only essential for optimizing the model but also becomes a necessity to effectively address Hughes' phenomenon.

3.3. Performance evaluation

One of the crucial tasks in machine learning is model evaluation, which compares models and helps practitioners make decisions by calculating performance measures. Fundamentally, the assessment method aims to gauge a model's performance on unobserved data by evaluating its generalization ability to minimize error and apply previously learned knowledge to new observations. In addition, choosing the right metric for a specific situation is only one aspect of proper result interpretation; another is the methodology employed in calculating the metric. 80% of the cleaned dataset is usually used for model selection, with the remaining 20% being set aside for testing or model evaluation. By assessing the model's performance using never-before-seen data, this division guarantees a thorough review of the model's efficacy in real-world scenarios. It offers valuable insights for decision-making in the machine learning workflow.

4. A CASE STUDY USING 3W PETROBRAS OIL PRODUCTION DATASET

Due to the scarcity of production data with labeled undesirable events from the geothermal fields, an open-source oil production dataset was used to demonstrate the applicability and efficiency of the proposed workflow. It is noteworthy that machine learning algorithms are often agnostic to the specific domain of production (oil or geothermal) if the underlying patterns and characteristics are captured by the data. Therefore, the choice of an oil production dataset for demonstration purposes does not compromise the generalizability of the proposed workflow to geothermal production scenarios, as the algorithms focus on learning patterns and relationships within the data rather than the specific production domain.

4.1. 3W Petrobras Dataset

The open-source 3W dataset, gathered and published by Petrobras[1], aims to optimize the identification of undesirable events in offshore well production. Its primary goal is to enhance the efficiency of monitoring well and subsea system integrity, crucial for preventing substantial losses to people, the environment, and the company's reputation. The detailed description of the dataset is published by Vargas et al. (2019). The dataset comprises three distinct types of instances categorized based on their sources: real, simulated, and hand-drawn. Real instances correspond to events that occurred in Petrobras' current wells during oil production. The inclusion of simulated and hand-drawn instances serves a fundamental purpose—to mitigate the initial imbalance in the dataset, which was initially dominated by real instances, a common characteristic in industrial data. This diversification enhances the dataset's representativeness and contributes to more robust machine learning model training by incorporating a broader range of scenarios. The dataset comprises following variables collected (or simulated) from temperature and pressure sensors located in downhole, wellhead, and the production platform (See Figure 3 below for sensor placement):

- P-PDG: pressure variable at the Permanent Downhole Gauge (PDG);
- P-TPT: pressure variable at the Temperature and Pressure Transducer (TPT);
- T-TPT: temperature variable at the Temperature and Pressure Transducer (TPT);
- P-MON-PCK: pressure variable upstream of the production choke (PCK);
- T-JUS-PCK: temperature variable downstream of the production choke (PCK);
- P-JUS-CKGL: pressure variable upstream of the gas lift choke (CKGL);
- T-JUS-CKGL: temperature variable upstream of the gas lift choke (CKGL);
- QGL: gas lift flow rate;
- Class: undesirable event observations labels

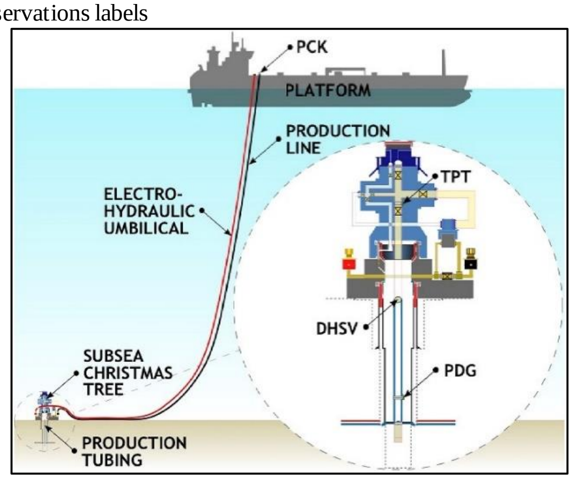

Figure 3. Simplified schematic of a typical offshore naturally flowing well (Vargas et al., 2019)

[1] https://github.com/petrobras/3W

The types of undesirable events selected to be predicted with Machine Learning algorithms, taken from 3W dataset, are outlined below. It is crucial to note that there may not always be a unanimous consensus on the terminology and interpretation of these undesirable events, even among experts. To validate the actual occurrences of each identified type of undesirable event, Petrobras well monitoring professionals typically analyze time windows of varying sizes. The estimated durations for these time windows are detailed in Table 2.

