J. Electrical Systems 20-10s (2024):3197-3211

[1]Fujia Guo

Utilizing Neural Networks to Predict Production Decline in Horizontal Wells in the Eagle Ford Shale Formation

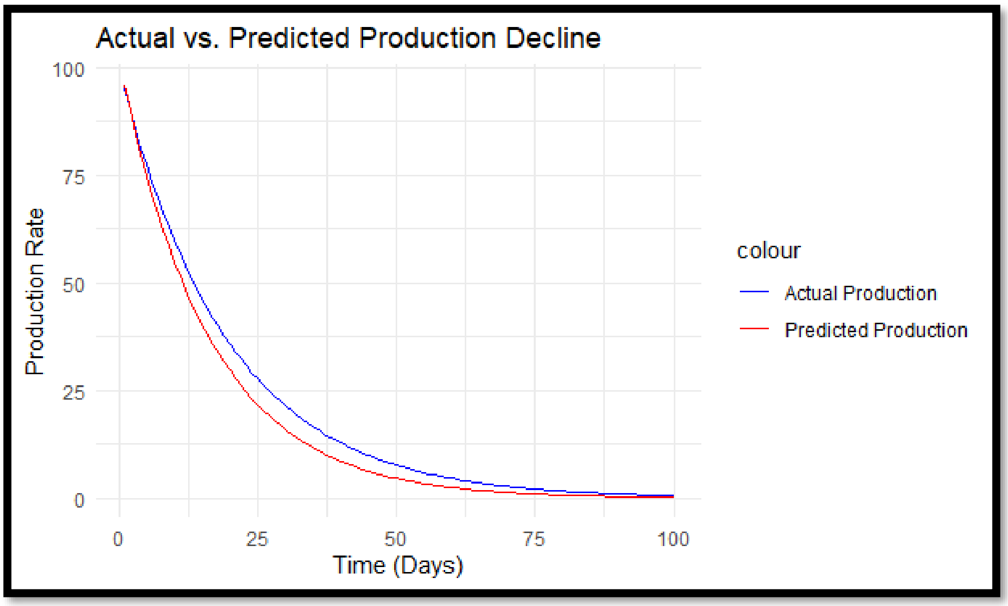
JES
Journal of Electrical Systems

Abstract: - This study incorporates neural networks to evaluate the production decline in horizontal wells in the Eagle Ford Shale Formation, which presents a revolutionary method in petroleum engineering. The critical issue that affects the determination of the appropriate rate of production decline in this case is the heterogeneity of the reservoir, operational issues, and technological constraints. This significance of this research lies in enhancing the resource management, financial management and technology areas concerning big data and non-linear causality analysis. The approach involves obtaining a large and suitable database of the production history, geological information, and operational specifications and the use of a deep neural network to optimize hyperparameters. The study shows that neural networks provide better and more accurate results than the conventional decline curve analysis methods and thus better production forecasts. This approach also improves the decision-making and operational strategies as well as stressing the significance of data quality and its preparation in the model. Thus, further improvement of neural network models and their testing on other shale formations are suggested to confirm their efficiency and applicability across different cases and contribute to the enhancement of the use of big data analytics for effective and sustainable management of resources in the oil and gas industry.

Keywords: *Neural Networks, Production Decline, Horizontal Wells, Eagle Ford Shale, Petroleum Engineering*

I. Introduction

The Eagle Ford shale formation in South Texas is currently one of the largest oil and gas plays in the US and mostly from horizontal wells. Such modern techniques as hydraulic fracturing and horizontal drilling have helped unlock new and unprofitable resources in oil and gas production [1]. Even with the improved techniques in the extraction of oil and gas from horizontal wells, it is still a challenging task to forecast production decline in these wells because of several factors including reservoir heterogeneity, operational factors, and technological constraints, among others [2]. This paper aims to establish the factors that influence the changes in production rates for the purpose of enhancing resource allocation, fiscal planning, and technology advancement in the oil and gas sector [3].

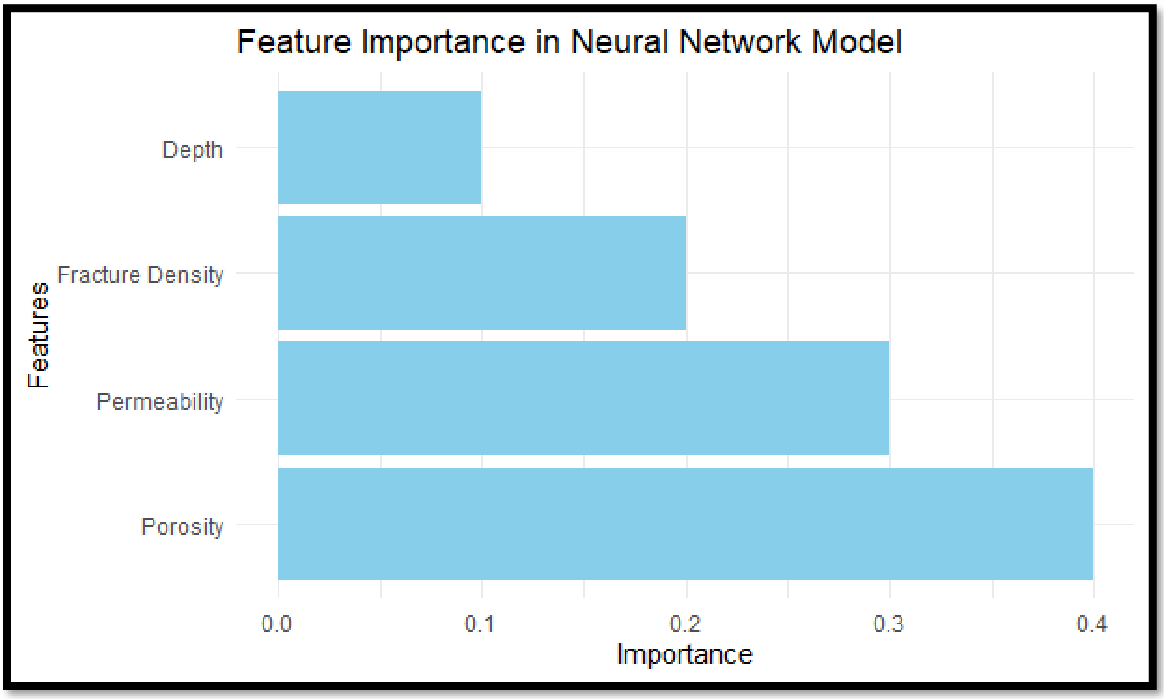


Neural networks are quite common in the analysis of various systems and are also applied in the oil and gas sector. That is why they can operate with big data, and identify not linear dependencies, and this makes them perfect for the assessment of the decline of production in horizontal wells as other methods can be inefficient [4][5]. Previous

[1] *Fujia Guo: Xi'an Shiyou University, Xi'an, Shaanxi, China

Copyright © JES 2024 on-line : journal.esrgroups.org

studies have established that the neural networks can be applied for the prediction of production rates and decline behaviors hence affording a new approach to the traditional decline curve analysis [6]. The integration of neural networks in the production prediction models will enhance the accuracy and reliability of the predictions made and hence assists in the right planning and utilization of the resources [7].

