IMPLEMENTATION OF REGRESSION ALGORITHMS FOR OIL RECOVERY PREDICTION

This paper presents the work of predicting oil production using machine learning methods. As a machine learning method, a multiple linear regression algorithm with polynomial properties was implemented. Regression algorithms are suitable and workable methods for predicting oil production based on a data-driven approach. The synthetic dataset was obtained using the Buckley-Leverett mathematical model, which is used to calculate hydrodynamics and determine the saturation distribution in oil production problems. Various combinations of parameters of the oil production problem were chosen, where porosity, viscosity of the oil phase and absolute permeability of the rock were taken as input parameters for machine learning. And the value of the oil recovery factor was chosen as the output parameter. More than 400 thousand synthetic data were used to test multiple regression algorithms. To estimate the quality of regression algorithms, the mean square error metrics and the coefficient of determination were used. It was found that linear regression does not cover all patterns in the data due to underfitting. Various degrees of polynomial regression were deployed and tested, and it was also found that for our synthetic data, the quadratic polynomial model trains quite well and perfectly predicts the value of the oil recovery factor. To solve the overfitting problem, L1 regularization known as the Lasso regression method was applied. For the quadratic polynomial regression model, the coefficient of determination was 0.96, which is a pretty good result for the test data. Thus, it is assumed that the data-driven machine learning methods discussed in the paper can be useful for predicting the oil recovery factor using practical data from oil fields at the stages of production

Keywords: machine learning, polynomial regression method, enhanced oil recovery, lasso regularization

Yerzhan Kenzhebek
Corresponding author
Master of Science*
E-mail: kenzhebekyerzhan@gmail.com
Timur Imankulov
PhD*
Darkhan Akhmed-Zaki
PhD, Rector
Astana IT University
Mangilik Yel ave., EXPO Business Center,
Block C.1, Nur-Sultan,
Republic of Kazakhstan, 010000
Beimbet Daribayev
PhD, Head of Department
*Department of Computer Science**
**Al-Farabi Kazakh National University
Al-Farabi ave., 71, Almaty,
Republic of Kazakhstan, 050040

1. Introduction

Recently, big data processing methods and machine learning algorithms have been widely used in all branches of science and industry. Oil company researchers are trying to find new solutions to improve the efficiency of oil production. Many fields have modern automated control systems from which a huge amount of data is received (various parameters of oil production). This amount of data cannot be analyzed and evaluated manually. Therefore, all over the world there is a question of using modern methods to improve the efficiency of oil production, such as machine learning and big data processing methods. Such approaches can significantly increase the recovery factor with minimal costs, optimize and reduce the cost of the technological process. From an economic point of view, most investments in oil fields are spent on obtaining as much information about the reservoir as possible. Therefore, the estimation of the expected oil production is an important aspect for planning the further development of oil fields. In this regard, the oil industry is in great demand for approaches related to the

processing of big data, machine learning approaches, and the development of artificial intelligence algorithms.

Currently, the application of different methods of machine learning in the oil and gas industry is becoming relevant. The data-driven approach makes it possible to build excellent oil prediction models to increase oil recovery. There is a lot of research related to increasing oil production using machine learning methods. In [1], the authors found out that the application of machine learning (ML) algorithms may turn out to be more productive in comparison with traditional calculations on a regular grid. The researchers of [2] described an approach to creating a proxy model based on machine learning methods, in particular, the random forest method was used. The authors reviewed two synthetic examples that used a reservoir simulation model to represent the true reservoir to generate production data for history matching and predict future reservoir performance.

The authors of [3] investigated the application of artificial intelligence methods to predict the performance evaluation of a polymer flooding operation. The proposed model of

the authors was tested using MLP, RBF and ANFIS neural networks. The authors of this study chose the main parameters of polymer flooding as input parameters of the neural network. The only MLP output parameter was oil recovery factor via polymer flooding. Thus, the authors of this work found that such parameters of polymer flooding as API gravity, salinity, permeability, porosity and salt concentration have the greatest effect on the characteristics of polymer flooding.

The study [4] considered the use of artificial neural network (ANN) to predict oil recovery and CO_2 storage capacity in the ROZs. The training dataset in this study was generated using geological factors and well operations. The dataset was collected from 351 numerical simulations jobs for the spatiotemporal database. The proposed data-driven model was applied to five ROZs fields in the Permian Basin as a real field application. The authors found that their ANN models provide excellent oil recovery predictions that are in excellent agreement with reports from [5].