Table 2. Estimates of time window sizes needed to confirm occurrences of undesirable events (Vargas et al., 2019)

TYPE OF UNDESIRABLE EVENT	TIME WINDOW TO CONFIRM THE EVENT
Label 1 – Abrupt increase of Basic Sediment and Water (BSW)	12 h
Label 2 – Spurious closure of the Downhole Safety Valve (DHSV)	5 min–20 min
Label 3 – Severe slugging	5 h
Label 4 – Flow instability	15 min
Label 5 – Rapid productivity loss	12 h
Label 6 – Quick restriction in Production Choke (PCK)	15 min
Label 7 – Scaling in Production Choke (PCK)	72 h
Label 8 – Hydrate formation in production line	30 min–5 h

Early identification of these events, either before issues occur or during transient stages would save a lot of production downtime and is essential for improving the efficiency of monitoring well and subsea system integrity.

4.2 Data pre-processing

The number of real, simulated and hand drawn instances is shown in Table 3 below. Each instance is recorded as a time series data, resulting in millions of observations. Fortunately, most of those datapoints can be used to train and test the machine learning algorithm. However, before feeding the data to the algorithm, missing values and outliers must be handled. The variables related to the gas lift are removed from the dataset since only a small number of the wells had the data. Next, any raw containing missing values is removed on a row-wise basis. While various techniques exist for handling missing data, such as propagating the last known number forward or utilizing the mean from the last known values, these methods may introduce additional noise to the signal. The decision to drop rows with missing values resulted in the removal of approximately 14.8% (1,472,177 observations) from the normal operation dataset. The identical event classes for steady state and transient states were combined. Since the number of merged clean datapoints is big, a subset of 50,000 data points, stratified by classes, was randomly selected from the total pool of 50,000,000 data points to save processing time, and avoid computing crashes. Additional filtering involved removing rows with P-TPT values exceeding 40,000,000 Pa (5800 psi) and T-JUS-CKP values above 150°C (302°F). The final number of observations for each undesirable event class is shown in Table 3 below. The matrix scatterplot in Figure 4 demonstrates the classification of labels 0 to 8 where the simulated data for class 8 stands out among others.

Table 3. The number of instances, datapoints and randomly selected folds for each class of undesirable event

<table>
<tr><th colspan="2" rowspan="2">Class</th><th colspan="3">Number of Instances</th><th rowspan="2">Number of all points after missing value removal</th><th rowspan="2">Merged</th><th rowspan="2">Randomly Selected Fold</th></tr>
<tr><th>Real</th><th>Simulated</th><th>Hand Drawn</th></tr>
<tr><td>Label 0</td><td>Normal Operation</td><td>597</td><td>-</td><td>-</td><td>9822473</td><td>9822473</td><td>11604</td></tr>
<tr><td rowspan="2">Label 1</td><td>Steady</td><td rowspan="2">5</td><td rowspan="2">114</td><td rowspan="2">10</td><td>2909702</td><td rowspan="2">8196087</td><td rowspan="2">9605</td></tr>
<tr><td>Transient</td><td>5286385</td></tr>
<tr><td rowspan="2">Label 2</td><td>Steady</td><td rowspan="2">22</td><td rowspan="2">16</td><td rowspan="2">-</td><td>348621</td><td rowspan="2">415528</td><td rowspan="2">493</td></tr>
<tr><td>Transient</td><td>66907</td></tr>
<tr><td>Label 3</td><td>Steady</td><td>32</td><td>74</td><td>-</td><td>4833360</td><td>4833360</td><td>5683</td></tr>
<tr><td>Label 4</td><td>Steady</td><td rowspan="2">344</td><td rowspan="2">-</td><td rowspan="2">-</td><td>2460270</td><td>2460270</td><td>2865</td></tr>
<tr><td rowspan="2">Label 5</td><td>Steady</td><td>10552143</td><td rowspan="2">12972403</td><td rowspan="2">15309</td></tr>
<tr><td>Transient</td><td>12</td><td>439</td><td>-</td><td>2420260</td></tr>
<tr><td rowspan="2">Label 6</td><td>Steady</td><td rowspan="2">6</td><td rowspan="2">215</td><td rowspan="2">-</td><td>12951</td><td rowspan="2">19203</td><td rowspan="2">23</td></tr>
<tr><td>Transient</td><td>6252</td></tr>
<tr><td rowspan="2">Label 7</td><td>Steady</td><td rowspan="2">4</td><td rowspan="2">-</td><td rowspan="2">10</td><td>110289</td><td rowspan="2">2233418</td><td rowspan="2">2602</td></tr>
<tr><td>Transient</td><td>2123129</td></tr>
<tr><td rowspan="2">Label 8</td><td>Steady</td><td rowspan="2">3</td><td rowspan="2">81</td><td rowspan="2"></td><td>603141</td><td rowspan="2">2041120</td><td rowspan="2">2430</td></tr>
<tr><td>Transient</td><td>1437979</td></tr>
</table>

