Proceedings World Geothermal Congress 2023
Beijing, China, April 17 – 21, 2023

An in-depth Review of Machine Learning Applications in Geothermal Reservoir Engineering.

Jesse Nyokabi[1] , Noah Oyugi[2] , Githae Mbugua[3] , Antony Hiuhu[4] , Kumbuso Joshua Nyoni[5]

[1] Green Energy Pacesetter, [2] Kabete National Polytechnic, [3] Innercore Geo-consultants, [4] Jomo Kenyatta University of Agriculture and Technology, [5] University of Edinburgh

Email: nyokabijesse@gmail.com[1], oyuginoah@gmail.com[2], johnmbugs008@outlook.com[3], mwangihiuhu2@gmail.com[4], kumbujosh@gmail.com[5]

Keywords: Geothermal, Reservoir Engineering, Machine Learning, Algorithms

ABSTRACT

Reservoir engineering constitutes a major part of the studies regarding geothermal exploration and production. Reservoir engineering comprises conducting experiments, constructing appropriate models, characterization, and forecasting reservoir dynamics. However, traditional engineering approaches started to face challenges as the number of raw field data increased. It pushed the researchers to use more powerful tools for data classification, cleaning, and preparing data to be used in models, which enhances a better data evaluation, thus making proper decisions. In addition, simultaneous simulations are sometimes performed, aiming to have optimization and sensitivity analysis during the history matching process. Machine learning techniques have been proposed as the best solution for strong learning capability and computational efficiency. Recently developed algorithms make it possible to handle a very large number of data with high accuracy.

In this study, we endeavor to review the machine learning approaches used in geothermal reservoir engineering and give the recommendation of more areas to explore. Areas of reservoir management and reservoir monitoring are the subcategories that have the most applications of machine learning under reservoir engineering. Most works involved predicting production data such as flow rates, temperature, and pressure. Though machine learning has been applied in geothermal reservoir engineering, this study shows that there are still opportunities to improve and expand the adoption of machine learning.

1. INTRODUCTION

1.1 Geothermal Reservoir Engineering

The main purpose of reservoir engineering is to investigate the physical properties and the processes occurring within the subsurface to approximate reserve estimates and advice on methods of maximum sustainable exploitation of the reservoir (Grant & Bixley, 2011). To fulfill these objectives, reservoir engineers are tasked to carry out numerical modeling, chemical analysis of reservoir fluids, well testing, and production testing. Geothermal reservoir engineers estimate the potential of the geothermal reservoir in terms of energy production.

Reservoir engineering is done in two phases. The first phase is to analyze the reservoir properties in its undisturbed state, i.e. during exploration, it is essential to collect and evaluate this information – fracture network, recharge system, and source of heat - since they cannot be obtained once exploitation start (Grant & Bixley, 2011). The second phase is during production, generally, the pressure transients of the reservoir and the flow of hydrothermal fluid and its relation to the hydro-geologic structure are established and monitored over the period of production.

There are two main types of high-temperature geothermal reservoirs – liquid-dominated and vapor dominated. Geothermal reservoirs have distinct features such as the reservoir having multiple zones of recharge and discharge, the fluid composition is not uniform, and the reservoir itself is not isothermal or static (Okoroafor, et al., 2022). For this reason, the flow of hydrothermal fluid (water, steam, or a mixture of the two) through the rock fractures is the unifying feature in the analysis of the properties of a geothermal reservoir.

Numerical models are used to simulate energy production, thermal output, temperature distribution, permeability distribution, and pressure transients of a reservoir (Okoroafor, et al., 2022). The models generated are applied to data from production wells to fully understand the processes occurring in the subsurface to see what roles they might play in the reservoir being exploited. These models take time to build and run, also there is a rising need to integrate multidisciplinary and multidimensional data in reservoir engineering. For this reason, the geothermal energy industry has been looking toward AI-based subsurface modeling to increase model computational time.

1.2 The Concept of Artificial Intelligence and Machine Learning

Artificial intelligence (AI) refers to the science and engineering of making intelligent machines, especially intelligent computer programs (McCarthy, 2004). There has been a boom in AI applications in many industries. Unfortunately, there has been less application in the energy industry compared to other industries. This is partly due to considerable reliance on physics-based models for describing processes related to energy resource exploitation (Okoroafor, et al., 2022).

Machine learning (ML) is a branch of artificial intelligence that applies statistical methods to train computational models from data. Machine Learning algorithms can be thought of broadly as mathematical functions containing parameters that map inputs (features) into one or more outputs (or targets) (Okoroafor, et al., 2022).

Each application of Machine Learning (ML) is unique, from the algorithms to the expected results. In reservoir engineering ML is used in the prediction of reservoir properties and energy production during exploration and exploitation. This involves analyzing production well data, geochemical data, and geophysical data.

The standard procedure followed in a Machine Learning project starts with data preparation, this is where the data is transformed and standardized to a form that is appropriate for modeling. ML algorithms require data to be in numerical format hence the data must be processed to numbers. The process of data preparation serves the purpose of exposing the unknown underlying structure of the problem to the learning algorithm. The main challenge in data preparation is the unique nature of each dataset, and when it comes to reservoir engineering this poses a great challenge and a lot of distinct attention is needed to ensure that the data available for prediction doesn't leave out important features.

The next step is to explore the data to establish the relationship between the various features of the data. This usually involves making visualizations of the features in the dataset (Brownlee, 2020). This step helps in framing the prediction task; the reservoir engineer can create a mental picture of the processes in the reservoir.

After we have taken a look at the data and established relationships between the features, it's now time to build the ML model. The model is built with respect to the statistics of the data. There are two broad categories of ML models to choose from – Regression models or Classification models. The choice and design of the model solely depend on the reservoir engineering problem (Brownlee, 2020). The models are then evaluated so that the results obtained can be trusted. This is done by selecting a performance matrix to evaluate the skill of the model, and resampling the split data into testing and training sets to simulate how the final model is going to be used. Hyper-parameter tuning and ensemble of models are also used to get more from the data at this stage.

Once a suite of models has been evaluated, the best model is chosen to represent the solution to the problem. Further evaluation can be done based on project-specific criteria such as model complexity. The model can be finally integrated into software or a data management system and may periodically require monitoring and maintenance (Brownlee, 2020).