The application of ML methods for predicting the oil recovery factor is a necessary part of field development planning. Therefore, research on the development of oil production using various effective methods of machine learning in the oil and gas industry is relevant.

2. Literature review and problem statement

The authors of the study [6] presented the application of machine learning in enhanced oil recovery (EOR) screening. A large database has been compiled from various surveys including over 1,000 experiences from worldwide enhanced oil recovery projects. The authors reviewed a variety of machine learning methods and deep artificial neural networks to predict the appropriate category of candidate enhanced oil recovery methods, where RF Deep ANN models performed best with an average accuracy of 90 %. The authors of the paper concluded that ML provides a very good indication of the primary screening for EOR, but this should not be considered the only predictive method. The analysis of this work shows that the screening of enhanced oil recovery methods using machine learning methods has several problems: lack of sufficient data for generalized learning, unbalanced noisy data, and insufficient input characteristics. These parameters can significantly affect the results.

The authors of [7] have defined the performance of oil wells using an artificial neural network. In this study, the authors applied ANN and Least Squares Support Vector Machines (LSSVM) to calculate shape-related skin factor or pseudoskin factor in horizontal wells. The authors found that the developed LSSVM approach gives the closest match to real data among the techniques of artificial neural networks. It is also suggested that the model proposed in this paper can reach petroleum engineers determine the optimal well location using the reverse engineering concept. The authors mentioned the efficiency of horizontal wells from the point of view of technical and economic perspectives, however, the authors found that the common studies conducted to accurately calculate this parameter have some inherent constraints and drew attention to numerical intelligent models.

The study [8] presented a data-driven methodology for estimating the oil recovery factor using reservoir parameters and statistics. The authors reviewed two datasets: 56 parameters from the TORIS dataset that contains a description of 1,381 USA oil fields and 199 parameters from a dataset

provided by a Russian oil company. The authors tested and analyzed the GradientBoosting and Random Forest models to estimate the expected final oil recovery factor. The authors of this study came to the conclusion that some groups of reservoirs have a strong dependence, which is characterized by higher porosity, permeability, and a difference in the geological age of the rock. Models of pre-production and post-production phases are considered. The authors found that the accuracy of the pre-production phase model is relatively low for the entire dataset. And also, it was revealed that the oil recovery factor significantly depends on the field development scheme and its efficiency. Therefore, one of the weighty reasons for the poor prediction of the model is the lack of complete data on the distributed parameters in the reservoir.

The work [9] considers machine learning algorithms for estimating the oil production rate using reservoir engineering parameters. The dataset was collected from 93 fields on the Norwegian continental shelf with 30 parameters for each reservoir. The authors of this study trained the model using linear regression and support vector machines. Thus, the authors of this work assume that the methods they have considered can be used to estimate the recovery factor for fields at the stages of appraisal and production. Analysis of the results of this work on the prediction of the oil recovery factor shows that the application of linear regression in choosing the most influential input parameters compared to the support vector machine algorithm is inefficient.

The authors of [10] considered the applicability of various machine learning methods for predicting some rock properties. As these properties, the authors chose porosity, absolute permeability and mass concentration of salts. The dataset was formulated from over 100 laboratory experiments. The authors found that the support vector machine algorithms and the two-layer neural network algorithms are the best algorithms for predicting porosity and permeability in their experiment. However, other algorithms considered by the authors predicted the data well only in some cases. In another paper [11], the authors reviewed the application of artificial intelligence algorithms to estimate the oil recovery factor from water drive sand reservoirs. The artificial intelligence was trained from data from 130 sand reservoirs, and data from another 38 reservoirs were used for testing. In this study, the authors used 10 parameters to estimate the oil recovery factor using four artificial intelligence algorithms: ANN, radial basis neural networks, adaptive neuro-fuzzy inference system and SVM. It was found that the ANN model had a low deviation compared to other considered algorithms, which were trained quite well, however, when predicting test data, the coefficient of determination decreased.