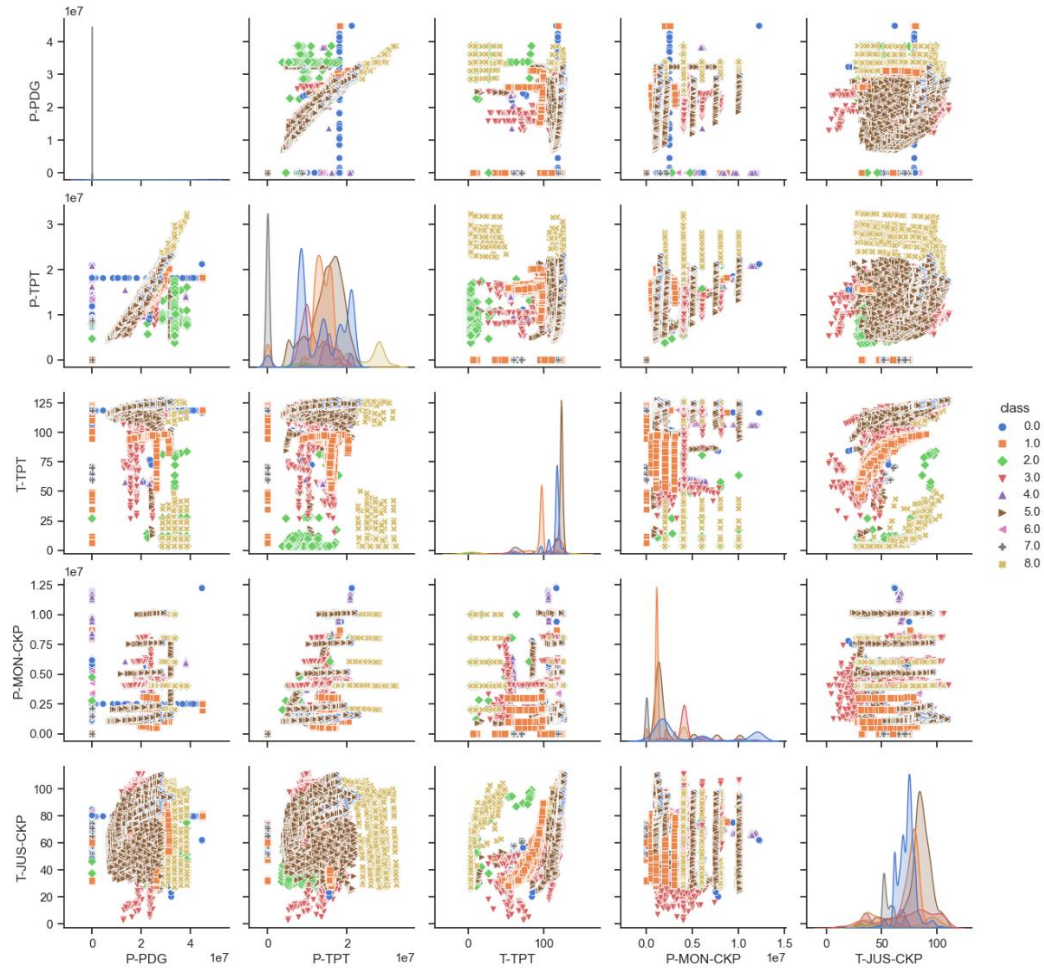

Figure 4. Matrix Scatterplot to visualize the classification of labels 0 to 8. Simulated data (class 8) stands out among others.