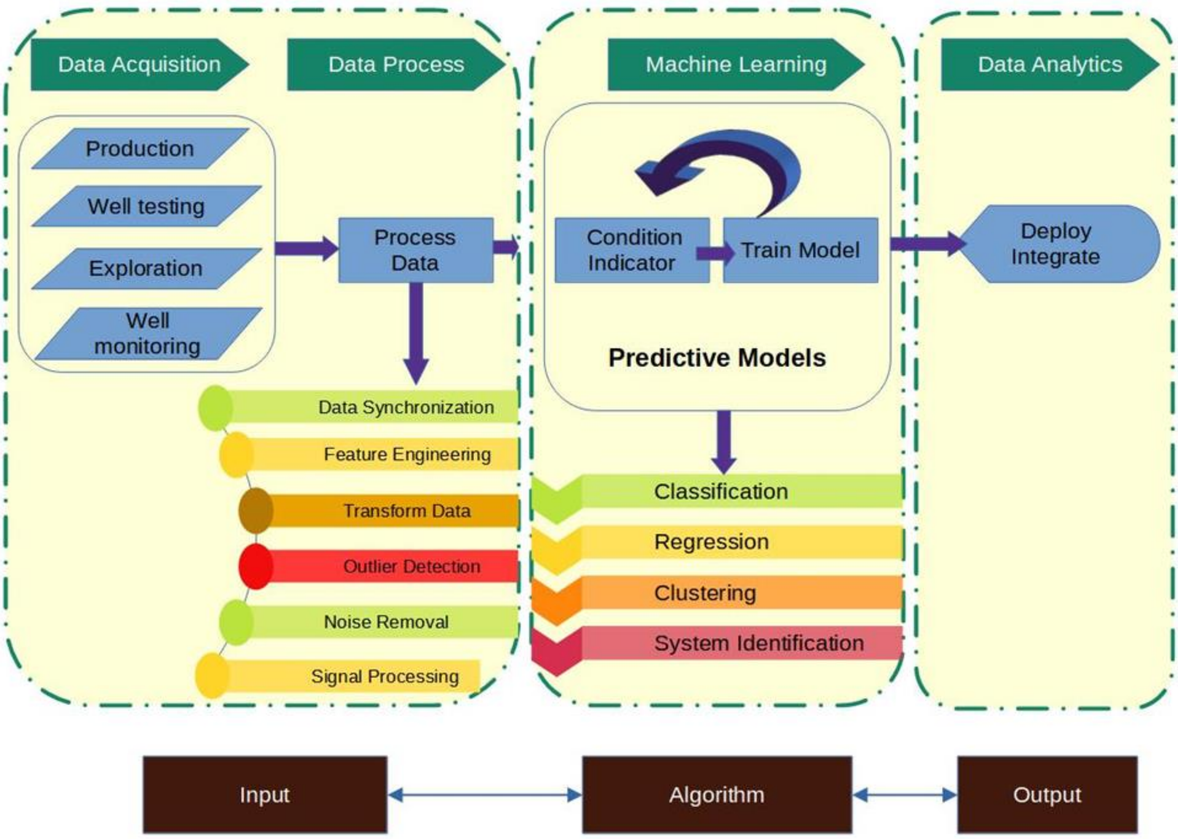

Figure 1: An illustration of Machine Learning workflow for reservoir engineering.

2. DATA ANALYTICS VS MACHINE LEARNING ANALYTICS

2.1 A Review of Data Analytics.

A large number of data sources, data types, and the rate at which the data is being provided has necessitated the use of data analytic techniques to cope with the complex challenges. Analytic solutions are strong tools that provide finer decision support to the users so that they can make better decisions by evaluating applicable information. There are four types of data analytics that can be listed as follows from the simplest to more sophisticated one: descriptive analytics, diagnostic analytics, predictive analytics, and perspective analytics.

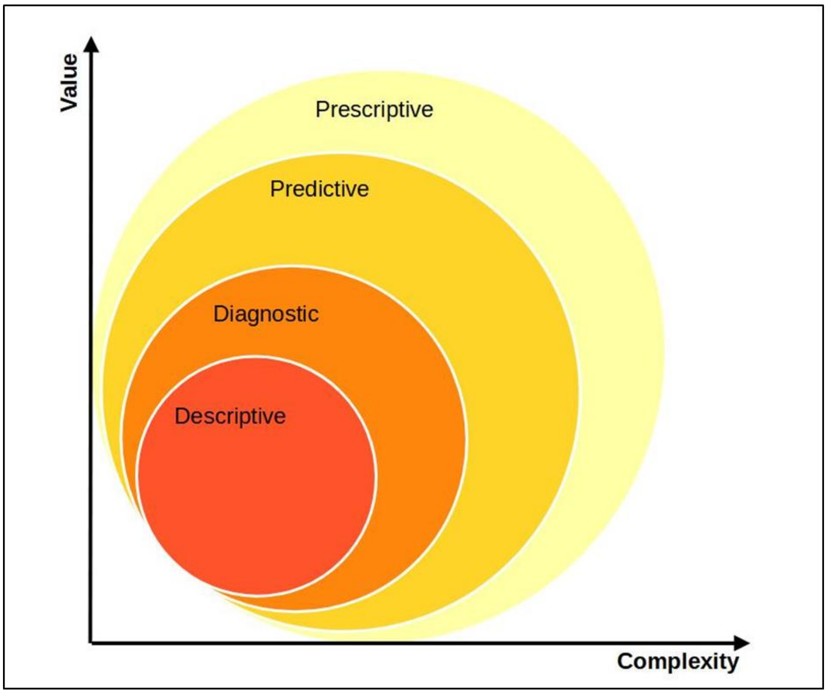

Figure 2: Data analytics value vs complexity (Balali et al 2020).

2.1.1 Different Forms of Data Analytics

Data analytics is the process of examining the data to draw a conclusion regarding the information they may contain. Data analytics have been widely used in both industry and academic areas. Industries seek to make more informed decisions, and researchers try to verify scientific models, hypotheses, and theories. Data analytics are so close to business analytics (BA) in terms of reporting and online analytical processing (Gandomi and Haider 2015).

Analytics has the power to change what organizations are going to do at which point in time. Data analytics can highly boost business performances by enhancing revenue, cost, resource utilization, and operational efficiency. Furthermore, data analytics enable businesses to quickly identify and respond to the changes in either market or demand. Depending on the application, the data which passes through an analytical tool may include a historical record or real-time information (Grover et al 2015).

Data analytics methodologies can include explanatory data analysis, to find any relation or pattern in the data sets, or confirmatory data analysis to apply statistical tools for verifying hypotheses regarding the data sets. On the other hand, the power of analytical tools, such as machine learning (ML) or artificial intelligence (AI) tools, could highly enhance the effectiveness of data analytics. Data analytic tools may have access to the managers at various levels to keep track of the processes more conveniently (Ousterhout et al).

i. Descriptive Analytics

Descriptive analytics is the task of providing a representation of the knowledge discovered without necessarily modeling a specific outcome. The tasks of cluster analysis, association and correlation analysis, and pattern discovery, can fall under this category. From a machine learning perspective, we might compare these algorithms to unsupervised learning (Berman and Israeli 2022). Descriptive analytics describes a phenomenon through different measures that could capture its relevant dimensions. The purpose is to simply unravel 'what happened' or alert on what is going to happen.

ii. Predictive Analytics

Often our task in data mining is to build a model that can be used to predict the occurrence of an event. The model builders will extract knowledge from historic data and represent it in such a form that we can apply the resulting model to new situations. We refer to this as predictive analytics. The tasks of classification and regression are at the heart of what we often think of as data mining and specifically predictive analytics.

Indeed, we call much of what we do in data mining predictive analytics. From a machine learning perspective, this is also referred to as supervised learning. The historic data from which we build our models will already have associated with specific outcomes (William 2011). Predictive analytics seeks options for future business imperatives, predicts potential future outcomes, and explains drivers of the observed phenomena using statistical or data mining techniques

iii. Diagnostic Analytics

Diagnostic analytics evaluates 'why' something happened. It needs exploratory data analysis of the existing data or additional data if required to be collected using tools such as visualization techniques to discover the root causes of a problem. Correlations, data mining, and discovery are some useful approaches used in this method.

iv. Prescriptive Analytics

Prescriptive analytics goes beyond describing, explaining, and predicting to suggest 'what courses of action may be taken for the future to optimize business processes to achieve business objectives. In other words, it associates decision alternatives with the prediction of outcomes. For prescriptive analytics, decision analysis is used which includes tools such as optimization and simulation (Banerjee et al 2013).