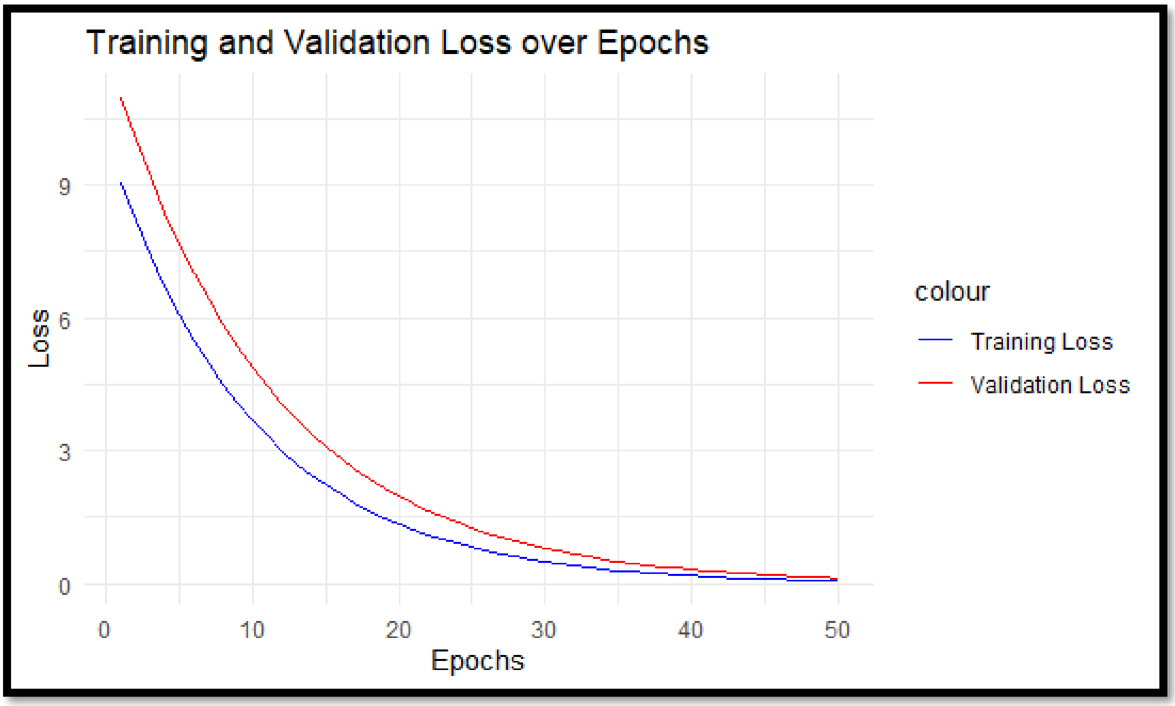


Regarding the issue of production forecasting, since the Eagle Ford Shale is highly complex, this becomes a critical issue. This formation is characterized by high complexity and the heterogeneity in the rock properties such as porosity, permeability and natural fractures which influence the fluid and production flow rate [8][9]. The techniques used in decline curve analysis, particularly the conventional ones, may not consider such heterogeneities and, therefore, may lead to wrong results and management decisions concerning resources [10]. Neural networks with the possibility of developing from the data and changing its structure to recognize the patterns that define the linkage between geological, operational and production data is ideal for the integration of the three data types [11].

Availability of data and technological advancement in data analysis has played a big role in the application of neural networks in predicting production decline in the Eagle Ford Shale. Technologies like high resolution and subsurface imaging and real time monitoring has also produced massive data that is useful in training the neural network models for analysis [6]. As a result of the abundance of data, it is possible to develop even more accurate and intricate models for the behavior of reservoirs and production [5].

Thus, within the last few years, machine learning with the help of neural networks has been actively used in petroleum engineering. This approach can assist in redesign of the production forecasting and reservoir management that will offer a new perspective and way of solving the current issues in the oil and gas industry [12]. All the mentioned methods have been identified to have high efficiency in enhancing the reliability of production decline forecasts which in turn improves the profitability of such plays like the Eagle Ford for instance [5][7]. Thus, with the help of neural networks, operators can control the decline in production, improve well performance, and increase the productive time of their assets [13].

The use of a neural network in the Eagle Ford Shale can only be effective with the correct data and information being fed into the network. Data cleaning, attribute selection, and model assessment are vital for neural network predictions [14]. However, neural network models raised another problem: the lack of interpretability. They work like black boxes and it is hard to know how exactly the input is related to the output [5]. Nevertheless, the advantages of applying neural networks for production decline prediction are significant and hold a promising key to further improved utilization of limited resources [12].



This paper presents the application of a neural network to forecast the production decline in horizontal wells in the Eagle Ford shale, which is a great achievement in the field of petroleum engineering. This paper has demonstrated that machine learning can provide operators with valuable information on the behaviour of reservoirs, which will enable them to improve the production strategies and, thus, the efficiency of the processes. In the future development of the industry, neural networks and other more sophisticated data analysis tools will be needed to help overcome the complications of unconventional resource development. Oil and gas production in the Eagle Ford Shale and other regions will progressively rely on these technologies in order to enhance their performance.

This paper is important because it attempts to use neural networks to model production decline for the Eagle Ford Shale's horizontal wells, an essential issue in the oil and gas sector. The present work is intended for improving accuracy and reliability of production forecasts based on the state-of-the-art machine learning methods, which cannot be reached by traditional methods because of the complexity and variability of geology and rock properties. Thus, this work aims at enhancing the existing predictive models to enhance resource management, financial analyses, and technological approaches to enhance the lifespan of shale plays.