The authors of [12] examined various machine learning methods for predicting downhole pressure, oil production, and predicting water cut in production tasks. The dataset in this study was generated using the ECLIPSE reservoir simulator. The authors applied ten different machine learning methods and took into account the effects of multiphase flow and data noise. In their study, ridge regression and support vector machines performed best at all noise levels. However, when predicting oil production, the ridge regression was less able to cope with water cut fluctuations compared to the performance of support vector regression. In addition, there are several works related to the application of machine learning methods for processing data from permanent downhole gauges (PDG). The authors of [13] used a simple core and approaches to data analysis based on the convolutional core method for inter-

preting data from permanent downhole gauges. These PDGs have noise due to the dynamics of the processes occurring in the well. The authors have dedicated this work to developing robust methods for processing big data from permanent downhole gauges. The authors divided the application of the algorithm into two stages: where the pressure data and flow rate from the PDG were used for training, and in the second stage, the algorithm predicts pressures depending on the flow rate. In this work, the convolutional kernel-based method was trained until the algorithm converged. The authors showed that the convolutional core method perfectly removes noise, but it turned out that this algorithm works very slowly. In the subsequent paper [14] by the same authors, a method for analyzing PDG data in the presence of significant data noise, outliers, and gaps was considered. The convolution kernel method was tested with data where there were significant outliers and aberrant segments and unknown pressure. As a result, with an incomplete history of oil production, the method could identify a reservoir model with an effective flow rate and find the initial pressure. The authors of [15] also examined the application of machine learning methods for interpreting pressure, flow rate, and temperature data from permanent downhole gauges. In this paper, three ML methods were applied: linear regression (LR), core method and ridge core regression. In addition, the authors showed that ML can simulate the generated data from the PDG, even when the physical model is complex. The authors found that when predicting reservoir pressure, the kernel method overfits due to high variance and predicts poorly compared to linear regression and kernel ridge regression due to a lack of interpretability property.

The authors of [16] built an ANN that predicts the productivity of wells using their own history. However, the study results do not claim that ANN prediction is a substitute for empirical or numerical simulation for predicting well production. The authors propose applying ANN prediction to provide confidence in data-driven prediction methods. Moreover, there is another work in which machine learning methods were built to predict Montney and Duvernay well production [17]. The authors found that of the several machine learning methods examined, random forest was the most accurate for their task. This method gave the authors higher prediction accuracy due to the absence of over-fitting problems.

Unlike other works, in this paper, regression algorithms were applied to a synthetic dataset that was manually generated with different values of the parameters (viscosity, porosity and permeability) of oil production in a mathematical model of Buckley-Leverett, which is used to calculate the hydrodynamics and determine the distribution of saturation in oil problems. Thus, various scenarios are considered for the influence of the characteristics used on the oil recovery factor to improve the efficiency of oil production. The development of an algorithm for predicting the oil recovery factor with high accuracy using regression algorithms on synthetic data is promising, because using a mathematical model, it is possible to interpret the simulation of various reservoirs with the tuning sets of oil parameters to estimate oil production in the initial stage of fields.

3. The aim and objectives of the study

The aim of this study is to develop an algorithm for predicting the oil recovery factor with high ac-

curacy using regression algorithms for synthetic data. This will make it possible to predict and improve the efficiency of oil recovery factor with regression algorithms.

To achieve this aim, it is necessary to solve the following objectives:

– to generate synthetic data using an ensemble scenario method based on the numerical 2D Buckley-Leverett model;

– to implement predicting the oil recovery factor using machine learning methods such as multiple linear regression and polynomial regression;

– to evaluate the quality of machine learning algorithms in order to determine the best regression model.

4. Materials and methods

4. 1. The Buckley-Leverett model
The Buckley-Leverett model is written as follows:

$$m\frac{\partial s}{\partial t}+div(\vec{v}_1)=q_1, \tag{1}$$

$$-m\frac{\partial s}{\partial t}+div(\vec{v}_2)=q_2, \tag{2}$$

$$\vec{v}_i=-K_0\frac{f_i(s)}{\mu_i}\nabla P, \tag{3}$$

where K_0 – permeability tensor, s – water saturation, q_i – source or sink, f_i, μ_i – relative phase permeabilities and viscosities of liquids of the corresponding phases, which are dependences of the following form:

$$f_1(s)=s^{3.5};\ \ f_2(s)=(1-s)^{3.5}. \tag{4}$$

The developed numerical model was run many times at different values of viscosity, porosity, absolute permeability to calculate the oil recovery factor, creating various scenarios for further application of ML methods to the data. To observe the change in the value of the oil recovery factor, the time iteration parameter for each pair was also recorded in the output data.