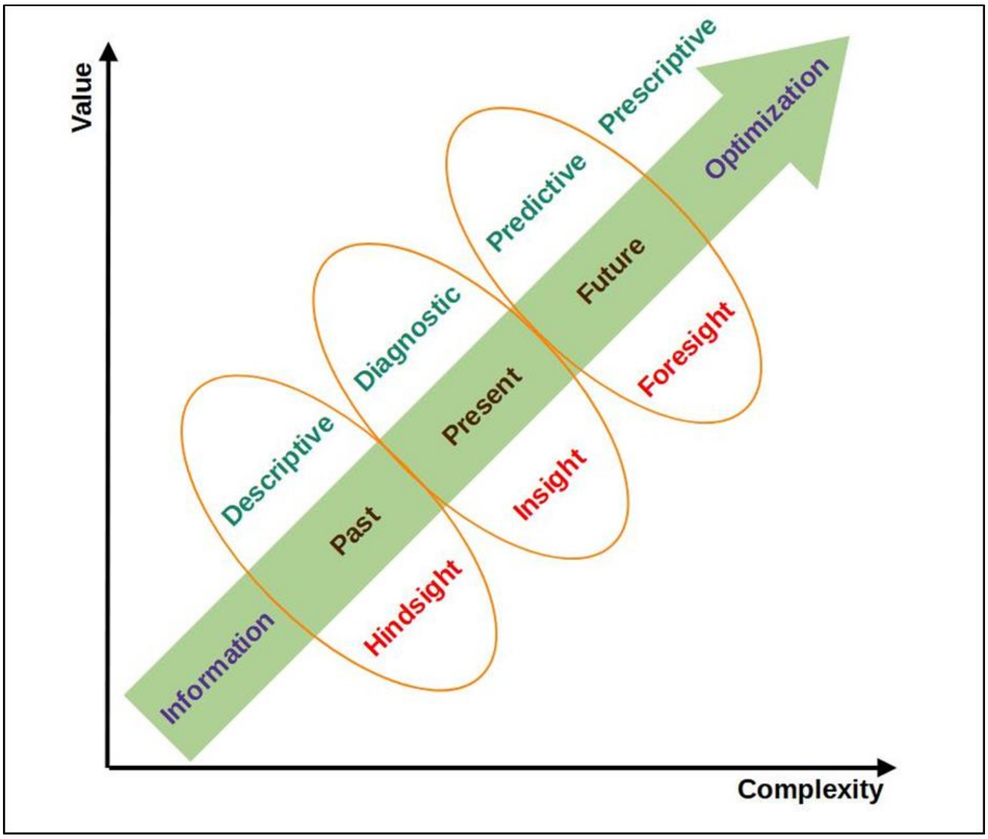

Figure 3: Data analytics information, value, and complexity.

The various forms of analytics play important roles and have their place. Data analytics seeks to bring more value to an organization. What brings advantages to the organizations is not the size or amount of data. Indeed, the efficient utilization of the data is key for an entity to be successful. Descriptive analytics try to explain the events, which have occurred in the past. Diagnostic analytics attempt to find out the reason for occurring each of these events. Predictive analytics take advantage of historical data and the current status of the system to predict the future condition or likelihood of occurrence of an event. Prescriptive analytics make optimum decisions based on the prediction outcomes and their consequences. It should be considered that the complexity and uncertainty of the data analytics would increase as we move forward in prediction time. Descriptive and prescriptive analytics have the least and the most level of complexity and uncertainty, respectively.

Artificial intelligence (AI) and big data could highly affect the effectiveness of data analytics, especially where the uncertainty and complexity are considerable. Statistical analyses are usually trying to approve or reject a hypothesis by assessing the correlation among the data. Machine learning (ML) is mostly about predicting the possible outcomes in the future based on various variables.

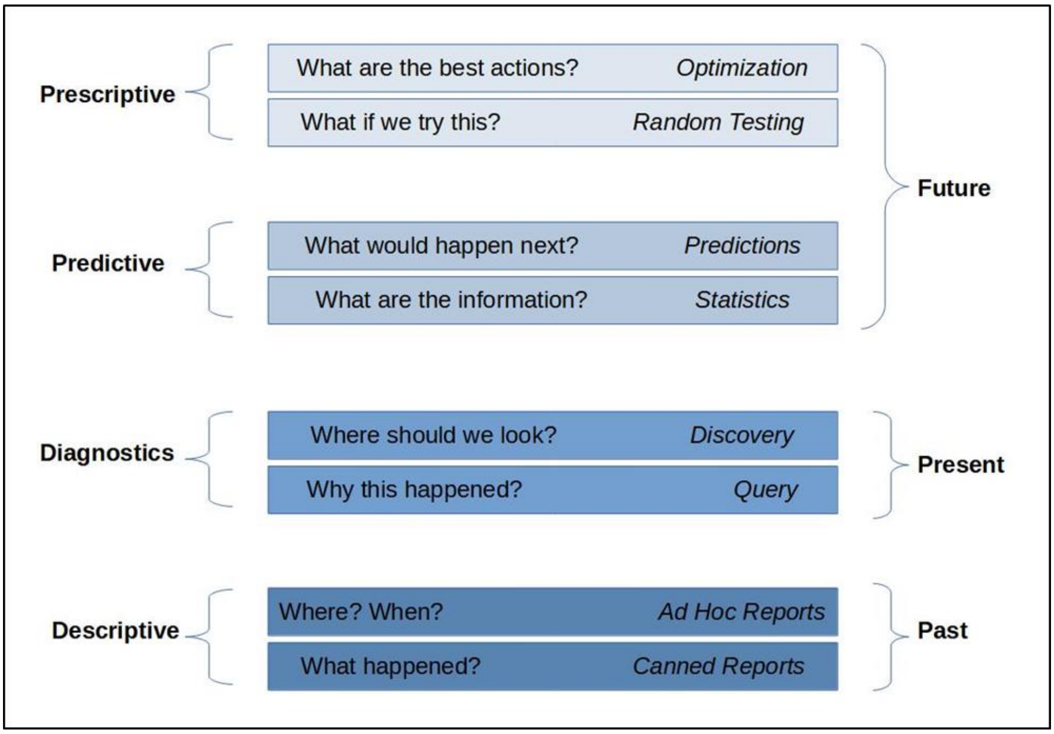

Figure 4: Data analytics in the context

Data analysts have advanced skill sets that they can't use effectively when they're spending their time stuck in a cycle of routine reports. The limitations of this process have paved the way for machine learning to take hold in analytics.

2.2 Review Machine Learning analytics.

Machine learning is a subset of artificial intelligence (AI) that leverages algorithms to analyze vast amounts of data. These algorithms operate without human bias or time constraints, computing every data combination to understand the data holistically. Further, machine learning analytics understands the boundaries of important information.