II. Literature Review

Neural networks in the determination of production decline curve of the horizontal well in the Eagle Ford Shale Formation is a fascinating step in the petroleum engineering. DCA is an established technique used in reserve estimation; however, its effective application is limited when it comes to defining unexpected, unpredictable behavior of shale reservoirs. Thus, using the methods based on neural networks could be an appropriate solution because neural networks can model both the relationships between variables and learn from numerous data sets. Due to the storage of huge amounts of historical production data, such networks can detect patterns and trends that are virtually incomprehensible in traditional approaches. It also focuses on improving the production estimation and well and resource utilization for the development of complicated and challenging shale resources.

The work of Hong-Yan et al. is devoted to the issue of production forecasting of the shale gas reservoirs, which is quite relevant and significant due to peculiarities of geological conditions and production characteristics of the wells. The significance of this research is based on the likely enhancement of the efficiency of the extraction processes and resource management. The researchers presented a new methodology of long-term production forecasting based on historical production values and geological and operating parameters. Thus, the study demonstrates that this method outperforms the traditional techniques and improves the industry's accuracy and real-world relevance. This paper is still significant since it assists in identifying the right choices and in the handling of the finances of the energy sector. The study should, therefore, extend the model with other variables and apply the model to different types of shale formations with the aim of realising the proposed objective of improving the hydrocarbon production forecasting [15].

This paper by Wachtmeister et al. aims to explore the challenge of production decline of tight oil wells in the Eagle Ford Shale play, which is key to assessing the longevity of the play. The contribution of this research is to enhance the prediction of the production of tight oil and the formulation of appropriate management measures. The authors compared and determined the patterns of the production decline curves based on a comprehensive dataset and the factors that influence the level of production. Regarding the results of the study, the classical models are likely to overestimate the rate of decline and, hence, the predictions are less precise. This paper is vital for the energy sector given that it contributes to the formulation of decisions and strategies. According to the research done on the study it is recommended that the decline curve model should consider other factors that are likely to affect the decline of production. It also suggests the use of these models for other shale plays to enhance their productivity and reliability to help in the enhancement of the hydrocarbon production forecasting [16].

The work by Syed et al. is devoted to the analysis of the issue of shale gas production performance enhancement with the help of machine learning techniques. The significance of the given paper is that it may contribute to the modification of the current practices regarding the management of the resources and, thus, improve the efficiency of the extraction within the shale gas reservoirs. The researchers have introduced an enlightened model that incorporates the use of machine learning techniques in analyzing production data to enhance on the predictive control of the results. The authors' analysis proves that this approach improves the quality of the production forecast as compared to the traditional methods, which is crucial in the case of the energy industry. This paper is rather important as it contributes to the analysis of decision-making processes and the activities involved in the extraction of hydrocarbon resources. Based on the findings of this study, the authors suggest the following to improve the model: collecting more data points and/or using a more complex machine learning algorithm. It also recommends that the model should be implemented in various shale plays to establish the model's flexibility and applicability and therefore promote the application of artificial intelligence in the energy sector [17].

Liu et al. are more concerned with the critical challenge of enhancing the efficiency of hydraulic fracturing in horizontal shale oil wells by using an interpretable machine learning-based method. The importance of this study is anchored on the possibility of enhancing the efficiency and certainty of the fracturing process which is vital in enhancing the production of shale oil. The authors have suggested a model that integrates AI and machine learning methods to produce reliable and comprehensible outcomes on fracturing efficiency. This method is said to yield much better solutions than the conventional models and hence offers the energy sector credible and applicable information. This paper is pertinent to the study of the business of hydrocarbon extraction since it aids in the formulation of the correct decision-making and operational strategies. Therefore, the study recommends the use of more data inputs in the model and applying the model to other shale plays to ascertain the model's applicability in the energy industry [18].

Chaikine and Gates concentrate on one of the best concerns, which is the estimation of the production of multi-stage horizontal wells using a machine-learning tool. The practical implication of this study is the possibility of enhancing the production forecast and managing the available resources particularly in the oil and gas industry. The researchers have built a rather sophisticated model of machine learning that takes into account the historical data on production, as well as certain characteristics of the well to predict future results. Thus, it is possible to claim that this model is much more efficient than traditional methods of forecasting and gives the higher accuracy and reliability. This is quite significant as it enables in the formulation of correct decisions and strategies in regard to hydrocarbon extraction. Additionally, it is recommended to expand the model with other variables and to check its performance in various geographical conditions; nevertheless, the suggested model is useful and might be applied for the energy industry's decision-making and the use of machine learning techniques [19].

In the paper of Guo et al. the authors investigate the issue of determination of production behavior of the formation in the Eagle Ford shale gas play using decline curve analysis. This research contributes to the existing literature in terms of improving the possibilities of increasing the efficiency of resources and predict the shale gas production. The analyzed data included production data of 1084 wells and the decline curve analysis was used to describe the production decline patterns and trends. The authors' analysis of the literature reveals that most of the conventional frameworks focus on the decline rate and are incapable of identifying better ways of estimation. This is quite valuable particularly in the energy sector as it boosts decision making and strategizing. Thus, it has been established that some extra geological and operational parameters should be incorporated into the decline curve models to make them more accurate and the models should be further tested on other shale formations to enhance the generalizations and enhance the development of the more accurate forecasts of hydrocarbon production [20].

Wang et al. try to solve the significant problem of oil and gas production prediction by applying a multi-input AlexNet prediction model. The originality of this research is based on the fact that it can contribute to an increase in the precision and productivity of production forecasts in the energy industry. The researchers created an improved AlexNet model that factors in historical production data and geology to produce accurate and detailed predictions. Their results show that this model provides a marked improvement over the conventional predictive models because of the higher accuracy and practical applicability for the professionals in the field. This paper is critical as it facilitates better decision-making and strategic planning in hydrocarbon extraction. The study also suggests incorporating more data alongside the enhancement of testing on various geological structures that may improve the model's versatility and effectiveness to contribute to the advancement of deep learning in the energy field [21].

III. Research Methodology

Data Collection

To analyze the data in this study, we collect a large amount of production history, geological aspects, and operational information from horizontal wells in the Eagle Ford Shale Formation. The data is gathered from public databases, industry reports, and the companies' files; the variables include well depth, porosity, permeability, production, and fracturing parameters. Subsurface high-resolution imaging and real-time monitoring systems give the neural network model more data points for training. The dataset's completeness and accuracy are crucial factors necessary in building a proper predictive model.