4. 2. Workflow of building a machine learning model
In this work, the obtained synthetic data from a mathematical model were divided into a training and test sample. Four parameters were taken as input parameters of the machine learning model, and the oil recovery factor was taken as the output parameter. Fig. 1 describes the process of building a machine learning model in this study.
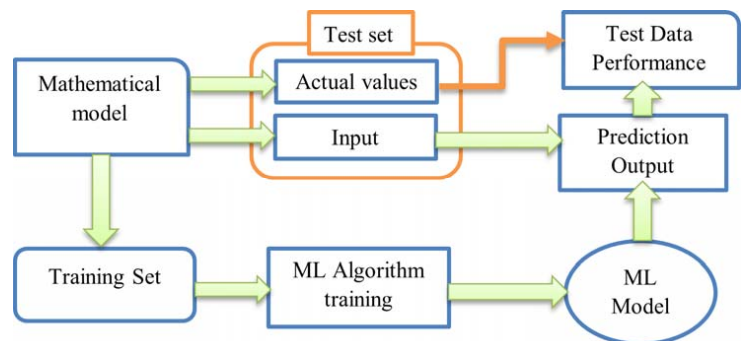

Fig. 1. Workflow of building a machine learning model

In this paper, we consider the task of supervised learning, which is one class of machine learning problems. Our task belongs to the class of regression problems in terms of machine learning methods. A synthetic data set was obtained from a mathematical model: absolute permeability k, porosity p, viscosity μ, time iteration t and oil recovery factor η. In our case, the oil recovery factor is represented as the objective function y, and the other four data are presented as signs of x.

$$x^{(i)} = \begin{bmatrix} k^i \\ p^i \\ \mu^i \\ t^i \end{bmatrix}, \quad i = 1,\dots,m, \tag{5}$$

where $x^{(i)}$ is the sign of the ith training example.

$$y^{(i)} = \eta^{(i)}, \quad i = 1,\dots,m, \tag{6}$$

where $k^{(i)}$, $p^{(i)}$, $\mu^{(i)}$ and $t^{(i)}$ are absolute permeability, porosity, viscosity, temporary iteration on i^{th} data, and m is the number of training examples (training example $m=403440$). Thus, x is an $(n_x+1) \times m$ matrix, and the objective function y is an $m \times 1$ vector. The regression model can be written as follows:

$$y^{(i)} = h\left(x^{(i)}\right) + \varepsilon^{(i)}, \quad i = 1,\dots,m, \tag{7}$$

where model h describes a pattern between x and y, and ε^i is a model error and measures some discrepancies. Consider the subgenus about methods.

4. 3. Multiple linear regression

In multiple linear regression, the hypothesis function h is described as follows:

$$h_\theta(x) = \theta^T x = \theta_0 + \theta_1 x_1 + \dots + \theta_{n_x} x_{n_x}, \tag{8}$$

where n_x is the number of features, in our case, $n_x=4$ given from (5). θ is a $(n_x+1) \times 1$ vector with model parameters (coefficients). The parameter θ_0 corresponds to $x_0=1$. (8) can be written as follows:

$$\begin{aligned} y^{(i)} &= \theta_0 + \theta_1 k^{(i)} + \theta_2 p^{(i)} + \\ &+ \theta_3 \mu^{(i)} + \theta_4 t^{(i)} + \varepsilon^{(i)}, \quad i = 1,\dots,m. \end{aligned} \tag{9}$$

To evaluate regression models, a quadratic loss function is often chosen. A coefficient of determination R^2 was also used to evaluate the results, which provides a measure of how well the observed results are reproduced by the model, based on the proportion of the total variation in the results explained by the model. The mean square error (MSE) is often used as an estimate of the loss function between the target and the predicted function:

$$MSE = \frac{1}{m} \sum_{i=1}^{m} \left(y^{(i)} - y_{pred}{}^{(i)} \right)^2. \tag{10}$$

Using linear regression, the model was trained with four input parameters and oil recovery factor. As a result, the trained model predicts the value of the oil recovery factor based on test data. Although multiple linear regression is very simple, the model has several good advantages. The linear regression model frees the engineer from the need for good physics knowledge in this study. This model is well-trained and highly interpreted, since all independent variables of multiple regression directly affect the target function. Consequently, the influence of input parameters is easily detected and visualized.