2.2.1 Applications of Machine Learning.

Many problems that ML sets out to solve require a bespoke approach. As such, the types of instructions needed for each will be very different. However, there are generally four main categories that these algorithms fall into:

- **Supervised learning –** With this method of machine learning, you train the algorithm using a labeled set of data to learn from. So, there are already some known answers, and it can determine whether new data matches it. As it produces results, it can evaluate them based on the information you've already provided. The more data you give it initially, the more it knows about unseen data.

- **Unsupervised learning –** In this type of machine learning algorithm, the program is trained with data that isn't labeled. It doesn't know what the data represents. Instead, the computer detects patterns, finds rules within them, and summarizes where there are relationships in the data.

- **Semi-supervised learning –** As you might expect, this type of algorithm uses elements of both of the above. The data you provide to teach the machine will have some labels, which are used to help process larger sets of unlabeled data.

- **Reinforcement learning –** This method of machine learning is focused on continuous learning and reward using unlabeled data. A useful way of thinking about this concept is with video games. If a computer wins a game, it receives positive feedback.

2.2.2 Typical Applications.

- Automation – Perhaps the most high-profile machine learning use is in the automation of tasks humans usually perform. The ability of a computer to think and act without being programmed has incredible potential.

- Insights – Machine learning algorithms can process and analyze huge sets of data. Often used in the field of big data, such insights can help businesses understand their customers and healthcare professionals understand their patients.

- Recommendation – Based on previous input data, machine learning can recommend products and services that users or customers might like. This is perhaps one of the most common forms of machine learning you'll see in your day-to-day life.

- Detection – The way that machine learning works makes it ideal for spotting anomalies in patterns. As algorithms learn what 'normal' is, they become more adept at detecting when things go wrong.

3. BACKGROUND ON MACHINE LEARNING COMPARATIVE APPLICATIONS IN GEOTHERMAL.

The main role of ML algorithms is to develop predictive models by using the training and verifying data sets. The output of the ML algorithms could be either a continuous or discrete variable. As the statistics present, machine learning (ML) algorithms have become one of the most attractive subjects for researchers during the last few decades. Various studies focus on either developing or enhancing the ML algorithms for various applications. It should be noted that the basics of ML algorithms are based on statistical principles. From the statistical point of view, ML algorithms have been developed as the integration of the statistical algorithms with computer science and data mining philosophies (Balali et al 2020).

The main difference between ML and traditional algorithms is in the way in which the analyst defines the rules. As Figure presents, traditional algorithms are based on predefined rules which are all coming from a series of logic that can extract the output with respect to the inputs. It means that there should not be any uncertainty or unknown entity for the traditional algorithms. It should be considered that more sophisticated rules are usually needed as the system becomes more complex. For this reason, it might be possible that the complex systems become unsustainable to maintain using the traditional algorithms. ML algorithms are supposed to overcome this issue since, in ML algorithms, the machine is in charge of defining the rules between the inputs and outputs.

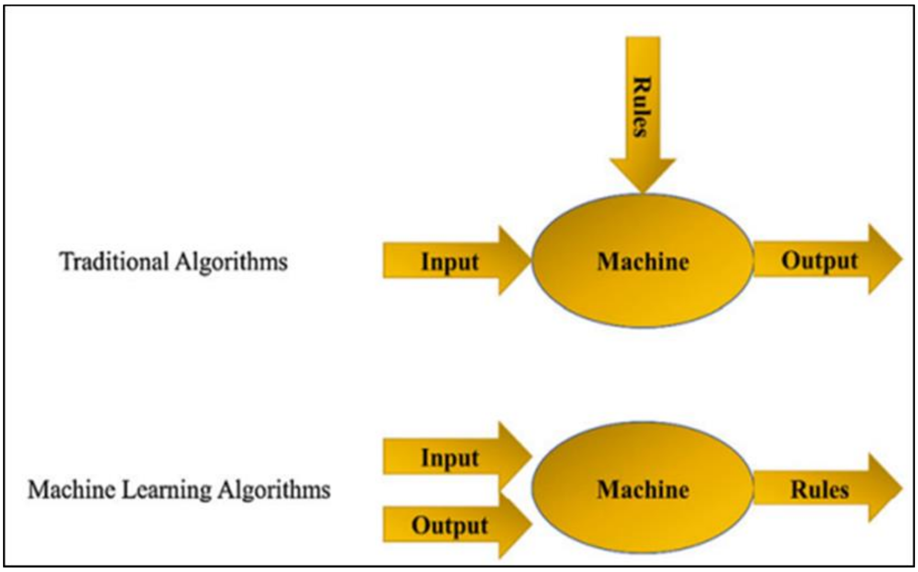

Figure 5: Traditional Algorithms versus Machine Learning Algorithms.

ML algorithms try to develop models based on the training data sets. Feature vectors are extracted based on the training data sets, and ML algorithms are trying to understand more beneficial information regarding the system based on the training data set. The built model can be tested using the verification data set which does not have any overlap with the training sets. If the model acquires acceptable performance measures, it can be applied to other cases to perform the prediction process.