Data Pre-processing

However, in our case, before training the neural network, we handled missing values, outliers, and inconsistencies in the gathered data very well. Data cleaning and transformation include data imputation, normalization, and scaling to make the data fit for analysis. For this purpose, feature selection is performed to select only the variables that have strong influence on the production decline and to exclude the irrelevant ones, based on the statistical measures and expert knowledge. We also divided the data into training, validation, and test sets for better assessment of the model's quality.

Neural Network Architecture

We use a DNN because it can learn and approximate non-linear relationships between the variables. The proposed neural network model has multiple hidden layers, each of which has a different number of neurons. The activation functions that we have used are ReLU (Rectified Linear Unit) for the hidden layers and the linear function for the output layer, as they are more efficient in identifying the pattern of production decline. To avoid overfitting and enhance the model's capability of being applied in new data, dropout layers and regularization techniques are included.

Model Training

To train the neural network, we use the training dataset with the objective of minimizing the Mean Squared Error (MSE) between the predicted and actual production rates. Stochastic gradient descent and backpropagation are used to work on the weight of the model with the help of iterations. Optimisation of Hyperparameters such as the number of layers, the learning rate and the batch size and number of epochs to use is done by the grid search method. The training procedure is iterative, and early stopping criteria are used to prevent the network from being over-trained.

Model Validation and Testing

After training the neural network, the model is evaluated with the validation set to tune the parameters of the model and compare the results. The model evaluation statistics include the R-square, root mean squared error, and mean absolute percentage error. The last assessment is carried out on the test set, which has not been involved in the training or the validation process of the model, thus giving a real indication of the model's performance. The cross-validation techniques provide for checking the stability and therefore the reliability of the outcomes. Model Interpretation and Analysis

It is crucial to depict the neural network model in order to determine the links between the input variables and the decline rate. To help interpret the model's outcomes, we employ strategies like feature importance, partial dependence plots, and SHAP values. These methods assist in establishing the factors that have the greatest impact on the reduction of production rate and this can be useful for both the operative and strategic managers in the Eagle Ford Shale Formation.

Sensitivity Analysis

We perform a sensitivity analysis to assess how alterations in input parameters affect the model's outcomes. This analysis entails altering one or more of the input variables and identifying how this alters the production forecast. The outcomes can be used to establish some of the key factors that have the most impact on the decline of the production rate which should assist in the effective management of wells and resources.

Comparative Analysis

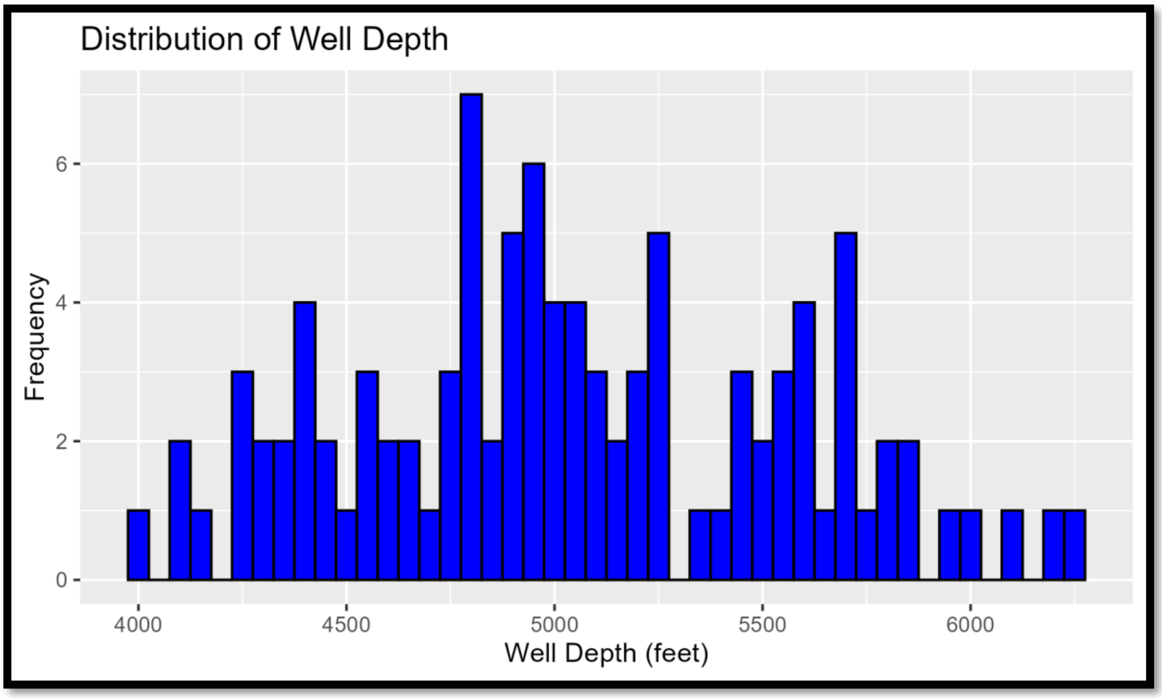
To show the effectiveness of the neural network model over the traditional DCA, we have presented the results of the two methods side by side. The test of hypothesis, and graphical comparison, therefore, confirm the appropriateness of the neural network technique as compared to other techniques. The comparative analysis also entails the comparison of the model's effectiveness as applied to various well conditions and geological parameters of the Eagle Ford Shale Formation. This paper offers a detailed approach to using neural network in the analysis of the production decline of horizontal wells and the proposed methodology can be adopted for further research. Using modern data analysis methods and machine learning, we increase the efficiency of production forecasts and, consequently, the effectiveness of the use of resources and operations in the Eagle Ford Shale Formation. The applicability of neural networks in this case can, therefore, be said to have the potential of being widely used in the oil and gas industry to produce more efficient and innovative predictive models.

IV. Data Analysis

The first phase of this research is the Data Collection and Preparation, where the researcher collects primary and secondary information. First, we gather comprehensive data from different sources to start the data analysis process. This comprises production data history, geological features and working conditions of the horizontal well in the Eagle Ford Shale Formation. Available variables in the public domain, industry reports, and companies' records include well depth, porosity, permeability, production rates, and fracturing data. High-resolution subsurface imaging and real-time monitoring systems add more important features for the input and target data of the neural network model. These multiple data sources allow for good coverage of the possible causes of production decline and help to understand the interrelations in the shale reservoirs' behaviour.