4. 4. Polynomial regression method

Polynomial regression is essentially a type of regression in which the ratio of the independent features of x and the dependent objective function y is modeled as a polynomial of n-th degree:

$$h_\theta(x) = \theta_0 + \theta_1 x + \theta_2 x^2 + \dots + \theta_n x^n, \quad i = 1,\dots,n, \tag{11}$$

where n is the degree of the polynomial that is used to transform the linear regression model. Polynomial regression may have a non-linear curve, but the model is still considered linear, since the model parameters associated with the attributes are linear.

Regression models are used to solve over-fitting problems in regression models. A regression model that uses L1 regularization is called Lasso Regression, and a model that uses L2 is called Ridge Regression. The regularization of L2 adds the coefficient of quadratic magnitude to the loss function and is presented in the following form:

$$L = \sum_{i=1}^{m} \left(y_i - \sum_{j=1}^{n_x} x_{ij} \theta_j \right)^2 + \lambda \sum_{j=1}^{n_x} \theta_j^2, \tag{12}$$

where λ is a setting parameter. A well-chosen value of parameter λ helps to avoid the problem of over-fitting. A regularization of L1 adds the absolute value of magnitude to the loss function and is presented in the following form:

$$L = \sum_{i=1}^{m} \left(y_i - \sum_{j=1}^{n_x} x_{ij} \theta_j \right)^2 + \lambda \sum_{j=1}^{n_x} \left| \theta_j \right|. \tag{13}$$

In this paper, polynomial regression (PR) is used as a modification of multiple linear regression. The use of polynomial regression makes it possible to increase the degree of the polynomial n, which in turn improves the multiple linear regression by adding non-linearity to the data. However, in many cases this is not a guarantee that as the degree of the polynomial increases, the model under consideration will learn even better. Because there are problems with under-fitting and over-fitting. To select the optimal model, it is necessary to find a compromise between bias and dispersion.

5. Results of the study of machine learning methods

5. 1. Dataset generation

The dataset is generated synthetically using an ensemble of scenarios based on the Buckley-Leverett 2D model. As input parameters, various combinations of parameters of the oil production problem (porosity, viscosity of the oil phase and absolute rock permeability, time iteration) were taken (Table 1). And as the output parameter, the value of the oil recovery factor was chosen. Thus, in this work, the number of sample pairs is 41*41*6=10,086. Using the Buckley-Leverett model, 6 synthetic data packets were generated for various permeability indices. Each data packet contains

the values of viscosity, porosity and oil recovery factor (if we take into account the data for each time layer, then the total amount of data is 403,440). Oil viscosity varies in the range 0.1–0.5, porosity in the range 0.1–0.3, and various permeability options.

Table 1

Input parameters

Parameters	Number of variation
Porosity	41
Viscosity of oil phase	41
Permeability	6
Time iteration	40

The data was divided into a training and test set. For training, 8,069 sets (80 %) of the total data were used, and for the test the remaining 2,017 pairs (20 %). Python was chosen as the runtime environment for machine learning. As mentioned earlier, the total number of sample pairs is 10,086 models. Each sample pair consists of 40 oil recovery factor values.

5. 2. Results of the machine learning methods

The following are the results of predicting multiple linear regression and polynomial regression. As mentioned earlier, the total number of sample pairs is over 10,000. However, the results will be shown for some sample pairs. Fig. 2 shows the results of one test sample pair for the linear and quadratic polynomial regression method.

Using polynomial regression (PR) increases the complexity of the model. For training with polynomial properties, it is important to choose the desired model, that is, the degree of the polynomial. Increasing and also modifying the degree of the model to cubic polynomial regression with the L1 regularization gives the following results (Fig. 3).