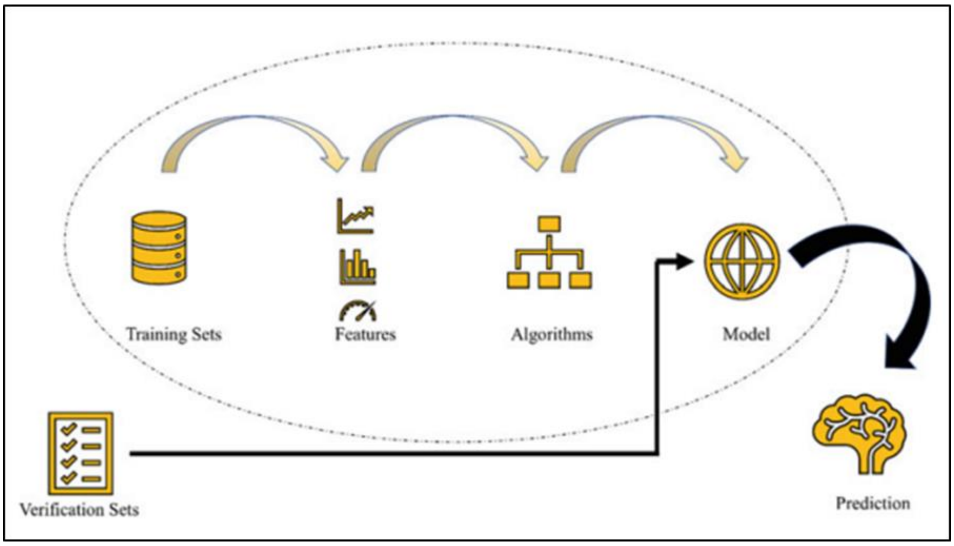

Figure 6: An overview of the Machine Learning process

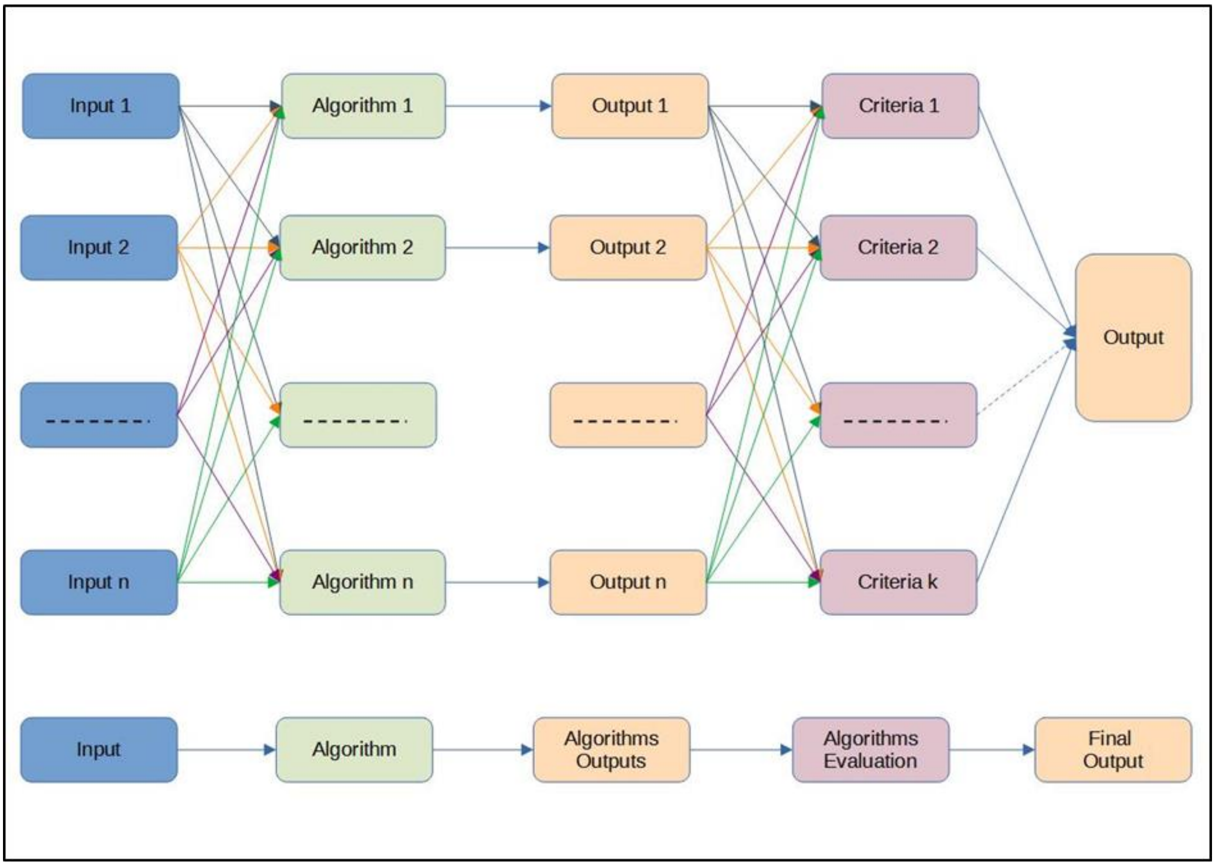

Figure 7: An overview of Machine Learning Algorithms.

3.1 Application of Machine Learning in geothermal reservoir engineering.

The main intention of reservoir engineering especially in geothermal is to generate reserve estimates and methods for maximum economical production. To reach this objective, the work of geothermal reservoir engineers is extensive, involving numerical reservoir modeling, production forecasting, and tracer testing. Also well testing, chemically analyzing the reservoir fluid, and economic modeling. Reservoir engineers evaluate the energy production potential of geothermal reservoirs. Geothermal reservoirs have many convolutions, such as they are not isothermal, not static, or uniform in fluid compositions. The phase behavior of the geothermal working fluid is complex, and the system is rarely closed with multiple zones of recharge and discharge. Subsurface geothermal numerical models take a significant time to build and run (Intelligent Solutions Inc, 2011). Therefore, the energy industry has been looking toward AI-based subsurface models. Notably, in geothermal systems, accounting for non-isothermal effects would increase the model computational time. Consequently, it is valuable to develop surrogate models that use reservoir description data to predict the extracted fluid temperature, which is time-series data.

The application of machine learning and deep learning in reservoir engineering under geothermal energy resources has been limited, but a sharp increase has been evident over the last three years showing that the areas of reservoir management and reservoir modeling are the subcategories that have the most applications of machine learning under reservoir engineering. Major works have involved predicting production data, such as flow rates, temperatures, and pressures, using injection histories.

Aydin et al. (2020) developed an artificial neural network (ANN) to estimate reservoir pressure and temperature using the inputs as wellhead data and non-condensable gas concentrations. They trained the ANN using synthetic data, consisting of 87,360 simulations with a broad spectrum of operating parameters. They delineate a model with less than 2% error between the synthetic data and the ANN predictions. In addition to synthetic data, and to further showcase the benefits of the model, they applied the model to data from a geothermal production well in the Alasehir geothermal field in Turkey.

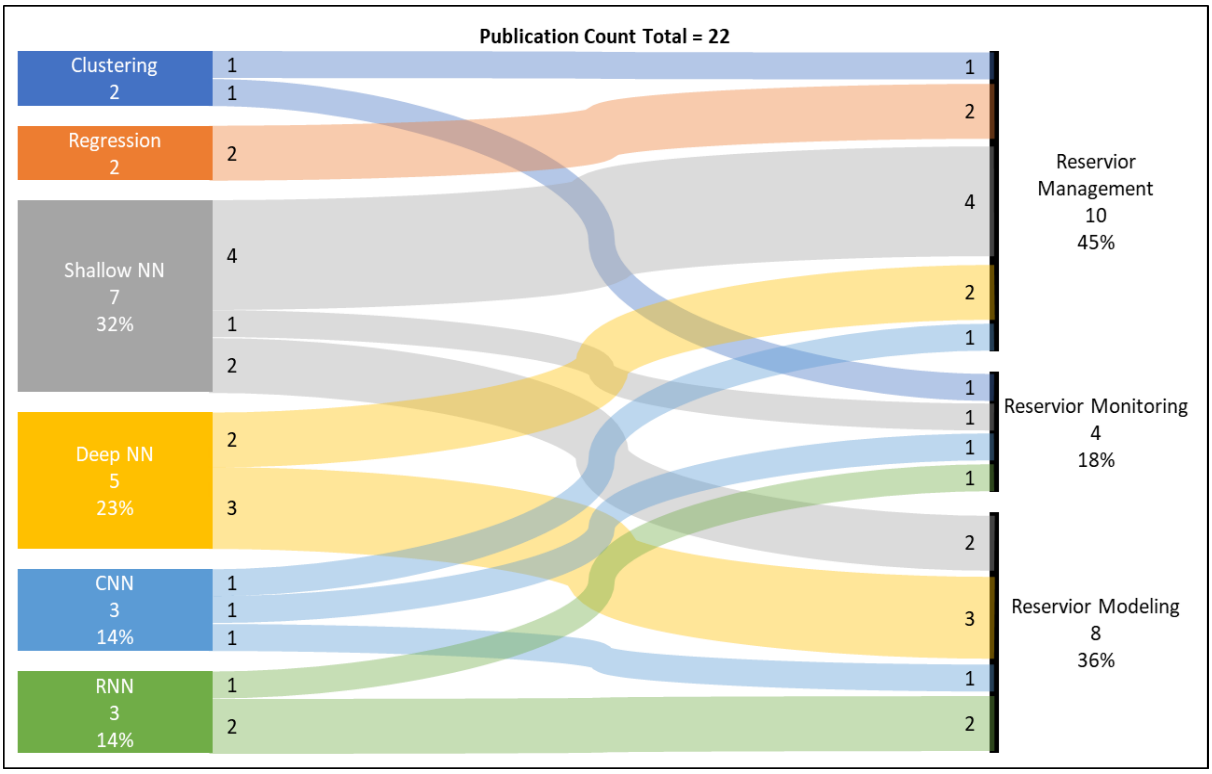

Figure 8: Summary of machine learning publications for subcategories of reservoir engineering in geothermal development (Okoroafor et al, 2022).