4.3. Algorithms for classification

The K-Nearest Neighbor (KNN), Random Forest, Artificial Neural Network (ANN), and Support Vector Machine (SVM) supervised classification models were exported from an open-source scikit-learn library[2]. The train and test split ratio were set to 80%/20% of the randomly selected fold for all models. Accuracy, Recall and F-1 scores were used to compare the model results and Hyperparameter optimization and Cross-validation were performed for all models to improve the performance. The proportion of event classes in randomly selected stratified dataset is shown in Figure 5. Class 6 and Class 2 have the lowest number of datapoints, thus lower classification performance is expected for them.

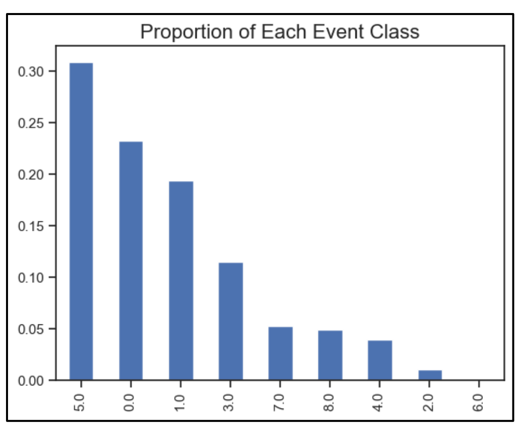

Figure 5. Proportion of event classes in randomly selected stratified dataset

[2] https://scikit-learn.org/stable/

4.3.1 K-Nearest Neighbor

K-Nearest Neighbors (KNN) is a non-parametric and versatile algorithm used for both classification and regression tasks in machine learning. It makes predictions by assigning the majority class or mean value of the k-nearest data points to the query point, where k is a user-defined parameter. K values from 5 to 500 were tested in Grid Search and best performance was observed when k was 50 without overfitting. The precision and recall values versus increasing K values were plotted for each class in Figure 6 (left and middle) below. As shown in the confusion matrix in Figure 6 (right), predicted labels match the True labels for all classes, except Class 6 (Quick restriction in Production Choke) due to a very low number of datapoints (0.46% of the dataset) used to train the model. F1 Scores obtained over 10-Fold Cross-validation using KNN Classification with an optimal set of parameters were consistent.

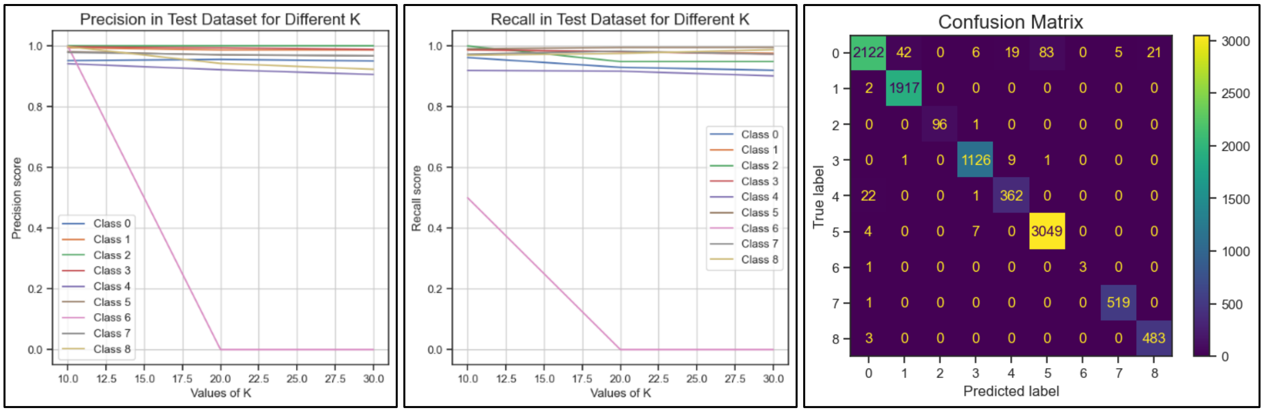

Figure 6. Precision (left), Recall (middle) and Confusion Matrix (right) for KNN Classification