To improve the cubic model, an L1 type regularization with the optimal value of λ was applied. The following figure shows cubic polynomial regression for other test data pairs (Fig. 4).
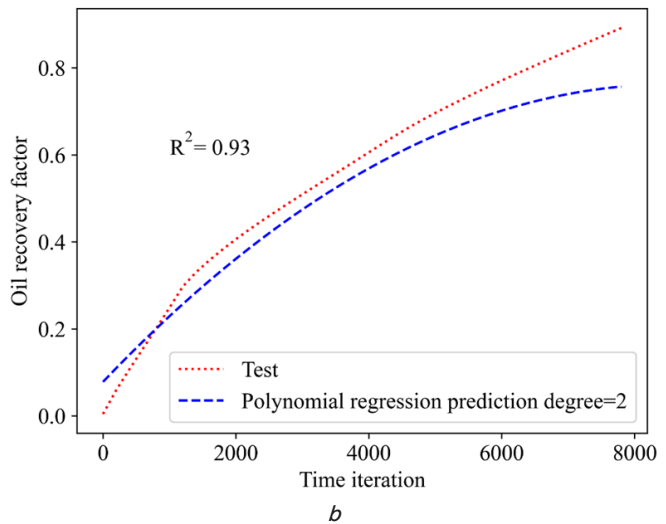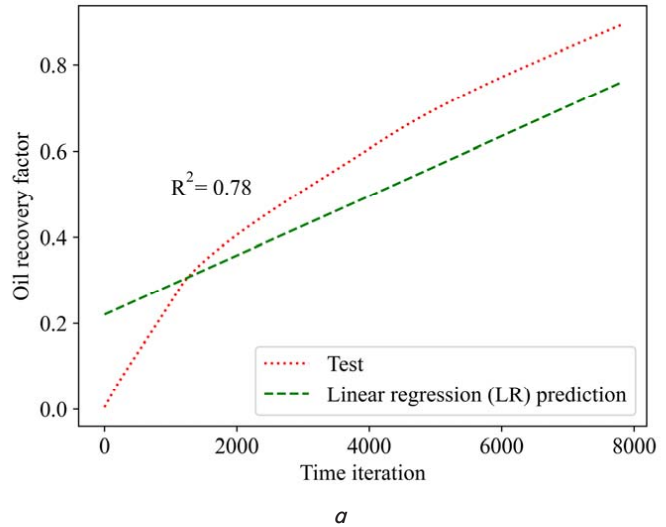

a

b

Fig. 2. Predictions of oil recovery factor on regression models: *a* — linear regression; *b* — quadratic polynomial regression
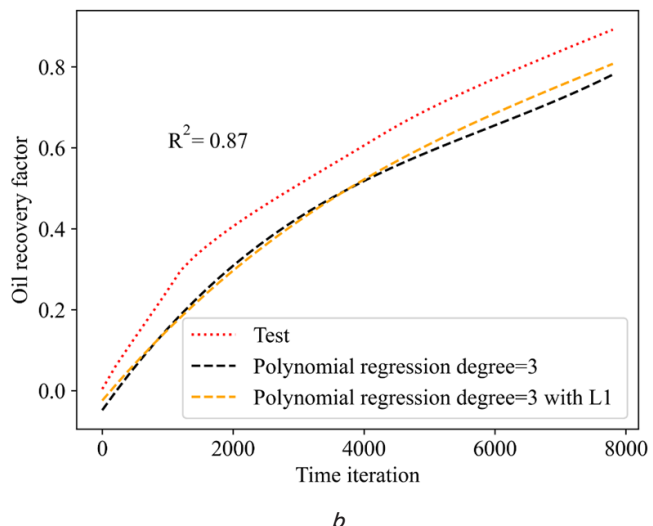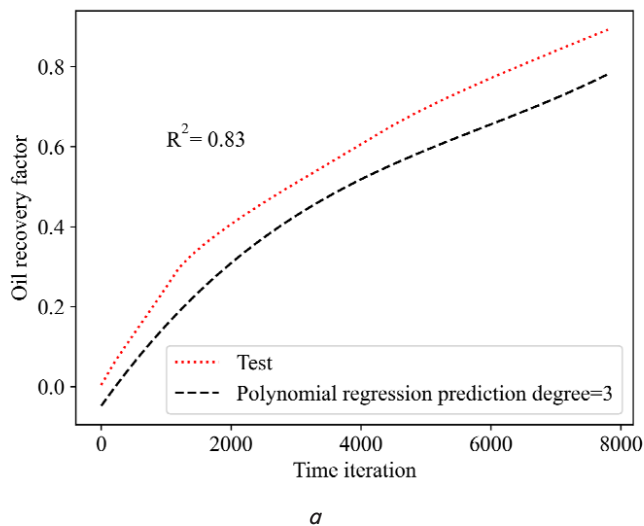