Suzuki et al. (2021) developed synthetic permeability distributions and simulated them within a TOUGH2 reservoir simulator. Support Vector Machine (SVM) was trained to predict permeability distributions with inputs being pressure and temperature measurements. Likewise, other researchers such as Akin et al. (2010), Gudmundsdottir and Horne (2020), Pandey and Singh (2021), and Beckers et al. (2021) have developed deep neural network models to serve as proxy models that predict production data. Additionally, Gudala and Govindarajan (2021) included geomechanics in a deep neural network and compared it to conventional ML models, decision trees, random forests, and support vector machines. Their results showed that deep neural networks obtained higher accuracy than traditional ML models.

Shi et al. (2021) utilized a more complex model structure, to not exclusively take into consideration the historical production data, but also reservoir properties and operational settings. The model structure consisted of a combination of Long Short Term Memory (LSTM) and multilayer perceptron (MLP) artificial neural networks. The MLP part of the model cultivated the relationship between the production data and the reservoir properties and operations, while on the contrary, the LSTM considered the sequential dependencies within the production data. The LSTM and MLP combinational neural network precisely predicted the productivity of a multilateral-well geothermal system and conspicuously outperformed the original MLP and LSTM neural networks.

Other remarkable techniques that have been employed are Ishitsuka et al. (2021) that compared Bayesian estimation and a feedforward neural network (FNN) to approximate temperature distribution in deep unexplored regions using electrical resistivity data, geological boundaries and micro-seismic observations, and a numerical model of the Kakkonda geothermal field in Japan. Li et al. (2017) used K-means clustering to group production wells based on their spatial relationships to reduce complexity in an aggregated parameter tank model. Ishitsuka et al. (2021) work showed that both Bayesian estimation and FNN performed well, and the authors suggested that selection between the approaches should be grounded on uncertainty in input parameters. In Li et al. (2017), data from the Laugarnes geothermal field and Reykir geothermal field in Iceland were used to show the method's effectiveness. Baser et al. (2021) also simulated different injection and production scenarios using a reservoir model of the Kizildere geothermal field in Turkey. The simulated data was applied to train a neural network to predict production. Shi et al. (2020) used a numerical model to generate simulation production data, production temperature, and thermal output under different constraint environments. The constraint conditions were reservoir temperature, injection mass flow rates, reservoir permeability, and lateral-well spacing. An LSTM and an MLP neural network were developed to forecast future production, with the LSTM having higher prediction accuracy than the MLP.

Kiran and Salehi (2020) used the FORGE well log data of already drilled wells and amalgamate the evolution of dynamic data with a 5-second interval when investigating the different data mining algorithms that can capture the pattern in the operational parameters considering the other petrophysical properties. The suitable algorithms were implemented to determine and eliminate the effect of operational parameters using a series of digital filtering techniques. Then, the filtered version of the geothermal well logs were used as input for unsupervised machine learning algorithms such as k-nearest neighbors, decision tree classification, and deep learning models with hidden layers. Decisively, hazardous zones were classified using the classifications, which were reported to improve confidence in the operation.

Beckers et al (2021), conducted hundreds of reservoir simulations in TETRAD-G and CMG STARS to explore different injection and production fluid flow rates and to develop a training data set for Machine Learning (ML). This process included simulating the documented injection and production since 1979 and forecasting future performance through 2040. ML networks were created and trained using TensorFlow based on multilayer perceptron, long short-term memory, and convolutional neural network architectures. These networks took these selected inputs: flow rates, injection temperatures, and historical field operation data, and produced estimates of future production temperatures. This approach was successfully tested on a simplified single-fracture doublet system, followed by the application to the Brady Hot Springs BHS reservoir. Using an initial BHS data set with 37 simulated scenarios, the trained and validated network predicted the production temperature for six production wells with the mean absolute percentage error of less than 8% achieved. In a complementary analysis effort, the principal component analysis applied to 13 BHS geological parameters revealed that vertical fracture permeability shows the strongest correlation with fault density and fault intersection density. The team developed the BHS reservoir model considering the fault intersection density as a proxy for permeability. This new reservoir model assisted to explore under-exploited zones in the geothermal reservoir.

Gudmundsdottir and Horne (2018) explored the applicability of applying statistical methods for reservoir characterization as well as prediction modeling. They explored three methods to analyze and apply to a synthetic library of fracture networks. First, the Alternate Conditional Expectation (ACE) algorithm was used to estimate well-to-well connectivity between injection and production wells using tracer return and temperature data. The outcomes obtained with tracer data were in good agreement with tracer transit times, for 80.5% of the fracture networks the ACE connectivity was within ±0.05 of the connectivity implied by transit time, while temperature data showed much less correlation to connectivity with the ratio reduced to 58.3%. Second, k-means clustering was applied where fractures of a similar character were grouped together and inter-well connectivity and thermal behavior estimated. The method displayed potential but the main constraints were deciding on the number of clusters and the growing complexity with added producers. Third, preliminary results using direct forecasting with Canonical Functional Component Analysis (CFCA) were presented. A momentous reduction in the predicted range of thermal responses for production wells was obtained but introducing more complex data is likely to cause data-prediction relationships to become less linear.

Gudmundsdottir and Horne (2020) investigated models to serve as supplementary for the full reservoir. They used deep learning algorithms for this purpose and explored two different architectures, standard feedforward neural networks, and recurrent neural networks. The mapping was accomplished with injection flow rates at the injectors as input and tracer concentration data at the producers as output. The models were trained on a synthetic geothermal reservoir and their analysis concluded that for this case the simple feedforward neural network outperformed the more complex recurrent neural network. For this synthetic geothermal reservoir, the simple feedforward neural network (MLP) is reported to outperform the more flexible structure of the recurrent neural network. Adding the target feed significantly is also reported to elevate the performance of the models and in the case of MLP, they were able to simplify the model architecture by reducing the number of hidden layers needed. The MLP was also much quicker to train, compared to the RNN, with fewer parameters to estimate. When creating one model to predict for all producers, compared to having separate models, they again found that the MLP performed better than the RNN.

Haklidir F., and Haklidir M., (2019) developed a Deep Neural Network (DNN) model, to predict the geothermal fluid temperatures based on hydro-geochemistry data from Western Anatolia (Turkey). This was an early stage study for reservoir prediction and emphasize data quality also. A comparative study of traditional machine learning algorithms was performed to measure the performance of DNN predictions. In their study, natural thermal springs at different temperatures were used to create the dataset which belongs to different geothermal systems.