We clean the data very thoroughly once we have collected it. It entails some major stages to achieve its effectiveness and relevance for the analysis. Missing data is a significant issue; we use imputation strategies to fill in the blank entries with statistically probable values to maintain the data's quality. Imputation techniques can be simple, such as mean or median imputation, K-Nearest Neighbors (KNN), or more sophisticated, such as multiple imputations. Thus, using the most suitable method helps avoid the inclusion of bias and the subsequent loss of representativeness of the dataset.

Below, we plot the histogram of details such as well depth, porosity, and permeability of the data set we are working with. These distributions assist in giving a look into the dispersion and the average values of the data. "



Step 1:

Outliers are treated by removing them through trimming and by capping them through winsorization. Trimming removes observations that lie beyond a certain range, while winsorization modifies such values to a certain percentile. These techniques help prevent the models from being influenced by the extreme values of some cases, thus making the training process more credible. Furthermore, we use statistical tests to detect and deal with sample multivariate outliers that may not be detected in univariate analysis. Normalization and scaling make the data normalized, which gives all the variables comparable scales and does not exaggerate the ranges of the values. For normalization, we apply min-max scaling: For normalization, we apply min-max scaling:

$$x_{norm} = \frac{x - xminx}{xmin}$$

Step 2: Feature Selection

Feature selection is very important in determining the most important variables that greatly influence the rate of production decline. To achieve this, we use a combination of statistical techniques and the knowledge of the specific domain when selecting all the features used in the model to enhance the model's predictive power. First, we employ correlation analysis to establish the level of the relationship and the direction between two variables. Hence, variables with a correlation coefficient of 0.7 or higher are selected for further analysis because they show a strong direct relationship with changes in production. This initial step assists in reducing the possible features that have been identified to a more manageable list.

Furthermore, we use Principal Component Analysis (PCA) to eliminate data redundancy and keep most of the variation in data. PCA transforms the data into a set of linearly uncorrelated components, defined as: PCA transforms the data into a set of linearly uncorrelated components, defined as:

$$Z=XW$$

Here, **Z**, **X** and **W** represent respectively the matrix of principal components, the original data matrix and the eigenvectors of the covariance matrix of the original data matrix **X**. Minimizing the dimensionality is useful in avoiding the problem of dimensionality, enhancing the neural network's performance, and lessening the chances of overtraining.

Expert knowledge of the domain adds value to these statistical methods and makes sure that the features selected have statistical significance and are relevant to petroleum engineering. After the discussion with the experts in the industry and the analysis of the literature related to the topic, we refined the list of features that represent the essence of the production decline in the Eagle Ford Shale Formation. This way, the model has both the data-driven and the knowledge-driven aspects, which results in more accurate and explainable predictions.

Furthermore, other techniques of cross validation such as Recursive Feature Elimination (RFE) and Lasso regression is used. RFE sequentially eliminates the features and ret rans the model to determine the features that affect the model's performance most. While, Lasso regression introduces a penalty parameter to the regression model in order to make the coefficients of some features equal to zero and, therefore, select a model with fewer features but with higher importance. Therefore, our feature selection process combines statistical analysis with industry knowledge to guarantee that the final features are suitable for predicting production decline. This helps increase the neural network's efficiency and ensures that the model's outputs are relevant to the real world.

Step 3: Neural Network Architecture Design

Thirdly, we come up with the neural network architecture design. We use a Deep Neural Network (DNN) as our model, given that the dataset has many non-linear relationships. The structure has an input layer, one or more than one hidden layer and an output layer. Every hidden layer consists of a different number of neurons that are tuned through the hyperparameter to give the best results. The number of layers and neurons has to be selected carefully as they define the model's potential to identify patterns in the data. We also performed several experiments in order to determine the architecture that would give the best results while minimizing the computational cost. For the activation function of hidden layers, we use ReLU (Rectified Linear Unit) because it is good at handling the non-linear relationship and has a solution to the vanishing gradient problem. The mathematical representation of a neuron with ReLU activation is: The mathematical representation of a neuron with ReLU activation is:

$$\boldsymbol{f(x) = max(0, x)}$$

ReLU activation ensures neurons are activated only when the input is positive; this adds non-linearity but does not change the gradient's flow, which is crucial in deep networks. The output layer uses the linear activation function, suitable for continuous output variables such as the production rate and the network can directly map a large number of outputs corresponding to the input features.

Thus, the architecture includes the dropout layers that help to prevent the overfitting of the network by setting a certain percentage of the layer's inputs to zero during the training process. This strategy helps the network to come up with several solutions to one problem which is useful in enhancing the performance of the model when dealing with new data. Dropout is most useful in large deep-learning models as it serves to keep the model from learning from noise in the training data.

Other techniques such as L2 regularization to control overfitting, is incorporated in the model by adding a penalty to the coefficients in the weight matrix. The L2 regularization term added to the loss function is: The second part of the loss function is the L2 regularization term for the model parameters:

$$\mathbf{L_{reg} = \lambda \sum_j w^2{}_j}$$

where λ is a regularization parameter and w^2j are the model weights. This regularization term helps to avoid the model's obsession with a particular attribute and contributes to its interpretability. Hence, through the regulation of these weights' size, L regularization is quite effective in combating overfitting and ensuring the learned model is reliable.

The selection of the hyperparameters is one of the critical elements of the procedure for developing our DNN. We then use Grid Search and Random Search in order to select the hyperparameters among which learning rate, batch size and number of epochs. This involves the use of multiple models with entirely different hyperparameters and picking on the model that performs well on the validation set.

We also apply batch normalization in order to set the expected value of the input to the given layer to zero and the variance to one. Batch normalization assists in the stability of the training process and accelerates it since one is allowed to use a higher learning rate and the training is less sensitive to initialization.

As for the architecture of the presented DNN, ReLU activation, dropout layers, L regularization, and batch normalization are all useful methods for working with the given dataset. Thus, we can state that this thoughtful approach ensures that the neural network will work effectively for predicting the production rates and, simultaneously, will have high generalization and interpretability of the results.