4.3.2 Random Forest

Random Forest is an ensemble learning algorithm that builds multiple decision trees and merges their predictions to enhance accuracy and reduce overfitting. By introducing randomness in both data sampling and feature selection during the tree-building process, Random Forest creates a robust and versatile model for classification and regression tasks. The Figure 7 (left) showed that T-JUS-CKP: temperature variable upstream of the gas lift choke (CKGL) is the least important variable in classification process, while T-TPT: temperature variable at the wellhead and P-PDG: pressure variable at the bottomhole are the most important variables. Maximum depth of 3 to 10 (a higher depth allows the trees to make more complex splits, capturing intricate patterns in the training data) and number of estimators of 100 to 500 (with higher number of trees, the model tends to become more robust and stable, reducing overfitting and improving the overall performance) was used in cross validation to optimize these parameters (Figure 7 (right)). Although seemed like the results got better, it was observed that the higher maximum depth resulted in overfitting and the higher number of estimators consumed significantly more computational power. Thus, maximum depth was set to 8 and number of estimators was set to 100.

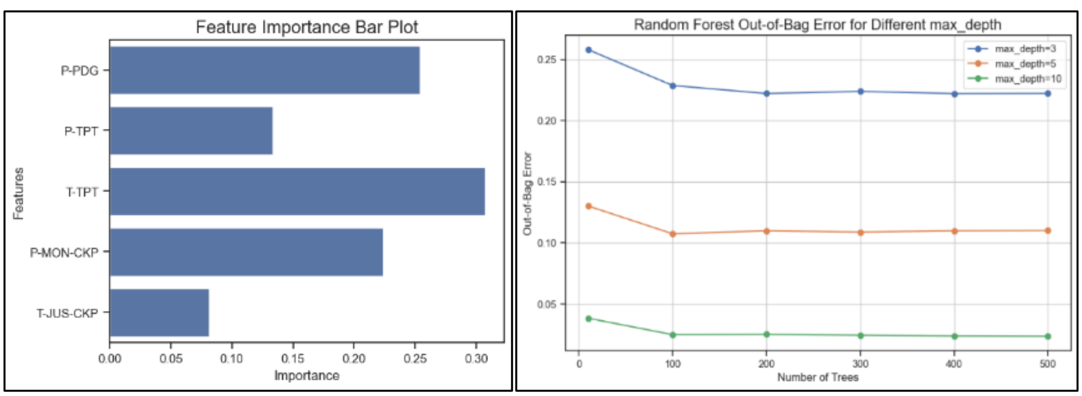

Figure 7. Feature Importance Bar Plot (left), Out-of-Bag Error for different maximum depth and number of trees (right)

	precision	recall	f1-score	support
0.0	0.98	0.85	0.91	2298
1.0	0.93	0.99	0.96	1919
2.0	1.00	0.98	0.99	97
3.0	0.94	0.99	0.96	1137
4.0	0.92	0.83	0.87	385
5.0	0.97	0.99	0.98	3060
6.0	1.00	0.50	0.67	4
7.0	0.96	1.00	0.98	520
8.0	0.93	1.00	0.96	486
accuracy			0.95	9906
macro avg	0.96	0.90	0.92	9906
weighted avg	0.96	0.95	0.95	9906

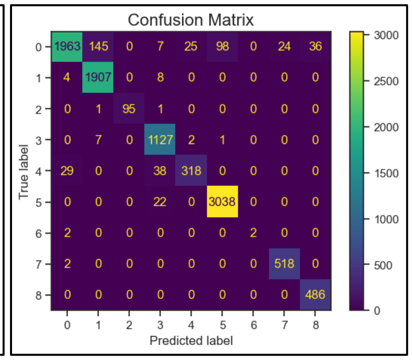

Figure 8. Precision, recall and f1-score for all classes (left) and Confusion Matrix (right) for RF Classification

The Random Forest model yielded a similar F-1 score across all classes except class 6 with overall 95% accuracy and according to the confusion matrix in Figure 8 (right), the model predicted the undesirable events very well. The highest confusion occurred when predicting the Class 0 - normal operation mainly as Class 1 or Class 5.