More than 60 thermal springs were selected for the training set and 16 set analyses of them are selected for the test data. Both regression and classification approaches were used in their study. The different algorithms were applied to the same data to reach a consistent comparison of results for each used approach. In the regression approach, neither the Linear Regression nor the linear SVM algorithms gave satisfactory results for the low and high-temperature prediction for the fluids. DNN algorithm gave more accurate values, especially for the High-temperature fluids which temperatures higher than 60 °C.

4. DISCUSSION

4.1 Challenges Encountered in the Applications of Machine Learning Methods to Geothermal reservoir engineering.

Despite the extraordinary advancement and accessibility of artificial intelligence algorithms and implementations, the adoption of this technology in geothermal reservoir engineering is associated with technical, logistical, and financial challenges (Okoroafor et al, 2022). Machine learning algorithms are data-driven and hence they cannot model a physical phenomenon without suitable datasets. Whereas geothermal reservoir engineering generates large volumes of raw data, these data should be characterized by accessibility, accuracy, structure, relevance, and security, among others (Ransbotham et al., 2017). The establishment of such big datasets requires months of planning and careful collection workflows alongside quality assurance and control protocols (Roh et al., 2019). For the espousal of machine learning to succeed in geothermal reservoir engineering, there is an urgent need for open and collaborative efforts across the geothermal industry to ensure that diverse datasets and lessons learned to reach talents to enable rapid and impactful development of machine learning models within the geothermal reservoir engineering.

In addition, data handling and preprocessing are other challenges facing the application of artificial intelligence algorithms within geothermal reservoir engineering, where large volumes of unstructured data are generated using various sensors and techniques. Such data complicates the training process of machine learning algorithms which aims to capture the true distribution of physical phenomena without overfitting the training and development datasets. This challenge can be resolved by upgrading geothermal facility instruments, allocating sufficient time for data collection, and recruiting a multidisciplinary team. Therefore, successful organizational adoption of machine learning within geothermal reservoir engineering demands a business model that is lean and agile in mindset and infrastructure.

4.2 Opportunities.

With the expeditious evolution of artificial intelligence algorithms and successful integration across many industries, there is an infinite of opportunities for machine learning applications within geothermal energy. From the review of the areas where machine learning has been applied, we see that there is a significant margin to improve and expand the adoption in geothermal reservoir engineering.

There are low-hanging opportunities in capitalizing on the successful utilization of machine learning algorithms across the oil and gas industry (Temizel et al, 2021), an industry that shares similar technical challenges with the geothermal industry. With the advent of transfer learning, where an artificial intelligence framework that has been trained on adequate data can be used for a similar domain that may not have enough training data, machine learning architectures from models in oil and gas reservoir engineering can be used as a starting point for similar tasks in the geothermal reservoir engineering.

Another opportunity lies in creating awareness and encouraging the utilization of machine learning algorithms within the geothermal reservoir engineering community through conferences, articles, geohackathon competitions, workshops, and courses. The collaboration across and within academic and professional entities is of paramount importance to expedite the progress and experimentation of machine learning algorithms on a broader range and to accumulate more diverse datasets.

5. CONCLUSION.

The paper includes a detailed review of machine learning applications in geothermal reservoir engineering. The trend in machine learning applications in geothermal reservoir engineering suggests that the adoption of machine learning is likely to increase in the near future. There is room for deep learning algorithms to be applied in geothermal reservoir engineering.

This study shows that there is still a significant margin to expand the usage of machine learning in geothermal reservoir engineering. In addition, the geothermal industry could leverage transfer learning from machine learning models of the oil and gas industry as starting points for similar domain tasks in the geothermal industry.

REFERENCES

Agostinelli, F., Hoffman, M., Sadowski, P. and Baldi, P., 2014. Learning activation functions to improve deep neural networks. arXiv preprint arXiv:1412.6830.

Akin, S., Kok, M., Uraz, I., 2010. Optimization of well placement geothermal reservoirs using artificial intelligence. Comput. Geosci. 776–785. https://doi.org/10.1016/j. cageo.2009.11.006

Aydin, H., Akin, S., Senturk, E., 2020. A proxy model for determining reservoir pressure and temperature for geothermal wells. Geothermics 88. https://doi.org/10.1016/j. geothermic.2020.101916

Balali, F., Nouri, J., Nasiri, A., and Zhao, T., 2020. Data analytics. In Data-intensive industrial asset management (pp. 105-113). Springer, Cham.

Banerjee, A., Bandyopadhyay, T., and Acharya, P., 2013. Data analytics: Hyped-up aspirations or true potential? Vikalpa, 38(4), pp.1-12.

Baser, A., Kucuk, S., Saracoglu, O., Senturk, E., Akin, S., 2021. Optimization of production and injection of geothermal fields: a machine learning approach. In: Proceedings of the World Geothermal Congress. Reykjavik, Iceland.

Beckers K., Duplyakin D., Martin M., Johnston H., Siler D., 2021. Subsurface Characterization and Machine Learning Predictions at Brady Hot Springs. Proceedings of the Forty-sixth Workshop on Geothermal Reservoir Engineering, Stanford University.

Berman, R. and Israeli, A., 2022. The Value of Descriptive Analytics: Evidence from Online Retailers. Marketing Science.

Breiman, L., 2001. Random forests. Machine learning, 45(1), pp.5-32.

Brownlee, J. (2020). *Data Preparation for Machine Learning*

Chai, C., Maceira, M., Santos-Villalobos, H., Venkatakrishnan, S. and Schoenball, M., 2020. Automatic Seismic Phase Picking Using Deep Learning for the EGS Collab project. Oak Ridge National Lab. (ORNL), Oak Ridge, TN (United States).

Chorowski, J.K., Bahdanau, D., Serdyuk, D., Cho, K. and Bengio, Y., 2015. Attention-based models for speech recognition. Advances in neural information processing systems, 28.

Cover, T., and Hart, P., 1967. Nearest neighbor pattern classification. IEEE transactions on information theory, 13(1), pp.21-27.

Denoeux, T., 2008. A k-nearest neighbor classification rule based on Dempster-Shafer's theory. In Classic works of the Dempster-Shafer theory of belief functions (pp. 737-760). Springer, Berlin, Heidelberg.

Deselaers, T., Hasan, S., Bender, O. and Ney, H., 2009, March. A deep learning approach to machine transliteration. In Proceedings of the Fourth Workshop on Statistical Machine Translation (pp. 233-241).

Dietterich, T.G., 2000. An experimental comparison of three methods for constructing ensembles of decision trees: Bagging, boosting, and randomization. Machine learning, 40(2), pp.139-157.

Friedman, J.H., 2002. Stochastic gradient boosting. Computational statistics & data analysis, 38(4), pp.367-378.

Gandomi A., Haider M., (2015). Beyond the hype: Big data concepts, methods, and analytics. Int. J. Inf. Manag. 35(2), 137–144.

Grant, M. A., & Bixley, P. F. (2011). Concepts of Geothermal Systems. In M. A. Grant, & P. F. Bixley, *Geothermal Reservoir Engineering* (pp. 11-12).

Grover V., Chiang R., Liang T., Zhang D., (2018). Creating strategic business value from big data analytics: A research framework. J. Manag. Inf. Syst. 35(2), 388–423.