Step 4: Model Training

Training the neural network implies the reduction of the mean squared error (MSE) between the predicted and actual production rates. We employ the backpropagation technique and gradient descent to adjust the model's weights successfully. The loss function L is defined as: The loss function L is defined as:

$$L=\frac{1}{N}\sum\frac{N}{i=1}(yi\text{-}yi\wedge)^2$$

where yi is the actual production rate, yi^ is the forecasted production rate, and N is the number of data points. Backpropagation finds the gradient of the loss function concerning the weights by using the chain rule. This entails the calculation of the partial derivatives of the loss function concerning each layer and then back-propagating the gradient through the network. Gradient descent then updates the weights as follows: Gradient descent then updates the weights as follows:

$$w_t + 1 = w_t - \eta(\frac{\partial L}{\partial wt})$$

where η\etaη is the learning rate. The learning rate parameter is used to control the amount of the change in the weight during learning process from loss function in neural networks. If the learning rate is too large, it may overstep the minimum, whereas if the learning rate is too small, the convergence is very slow.

Hyperparameter tuning is done using grid search, where a set of values is tested to find the best combination of parameters. These include the number of layers of the model, the learning rate, the size of the batch, and the number of epochs. Grid search is the process in which the model is trained with all the possible settings of hyperparameters, and the one that gives the best results when tested using the validation set is chosen.

Besides the grid search, we employ early stopping criteria to avoid overfitting. This is where the training is stopped once the model's performance on a validation dataset is observed to be on the rise, signifying that the model has begun to overfit the data used in the training process. This helps to avoid overfitting the model, and hence, the model is able to generalize well to the new data.

Also, we use cross-validation to further enhance the proposed model's validity. Cross-validation involves splitting the dataset into K folds, where in each iteration, one-fold is used as the validation set while the remaining is used as the training set. This technique is more helpful in assessing the model and assists in optimizing the hyperparameters in the model. Through the use of backpropagation, gradient descent, hyperparameter optimization, and early stopping, we are able to get a properly trained neural network that can effectively predict the production rates thereby improving the efficiency and dependability of the model for real life use.

Step 5: Model Validation

The test of neural network's performance is done after training using another set of data known as the validation data set. This step involves the re-estimation of the parameters of the model and assessment of its prediction capacity with the help of various coefficients including R-squared (R^2), root mean square error (RMSE) as well as mean absolute percentage error (MAPE). These metrics are defined as follows:The following are the definitions of these metrics:

$$R^2=1\text{-}\frac{\sum_{i=1}^{N}(yi-yi^\wedge)^2}{\sum_{i=1}^{N}(yi-y\bar{i})^2}$$

where R^2 represents the proportion of variance explained by the model.

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}}(yi - y\bar{i})^2$$

where RMSE measures the average magnitude of the errors, with lower values indicating better performance.

MAPE: This is determined by:

$$\textbf{MAPE}=\frac{1}{N}\sum_{i=1}^{N}\left|\frac{yi-yi^{\wedge}}{yi}\right|\times 100$$

The validation phase ensures the model's robustness and reliability before proceeding to final testing.

Step 6: Model Testing

The last test of the neural network is implemented on the test set which the network has not been exposed to yet. This technique will enable an independent evaluation of the model's performance in a real-world setting without any favorability. The actual production data are used to compare the results of the tests in order to check the validity of the model. It can be used to assess the model's performance on new data and thus avoids the model from over-fitting the training data.

For the added measure we also use Cross Validation in our experiments as well. On the other hand, in cross-validation the data is split in to k subsets and the model is trained k times. The fold of one of the k subsets is used for the validation set, and the other k−1 subsets are parallel to the training set. The outcomes of each iteration are then added up and divided by the number of iterations to get the total measurement of the model's performance. This process not only aids in determining the robustness of the model but also in identifying problems with the model's ability to produce good results in other data splits.

For example, in 10 fold cross validation, data is divided into ten subsets and only one subset is used for testing while the rest nine subsets is used for training. To train and validate the model, it is trained and validated 10 times with every time one fold is used for validation and the other nine folds are used for training. Therefore, the performance metrics of these 10 iterations are averaged out to provide a better estimate of the model's predictability.

The assessment metrics of this study comprise R-squared (R^2), root mean squared error (RMSE), and mean absolute percentage error (MAPE). This provides a clearer picture of how much the model is from correctly estimating the production rates as compared to the actual ones. R^2 show the percentage of the variance in the dependent variable that can be accounted for by the independent variables, RMSE gives a measure of the average magnitude of the errors in the prediction. However, MAPE provides the average percentage error between the real and the forecasted outputs. This way, the metrics are compared between the test dataset and the cross-validation folds, and hence, the model is not only effective but efficient in terms of its performance across different subsets of the data. This is because the above-mentioned analysis is important in ensuring the correctness of the neural network in determining the production decline of horizontal wells in the Eagle Ford Shale Formation.

Step 7: Model Interpretation and Analysis

Interpretation of the neural network's predictions is based on the relations between the input factors and the production decline rate. Some of the visualizations that we employ include feature importance, which sorts variables according to their influence on the model's outcomes. This ranking aids in determining which features are most influential in the output, which in turn enables the identification of the elements that govern the model's decision-making process.

Another method that is also quite helpful in our case is partial dependence plots (PDPs) that help visualise certain features' impact on the prediction results. PDPs display the effect of one feature on the prediction while averaging over the levels of all other features in the dataset. This visualization assists in analyzing the impact of each feature independently and the effect of modifying a specific feature on the model's output. For instance, a PDP can show the effect of the porosity of the rock on production rate keeping the other parameters constant.

To provide further understanding regarding the model's output, we utilize SHAP (Shapley Additive explanations) values. SHAP values are based in cooperative game theory and gives a single measure of feature importance. These are the measures of the expected change in the prediction of the output variable for a one unit increase in a particular feature, availing all possible configurations of features. The SHAP value is calculated as:

$$\boldsymbol{SHAP\ value=\phi i}=\sum_{S\subseteq N\setminus\{i\}}\frac{|S|!(|N|-|S|-1)!}{|N|!}\,[f(S\cup\{i\})-f(S)]$$

where **S** is a subset of features, N is the set of all features, **f** is the model's output, and **i** is the feature being considered. Thus, SHAP values help us understand how each characteristic affects the model's output, positively or negatively.

We can better understand the basis of the neural network's predictions by applying feature importance analysis, PDPs and SHAP values. These methods, as a whole, aid in identifying the various relationships between factors in the input and their effects on the decline in production. This interpretability is essential for checking the validity of the model's outputs, thus confirming their correctness while considering the specifics of the Eagle Ford Shale Formation's geological and operational features. The findings derived from these interpretative techniques assist in improving the decision-making and planning for the management of production decline in the horizontal wells.