a

b

Fig. 3. Predictions of oil recovery factor on regression models: *a* — cubic polynomial regression; *b* — cubic polynomial regression with L1 regularization
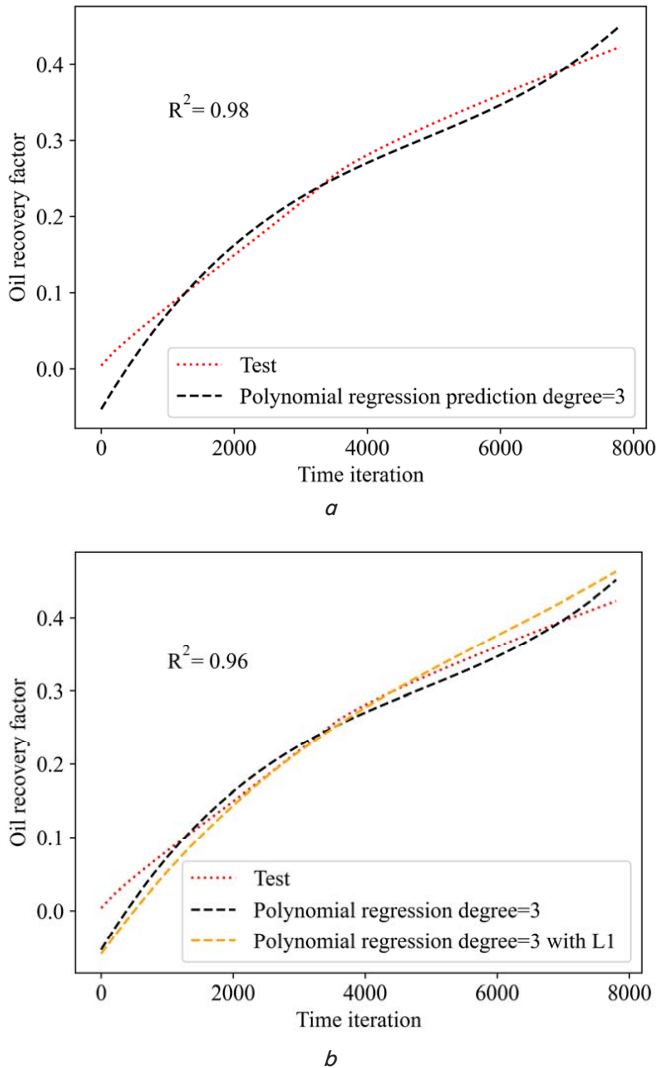
Fig. 4. Predictions of oil recovery factor on regression models
for other test data pairs: *a* — cubic polynomial regression;
b — cubic polynomial regression with L1 regularization

As can be seen from the tested results, as the degree of the polynomial increases, the model captures more data and predicts better than linear regression. The addition of L1 type regularization to the cubic polynomial regression is motivated by the fact that in most test data, the coefficient of determination decreases, this is due to over-fitting. However, in some cases where the oil recovery factor is small, the model predicts very closely with the test results.

5. 3. Evaluation of the quality of machine learning algorithms

The mean square error (MSE) and the coefficient of determination R^2 were used to evaluate machine learning regression methods. The following Table 2 shows the average MSE score for all 20 % of the test sets.

The following Table 3 shows the average R^2 score for 80 % of training sets and 20 % of test sets.

From Table 3, it is noticeable that cubic polynomial regression is trained fairly well with the training set, but the determination coefficient R^2 on test data decreases due to the high dispersion between the data sets at degree=3.

Table 2

MSE score for all test case pairs

Machine learning algorithms	Test sets (20 %) MSE
LR	0.0037
Polynomial Regression (PR) degree=2	0.0016
Polynomial Regression (PR) degree=3	0.0084
Polynomial Regression (PR) degree=3 with L1	0.0033

Table 3

Evaluation of R^2 for all pairs of training and test set

Machine learning algorithms	Train sets (80 %) R^2	Test sets (20 %) R^2
LR	0.87	0.91
PR degree=2	0.95	0.96
PR degree=3	0.97	0.79
PR degree=3 with L1	0.96	0.92

6. Discussion of the regression algorithms results

From Fig. 2, we can see that the predicted LR function does not capture all patterns in the data. Consequently, the multiple linear regression model has an example of under-fitting. Moreover, it is clear that the quadratic polynomial model trains data better than a linear model. This is evidenced by the MSE estimates, where the error of the quadratic polynomial regression is reduced. In addition, it can be noted that the determination coefficient R^2 has increased compared to the linear model.