Gudala, M., Govindarajan, S., 2021. Numerical investigations on a geothermal reservoir using fully coupled thermo-hydro-geomechanics with integrated RSM-machine learning and ARIMA models. Geothermics 96. https://doi.org/10.1016/j. geothermics.2021.102174.

Gudmundsdottir H. and Horne R., 2018. Reservoir Characterization and Prediction Modeling Using Statistical Techniques. Proceedings of the Forty-third Workshop on Geothermal Reservoir Engineering, Stanford University.

Gudmundsdottir H. and Horne R., 2020. Prediction Modeling for Geothermal Reservoirs Using Deep Learning. Proceedings of the Forty-fifth Workshop on Geothermal Reservoir Engineering, Stanford University.

Gul, S., Aslanoglu, V., Tuzen, M.K. and Senturk, E., 2019. Estimation of bottom hole and formation temperature by drilling fluid data: a machine learning approach. In 44th Workshop on Geothermal Reservoir Engineering.

Haklidir F., and Haklidir M., 2019. The Fluid Temperature Prediction with Hydro-geochemical Indicators Using A Deep Learning Model: A Case Study Western Anatolia (Turkey). Proceedings of the Forty-fourth Workshop on Geothermal Reservoir Engineering, Stanford University.

Intelligent Solutions Inc. (2011). Surrogate reservoir models: an alternative to traditional numerical reservoir simulation and modeling. Morgantown, West Virginia. Retrieved from http://intelligentsolutionsinc.com/PDFs/WhitePaper-SurrogateReservoirMode ls.pdf

Ishitsuka, K., Y, K., N, W., Yamaya, Y., E, B., Suzuki, A., Saito, R., 2021a. Bayesian and neural network approaches to estimate deep temperature distribution for assessing a supercritical geothermal system: evaluation using a numerical model. Nat. Resour. Res. 30 (5), 3289–3314. https://doi.org/10.1007/s11053-021-09874-w

Kingma, D.P. and Ba, J., 2014. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.

Kiran R. and Salehi S., 2020. Assessing the Relation between Petrophysical and Operational Parameters in Geothermal wells. A machine Learning Approach. Proceedings of the Forty-fifth Workshop on Geothermal Reservoir Engineering, Stanford University.

Krizhevsky, A., Sutskever, I. and Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems, 25.

Li, Y., Júlíusson, E., Palsson, ´ H., Stefansson, ´ H., Valfells, A., ´ 2017. Machine learning for the creation of generalized lumped parameter tank models of low-temperature geothermal reservoirs systems. Geothermics 70, 62–84. https://doi.org/10.1016/j. geothermics.2017.05.009.

Natekin, A. and Knoll, A., 2013. Gradient boosting machines, a tutorial. Frontiers in neurorobotics, 7, p.21.

Okoroafor, R. E., Smith, C. M., Ochie, K. I., Chinedu, J. N., Gudmundsdottir, H., & Aljubran, M. (2022). Machine Learning in Subsurface Geothermal Energy: Two Decades in Review. *Geothermics*.

Ousterhout K., Rasti R., Ratnasamy S., Shenker S., Chun B., Making sense of performance in data analytics frameworks, pp. 293–307.

Pal, M., 2005. Random forest classifier for remote sensing classification. International journal of remote sensing, 26(1), pp.217-222.

Pandey, S.N., Singh, M., 2021. The artificial neural network to predict the thermal drawdown of an enhanced geothermal system. J. Energy Resour. Technol. (1), 143. https://doi.org/10.1115/1.4048067

Polat, K. and Güneş, S., 2007. Classification of epileptiform EEG using a hybrid system based on a decision tree classifier and fast Fourier transform. Applied Mathematics and Computation, 187(2), pp.1017-1026.

Pollack, A., 2020. Quantifying Geological Uncertainty and Optimizing Techno-economic Decisions for Geothermal Reservoirs. Stanford University.

Ransbotham, S., Kiron, D., Gerbert, P., Reeves, M., 2017. Reshaping business with artificial intelligence: closing the gap between ambition and action. MIT Sloan Manag. Rev. (1), 59

Rodriguez-Galiano, V.F., Ghimire, B., Rogan, J., Chica-Olmo, M. and Rigol-Sanchez, J.P., 2012. An assessment of the effectiveness of a random forest classifier for land-cover classification. ISPRS journal of photogrammetry and remote sensing, 67, pp.93-104.

Roh, Y., Heo, G., Whang, S.E., 2019. A survey on data collection for machine learning: a big data-AI integration perspective. IEEE Transactions on Knowledge and Data Engineering.

Sankran S., Matringe S., Sidahmed M., Saputelli L., Wen X., Popa A., Dursun S., 2020. Data Analytics in Reservoir Engineering, Society of Petroluem Engineering.

Sinha, S., de Lima, R.P., Qi, J., Infante-Paez, L. and Marfurt, K., 2018, October. Well-log attributes to map upward-fining and upward-coarsening parasequences. At the 2018 SEG International Exposition and Annual Meeting. OnePetro.

Shi, Y., Song, X., Li, G., 2020. Productivity prediction of a geothermal system using a LSTM Neural Network. GRC Trans. 44

Shi, Y., Song, X., Song, G., 2021. Productivity prediction of a multilateral-well geothermal system based on long short-term memory and multi-layer combinational neural network. Appl. Energy, p. 282.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. and Salakhutdinov, R., 2014. Dropout: a simple way to prevent neural networks from overfitting. The journal of machine learning research, 15(1), pp.1929-1958.

Stein, G., Chen, B., Wu, A.S. and Hua, K.A., 2005, March. Decision tree classifier for network intrusion detection with GA-based feature selection. In Proceedings of the 43rd annual Southeast regional conference-Volume 2 (pp. 136-141).

Suzuki, A., Konno, M., Watanabe, K., Inoue, K., Onodera, S., Ishizaki, J., Hashida, T., 2021. Machine learning for input parameter estimation in geothermal reservoir modeling. In: Proceedings of the World Geothermal Congress. Reykjavik, Iceland.

Swain, P.H. and Hauska, H., 1977. The decision tree classifier: Design and potential. IEEE Transactions on Geoscience Electronics, 15(3), pp.142-147.

Temizel C., Canbaz C., Palabiyik Y., Aydin H., Minh T., Ozyurtkan M., Yurukcu M., and Johnson P., 2021. A thorough review of machine learning applications in the Oil and Gas Industry. Society of Petroluem Engineers. SPE-205720-MS

Torlay, L., Perrone-Bertolotti, M., Thomas, E. and Baciu, M., 2017. Machine learning–XGBoost analysis of language networks to classify patients with epilepsy. Brain informatics, 4(3), pp.159-169.

Vesselinov, V.V., Mudunuru, M.K., Ahmmed, B., Karra, S. and Middleton, R.S., 2020. Discovering signatures of hidden geothermal resources based on unsupervised learning.

Williams, G., 2011. Descriptive and predictive analytics. In Data Mining with Rattle and R (pp. 171-177). Springer, New York, NY.

Zhang, X., Zhao, J. and LeCun, Y., 2015. Character-level convolutional networks for text classification. Advances in neural information processing systems, 28.