4.3.3 Artificial Neural Networks

Artificial Neural Networks (ANN) are computational models inspired by the structure and function of the human brain. Comprising interconnected nodes organized in layers, ANNs are designed for tasks such as pattern recognition, classification, and regression. Through a process of training and learning from labeled data, neural networks adapt their internal weights to make accurate predictions and uncover complex patterns in diverse datasets. Hyperparameter optimization was carried out for the number of hidden layers, learning rate and epochs. Increasing the size of the hidden layers increased the classification performance but it became much more time-consuming during model running. Meanwhile, the optimized hyperparameters resulted in consistent F-1 score over 10-fold cross-validation and overall 93% accuracy, 93% precision and 86% recall were achieved (Figure 9).

	precision	recall	f1-score	support
0.0	0.91	0.82	0.87	2298
1.0	0.90	0.97	0.93	1919
2.0	1.00	0.99	0.99	97
3.0	0.99	0.98	0.99	1137
4.0	0.82	0.73	0.77	385
5.0	0.96	1.00	0.98	3060
6.0	1.00	0.50	0.67	4
7.0	0.82	0.87	0.84	520
8.0	0.94	0.91	0.93	486
accuracy			0.93	9906
macro avg	0.93	0.86	0.88	9906
weighted avg	0.93	0.93	0.93	9906

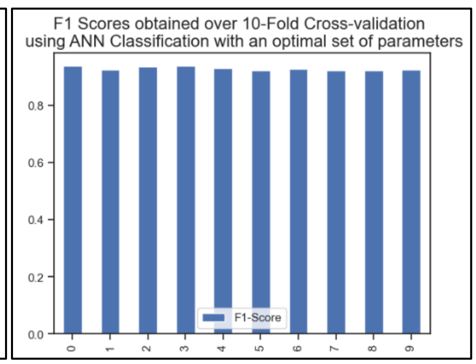

Figure 9. Precision, recall and f1-score for all classes (left) and F1 Scores obtained over 10-Fold Cross-validation (right)

4.3.4 Support Vector Machine

Support Vector Machines (SVM) works by finding the optimal hyperplane that maximally separates different classes in the feature space. It is particularly effective in high-dimensional spaces and can handle both linear and non-linear relationships through kernel functions. SVM algorithm was originally developed for classification problems of two classes. Later, it was extended for regression and multi-class classification.

Cross-validation was used to find the optimal values for kernel and gamma hyperparameters only, because optimizing all other hyperparameters required huge run time. However, consistent F-1 scores were obtained over 10-fold cross-validation. The 92% accuracy and 86% recall values provided by the SVM are the smallest in comparison with previous models and it was not able to predict Class 6 at all.

4.3.5 ML Performance Summary

Table 4 below summarizes the performance of four machine learning models used in this study. Overall, KNN had the best performance in terms of precision, recall, f1-score and accuracy. Another advantage of KNN was that it required the least computational power. On the other hand, SVM took the most amount of time to run, had a lot of hyperparameters to optimize and performed with the smallest accuracy compared to three other models used in this study.

Table 4. Summary of the performance of machine learning methods used in the study

Model	Precision	Recall	F-1 Score	Accuracy	Computational Power Requirement
K-Nearest Neighbor	0.97	0.89	0.91	0.97	Low
Random Forest	0.96	0.95	0.95	0.95	Moderate
Artificial Neural Networks	0.93	0.86	0.88	0.93	High
Support Vector Machine	0.81	0.79	0.8	0.92	High

CONCLUSION

In conclusion, this research pioneers a data-driven approach to address challenges in geothermal fluid/steam production, focusing on identifying and predicting undesirable events. Leveraging machine learning techniques, including K-Nearest Neighbor (KNN), Decision Tree (DT), Random Forest (RF), Artificial Neural Networks (ANNs), and Support Vector Machines (SVMs), our study demonstrates the efficacy of these methods in enhancing the reliability and efficiency of geothermal energy production.