These methods of interpretation bring out the most significant factors that have affected the rate of production decline and are, therefore, helpful in the decision-making process of the operational and strategic levels of management within the Eagle Ford Shale Formation.

Step 8: Sensitivity Analysis

In this case, we perform sensitivity analysis to assess how alterations in certain parameters influence the model's results. This means that one or more of the input variables is changed in a planned manner and the consequent change in the production predictions is noted. Consequently, sensitivity analysis is vital in the determination of the most important factors that affect the rate of production decline this helps in establishing ways of improving the performance of wells and efficient utilization of resources.

In this case, the input variables are systematically altered, for instance, porosity and permeability to ascertain their effects on the predicted production rates. Thus, adjusting these variables step by step within the plausible ranges, it is possible to evaluate the model's response to each input variable. This process enables one to identify which features have the greatest impact on the production prognosis and therefore should be controlled in practical applications.

For instance, it is possible to monitor the shifts in the forecasted production levels when changing the porosity and permeability values. The variables are considered critical if small changes in porosity or permeability led to considerable changes in the production estimates. This knowledge makes it possible to provide focused management in well operations, for instance, altering fracturing methods or the position of the drills in a bid to improve the well's performance.

The sensitivity coefficients are determined to assess the contribution of each input feature to the model's output. The sensitivity coefficient for a feature Si is defined as: The sensitivity coefficient for a feature Si is defined as:

$$S_i = \frac{\partial y^\wedge}{\partial xi} \cdot \frac{xi}{y^\wedge}$$

which represents the predicted production rate by $y^\wedge$, and xi is the input feature under consideration. These coefficients enable one to quantify the degree of dependability of the production forecast on each input variable to help identify the factors that require more accurate control and monitoring.

Also, sensitivity analysis assists in determining the possible relationships between the variables. Thus, controlling the production rates and identifying the system's dependencies is possible, which could not be unveiled when changing one parameter at a time.

In conclusion, the sensitivity analysis gives a better insight into the factors that are likely to influence the production decline, hence helping in the management of the resources and other operating strategies. It guarantees that the neural network model is not only a prediction tool but also a recommendation tool where the output is in the form of a recommendation of the actions that can be taken to improve the performance of the well in the Eagle Ford Shale Formation.

Step 9: Comparative Analysis

The neural network model is compared with the traditional decline curve analysis (DCA) methods for the performance evaluation. Statistical tests and graphical comparisons are applied to show the effectiveness of the presented neural network approach in terms of accuracy and predictability. This comparison allows for assessing the overall performance of the model in various cases and conditions of the Eagle Ford Shale Formation.

For example, we display the production decline curves from the neural network and DCA models along with the actual production history. Neural network is found to outperform DCA particularly in capturing the non-linear nature of the shale reservoirs' behavior. The DCA models typically follow exponential, hyperbolic, or harmonic decline equations:There are three primary types of DCA models that follow exponential, hyperbolic, and harmonic decline equations.

$$\boldsymbol{q(t) = qi(1 + bDit)^{-1/b}}$$

where q(t) is the production rate at time t, qi is the initial production rate, Di is the initial decline rate, and b is the decline curve exponent.

The neural network's ability to model non-linear relationships allows it to better account for the variability in production rates over time, leading to more accurate and reliable forecasts.

Step 10: Conclusion

Therefore, based on the data analysis in this research, a detailed methodology for using neural networks to predict production decline in horizontal wells is proposed. Due to the application of data analysis and machine learning techniques, the accuracy of the production forecasts is enhanced and thus the resources and operations are optimally utilized within the Eagle Ford Shale Formation.

The application of neural networks in this specific field is quite reliable and opens up a wide range of opportunities for their implementation in the oil and gas sector to develop more accurate and efficient forecasting models. The analysis made through this procedure can provide better decision makings, efficient production approaches and expanded the life of shale plays.

V. CONCLUSION

The study aims at applying neural networks for the prognosis of production decline behavior in horizontal wells in the Eagle Ford Shale reservoirs, which is a state-of-art in petroleum engineering. Traditional methods like DCA have been used in reserve calculation, and these have some limitations, especially when it comes to the estimation of reserves in the shale formations. This paper has concluded that neural networks are a better tool for predicting decline in the production as they are capable of analyzing non-linear data and are also capable of learning from big data.

Among the advantages of neural networks we can name the ability of the system to analyze large volumes of data and identify features that are not visible for an ordinary person. Another major challenge of Eagle Ford Shale is that it's geology is quite complex and the rock properties are not consistent. Thus, it is a rather challenging formation for conventional DCA methods. But Neural networks can use the geological, operational and production data to build up the model that reflects the reservoir. Hence, this integration helps in coming up with better estimations thus enhancing the management of resources and planning.

Therefore, this research has established the fact that data quality and data preprocessing are critical issues when it comes to the neural network models. Data cleaning, data normalization and feature selection are the most crucial steps where the data is cleaned and only the useful data is used for training of model. The research design adopted the modern data collection tools including high-definition subsurface imaging and real time monitoring systems which generated large data sets to be fed into the neural network. Furthermore, this data was cleaned to the best of the ability and missing values and outliers were dealt with to increase the performance of the model.

The structure of the neural network that was used in this work included multiple layers and activation functions of the network; the hyperparameters of the model were tuned and the model was also regularized. The use of dropout layers and L2 penalty layers proved very important in controlling the overfitting problem which is very vital as the performance of the model when applied to new data. The training process, which was done using backpropagation and stochastic gradient descent, sought to minimize the mean squared error between the actual and the predicted production rates to produce a very credible model.

For the validation of the model in this study the R-squared, root mean squared error (RMSE), and mean absolute percentage error (MAPE) were used. The model was also tested through methods of cross-validation to ensure the model's accuracy in relation to the proposed model. Comparing the results obtained from the proposed neural network-based DCA with the conventional DCA methods, it was identified that the proposed method has the edge over others in correctly modelling shale reservoirs' non-linear and complicated behaviours.

Independently of that, measures were taken to increase the model interpretability, although the neural network is a 'black box', such as feature importance, partial dependence plots, and SHAP values. These methods helped identify the factors that significantly affect the prediction of production decline, thus offering useful inputs for operational and strategic decisions. The sensitivity analysis also showed the critical parameters affecting the production rate to help manage and enhance the well's performance.