The analysis, based on the 3W Petrobras Oil Production Dataset, underscores the superior performance of KNN in precision, recall, F1-score, and accuracy. Notably, KNN exhibits exceptional results while requiring minimal computational power, making it a practical choice for real-world applications. DT and RF also showcase commendable performance, offering a balance between accuracy and computational efficiency. In contrast, SVM exhibits drawbacks, such as prolonged runtime, numerous hyperparameters, and lower accuracy. The study emphasizes optimizing computational resources, with KNN emerging as the optimal choice for efficiency and performance.

Insights gained extend to the practical implementation of machine learning algorithms in the geothermal industry. The successful identification of undesirable events, demonstrated using temperature and pressure data, provides a foundation for proactive measures in geothermal energy production. Continuous monitoring and expertly labeled training datasets are crucial for algorithm effectiveness.

Future research can explore unsupervised classification methods to evaluate the potential of machine learning algorithms in identifying clusters associated with such production challenges. The findings contribute significantly to advancing sustainability and efficiency in geothermal power generation, marking a crucial step towards ensuring the success of geothermal energy production processes.

ACKNOWLEDGEMENT

The authors of this paper would like to thank DeepPower Inc for funding this research. The opinions, findings, conclusions, or recommendations presented in this publication are solely those of the authors and do not necessarily represent the views or opinions of DeepPower Inc.

DATA AND CODE AVAILABILITY

The data used in the study is available online at https://github.com/petrobras/3W and the code developed through this study can be provided upon request by emailing the primary author at orkhan@ou.edu.

REFERENCES

Beckers, K. J., Young, K. R., Johnston, H., & Snyder, D. M. (2017). Analysis of Geothermal Reservoir and Well Operational Conditions using Monthly Production Reports from Nevada and California. *Geothermal Resources Council 41st Annual Meeting*.

Carvalho, B. G. (2021). Evaluating machine learning techniques for detection of flow instability events in offshore oil wells. *Master's degree dissertation*.

Debie, E., & Shafi, K. (2019). Implications of the curse of dimensionality for supervised learning classifier systems: theoretical and empirical analyses. *Pattern Analysis and Applications*, *22*, 519-536.

Friðleifsson, G. Ó., Elders, W. A., Zierenberg, R. A., Stefánsson, A., Fowler, A. P. G., Weisenberger, T. B., Harðarson, B. S., & Mesfin, K. G. (2017). The Iceland Deep Drilling Project 4. km deep well, IDDP-2, in the seawater-recharged Reykjanes geothermal field in SW Iceland has successfully reached its supercritical target. *Sci. Dril.*, *23*, 1-12.

Friðleifsson, G. Ó., Pálsson, B., Albertsson, A., Stefánsson, B., Gunnlaugsson, E., Ketilsson, J., & Gíslason, Þ. (2015). IDDP-1 Drilled Into Magma - World's First Magma-EGS System Created.

Hauksson, T., Markússon, S. H., Einarsson, K., Karlsdóttir, S. N., Einarsson, Á., Moeller, A., & Sigmarsson, Þ. (2014). Pilot testing of handling the fluids from the IDDP-1 exploratory geothermal well, Krafla, N.E. Iceland. *Geothermics*, *49*, 76-82.

Hu, Z., Vivas, C., & Salehi, S. (2023). Computer Vision in Predicting Thermal Diffusivity of Subsurface Rocks. *2023 Geothermal Rising Conference Transactions*, *47*, 2299-2308.

Kalvenes, K. L. (2017). Investigation of Corrosion in Geothermal Wells - A Qualitative Risk Assessment.

Karlsdóttir, S. N., Krogh, B., Sæther, S., & Pórólfsson, G. (2019). Corrosion Damage of Injection String From the Deep Geothermal Well IDDP-2 in Reykjanes Iceland. *Corrosion*.

Karlsdóttir, S. N., & Thorbjornsson, I. O. (2012). Hydrogen Embrittlement And Corrosion In High Temperature Geothermal Well. *Corrosion*.

Khankishiyev, O., & Salehi, S. (2023). Technology Review and Techno-Economic Analysis of Super-Hot EGS Development in Hard Rock Formations. *Geothermal Resources Council Transactions*, *47*, 2979-2997.

Khankishiyev, O., Salehi, S., Vivas, C., Nygaard, R., & Rehg, D. (2023). Techno-Economic Investigation of Geothermal Development in Sedimentary Basins. *ASME Open Journal of Engineering*.

Kumar, V., & Minz, S. (2014). Feature selection: a literature review. *SmartCR*, *4*(3), 211-229.

Kurata, Y., Sanada, N., Nanjo, H., & Ikeuchi, J. (1992). Material damages in geothermal power plants. *Proceedings of the 14th New Zealand geothermal workshop*, *1*, 159-164.

Lerman, P. (1980). Fitting segmented regression models by grid search. *Journal of the Royal Statistical Society Series C: Applied Statistics*, *29*(1), 77-84.

Marins, M. A., Barros, B. D., Santos, I. H., Barrionuevo, D. C., Vargas, R. E., Prego, T. d. M., de Lima, A. A., de Campos, M. L., da Silva, E. A., & Netto, S. L. (2021). Fault detection and classification in oil wells and production/service lines using random forest. *Journal of Petroleum Science and Engineering*, *197*, 107879.

Markússon, S. H., & Hauksson, T. (2015). Utilization of the Hottest Well in the World, IDDP-1 in Krafla.

Nogara, J. B., & Zarrouk, S. J. (2018). Corrosion in geothermal environment: Part 1: Fluids and their impact. *Renewable & Sustainable Energy Reviews*, *82*, 1333-1346.

Ocampo, J. d. D., Valdez-Salaz, B., Shorr, M. J., & Sauceda, I. (2005). Review of Corrosion and Scaling Problems in Cerro Prieto Geothermal Field over 31 Years of Commercial Operations.

Okoroafor, E. R., Smith, C. M., Ochie, K. I., Nwosu, C. J., Gudmundsdottir, H., & Aljubran, M. J. (2022). Machine learning in subsurface geothermal energy: Two decades in review. *Geothermics*, *102*, 102401.

Phi, T. H., Elgaddafi, R., Ramadan, M., Ahmed, R., & Teodoriu, C. (2019). Well Integrity Issues: Extreme High-Pressure High-Temperature Wells and Geothermal Wells a Review.

Stahl, G., Pátzay, G., Weiser, L., & Kálmán, E. (2000). Study of calcite scaling and corrosion processes in geothermal systems. *Geothermics*, *29*, 105-119.

Tester, J. W., Anderson, B. J., Batchelor, A. S., Blackwell, D. D., DiPippo, R., & Drake, E. M. (2006). The Future of Geothermal Energy: Impact of Enhanced Geothermal Systems (EGS) on the United States in the 21st Century.

Vargas, R. E. V., Munaro, C. J., Ciarelli, P. M., Medeiros, A. G., do Amaral, B. G., Barrionuevo, D. C., de Araújo, J. C. D., Ribeiro, J. L., & Magalhães, L. P. (2019). A realistic and public dataset with rare undesirable real events in oil wells. *Journal of Petroleum Science and Engineering*, *181*, 106223.

Vivas, C., & Salehi, S. (2021). Rheological investigation of effect of high temperature on geothermal drilling fluids additives and lost circulation materials. *Geothermics*, *96*, 102219.

Zais, E. J., & Bodvarsson, G. (1980). Analysis of production decline in geothermal reservoirs. *Lawrence Berkeley National Laboratory*.

Zhao, H., Huang, Y., Deng, S., Wang, L., Peng, H., Shen, X., Ling, D., Liu, L., & Liu, Y. (2023). Research progress on scaling mechanism and anti-scaling technology of geothermal well system. *Journal of Dispersion Science and Technology*, *44*(9), 1657-1670. https://doi.org/10.1080/01932691.2022.2033625