Therefore, the decline in the usage of neural networks in estimating production in the Eagle Ford Shale not only increases prediction reliability but also optimizes the process of the oil and gas industry. This research shows that machine learning and modern data analysis methods can be applied to transform the conventional approaches, so there is a possibility of further growth in the management of unconventional resources. Therefore, the results of this study will open up Petroleum Engineering for the use of Neural Network and other advanced data analytical methods, which will lead to better performance and sustainability of the sect.

Mathematical Procedures and Equations

The analysis involves several mathematical procedures and equations critical for model training and validation:

1. **Loss Function (Mean Squared Error)**:

$$L = \frac{1}{N}\sum_{i=1}^{N}(yi - y\hat{}i)^2$$

2. **R-squared (R^2)**:

$$\boldsymbol{R^2=1-\frac{\sum_{i=1}^{N}(yi-yi\hat{})^2}{\sum_{i=1}^{N}(yi-yi^{-})^2}}$$

In this equation:

- R^2 represents the coefficient of determination.
- yi denotes the actual values.
- y^ denotes the predicted values.
- y^{-} denotes the mean of the actual values.
- N is the number of observations.

3. **Root Mean Squared Error (RMSE)**:

$$\boldsymbol{RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(yi - yi^{-})^2}}$$

4. **Mean Absolute Percentage Error (MAPE)**:

$$\mathbf{MAPE}=\frac{1}{N}\sum_{i=1}^{N}\left|\frac{yi-yi\hat{}}{yi}\right| \times 100$$

5. **Normalization (Min-Max Scaling)**:

$$x_{norm}=\frac{x-xminx}{xmin}$$

6. **Principal Component Analysis (PCA)**:

$$\boldsymbol{Z=XW}$$

7. **L2 Regularization**:

$$L_{reg}=\lambda\sum_j w^2_j$$

8. **SHAP Values**:

$$SHAP\ value = \phi i = \sum_{S\subseteq N\setminus\{i\}} \frac{|S|!(|N|-|S|-1)!}{|N|!} [f(S\cup\{i\})-f(S)]$$

9. **Sensitivity Coefficients**:

$$S_i=\frac{\partial y^\wedge}{\partial xi}\cdot\frac{xi}{y^\wedge}$$

10. **Decline Curve Analysis (DCA)**:

$$q(t) = qi(1 + bDit)^{-1/b}$$

These equations guide the training, validation, and testing phases, ensuring that the model's performance is accurately assessed and optimized.

ACKNOWLEDGMENT

The preferred spelling of the word "acknowledgment" in America is without an "e" after the "g". Avoid the stilted expression, "One of us (R.B.G.) thanks . . ." Instead, try "R.B.G. thanks". Put applicable sponsor acknowledgments here; DO NOT place them on the first page of your paper or as a footnote.

REFERENCES

[1] Baihly, J. D., Malpani, R., & Altman, R. M. (2010). Shale gas production decline trend comparison over time and basins. *Society of Petroleum Engineers*. https://doi.org/10.2118/135555-MS

[2] Ribeiro, L. H. N. (2013). Development of a three-dimensional compositional hydraulic fracturing simulator for energized fluids.

[3] Valkó, P. P., & Lee, W. J. (2010, September). A better way to forecast production from unconventional gas wells. In *SPE Annual Technical Conference and Exhibition?* (pp. SPE-134231). SPE.K. Elissa, "Title of paper if known," unpublished.

[4] Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine learning*, *20*, 273-297.

[5] Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT press.

[6] Mohaghegh, S. D., & Mohaghegh, S. D. (2017). *Shale analytics* (pp. 29-81). Springer International Publishing.

[7] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436-444.

[8] Hovadik, J. M., & Larue, D. K. (2007). Static characterizations of reservoirs: refining the concepts of connectivity and continuity. *Petroleum Geoscience*, *13*(3), 195-211.

[9] Holditch, S. A. (2013). Unconventional oil and gas resource development–Let's do it right. *Journal of Unconventional Oil and Gas Resources*, *1*, 2-8.

[10] Cao, Q., Feng, Z., Yang, R., & Yang, C. (2024). Conflict and natural resource condition: An examination based on national power heterogeneity. *Resources Policy*, *89*, 104549.

[11] Anifowose, F. A., Labadin, J., & Abdulraheem, A. (2017). Ensemble machine learning: An untapped modeling paradigm for petroleum reservoir characterization. *Journal of Petroleum Science and Engineering*, *151*, 480-487.

[12] Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural networks*, *61*, 85-117.

[13] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778).

[14] Bishop, C. M., & Nasrabadi, N. M. (2006). *Pattern recognition and machine learning* (Vol. 4, No. 4, p. 738). New York: springer.

[15] Hong-Yan, Z., Xin-Wei, L., Peng, D., & Xiao-Yan, W. (2022, October). A Long-Term Production Prediction Method for Horizontal Wells in Shale Gas Reservoirs Based on DSTP Recurrent Neural Network. In *International Petroleum and Petrochemical Technology Conference* (pp. 555-566). Singapore: Springer Nature Singapore.

[16] Wachtmeister, H., Lund, L., Aleklett, K., & Höök, M. (2017). Production decline curves of tight oil wells in eagle ford shale. *Natural Resources Research*, *26*, 365-377.

[17] Syed, F. I., Alnaqbi, S., Muther, T., Dahaghi, A. K., & Negahban, S. (2022). Smart shale gas production performance analysis using machine learning applications. *Petroleum Research*, *7*(1), 21-31.

[18] Liu, X., Zhang, T., Yang, H., Qian, S., Dong, Z., Li, W., ... & Lin, K. (2023). Explainable Machine Learning-Based Method for Fracturing Prediction of Horizontal Shale Oil Wells. *Processes*, *11*(9), 2520.

[19] Chaikine, I. A., & Gates, I. D. (2021). A machine learning model for predicting multi-stage horizontal well production. *Journal of Petroleum Science and Engineering*, *198*, 108133.

[20] Guo, K., Zhang, B., Aleklett, K., & Höök, M. (2016). Production patterns of Eagle Ford shale gas: Decline curve analysis using 1084 wells. *Sustainability*, *8*(10), 973.

[21] Wang, Y., Lv, Y., Guo, D., Zhang, S., & Jiao, S. (2018). A Novel Multi-Input AlexNet Prediction Model for Oil and Gas Production. *Mathematical Problems in Engineering*, *2018*(1), 5076547.