It is noticeable that a cubic polynomial model predicts data worse than a quadratic model (Fig. 3). Moreover, the determination coefficient R^2 decreased compared to the quadratic model. Even adding L1 regularization did not improve the cubic model that much. Thus, the quadratic model is the most optimal for this test pair. However, this pattern is true only for this pair. For the remaining pairs from the entire test sample, the results may be different. For example, this is illustrated in Fig. 4 where the cubic model trains much better and predicts very closely with the test data in both cases. This is because the polynomial cubic model in our case has over-fitting due to its high variance. By applying L1 regularization, the cubic polynomial model predicts the function for all test data rather well than the simple cubic model. This can be seen from Table 3.

A feature of the considered methods is the use of regression algorithms for the generated data, which were obtained from the launches of the implementation of the Buckley-Leverett mathematical model using an ensemble of scenarios.

This paper discusses a data-driven approach that can perfectly predict the output parameter using big data, however, the slight disadvantage of this method is the difficulty of interpretation. Therefore, in the future, for the development of this study, there is a motivation to consider the direction of scientific machine learning based on physical modeling, which takes into account physics. In the following works, it is planned to conduct research in the direction of physics-informed neural networks (PINN) for solving problems of fluid flow in a porous medium.

7. Conclusions

1. The dataset was generated synthetically using the scenario ensemble method from the Buckley-Leverett

mathematical model, where for different oil input parameters (41 viscosity and porosity values, 6 permeability types, 40 time iteration of mathematical model values) different output values of the oil recovery factor were obtained. More than 400,000 synthetic datasets were used to train regression methods.

2. Multiple linear regression was implemented to predict the oil recovery factor. The average R^2 coefficient of determination LR for all test samples was 0.91, which is a good result. However, for some test data pairs, there are examples of under-fitting, which affects the prediction of the model. Therefore, different degrees of polynomial regression are implemented and tested to improve the linear model. Quadratic and cubic polynomial regression models with modified L1 type regularization were implemented, thereby improving the R^2 score to 0.96 for all test pairs of the sample.

3. To evaluate the quality of the considered regression methods, the metrics mean squared error and the coefficient of determination R^2 were used. It was found that for some test pairs where the recovery factor is small, the cubic model at 0.98 accurately predicts with respect to the test data. However, for our synthetic data, quadratic polynomial regression is the best model for predicting the oil recovery factor for all test pairs. In future works, it is planned to add noise in the form of practical data from real oil fields.

References

1. Krasnov, F., Glavnov, N., Sitnikov, A. (2017). A Machine Learning Approach to Enhanced Oil Recovery Prediction. Analysis of Images, Social Networks and Texts, 164–171. doi: https://doi.org/10.1007/978-3-319-73013-4_15

2. Guo, Z., Reynolds, A. C., Zhao, H. (2017). A Physics-Based Data-Driven Model for History-Matching, Prediction and Characterization of Waterflooding Performance. Day 3 Wed, February 22, 2017. doi: https://doi.org/10.2118/182660-ms

3. Saberi, H., Esmaeilnezhad, E., Choi, H. J. (2021). Artificial Neural Network to Forecast Enhanced Oil Recovery Using Hydrolyzed Polyacrylamide in Sandstone and Carbonate Reservoirs. Polymers, 13 (16), 2606. doi: https://doi.org/10.3390/polym13162606

4. Vo Thanh, H., Sugai, Y., Sasaki, K. (2020). Application of artificial neural network for predicting the performance of CO_2 enhanced oil recovery and storage in residual oil zones. Scientific Reports, 10 (1). doi: https://doi.org/10.1038/s41598-020-73931-2

5. Koperna, G. J., Melzer, L. S., Kuuskraa, V. A. (2006). Recovery of Oil Resources From the Residual and Transitional Oil Zones of the Permian Basin. All Days. doi: https://doi.org/10.2118/102972-ms

6. Cheraghi, Y., Kord, S., Mashayekhizadeh, V. (2021). Application of machine learning techniques for selecting the most suitable enhanced oil recovery method; challenges and opportunities. Journal of Petroleum Science and Engineering, 205, 108761. doi: https://doi.org/10.1016/j.petrol.2021.108761

7. Ahmadi, M. A., Soleimani, R., Lee, M., Kashiwao, T., Bahadori, A. (2015). Determination of oil well production performance using artificial neural network (ANN) linked to the particle swarm optimization (PSO) tool. Petroleum, 1 (2), 118–132. doi: https://doi.org/10.1016/j.petlm.2015.06.004

8. Makhotin, I., Orlov, D., Koroteev, D., Burnaev, E., Karapetyan, A., Antonenko, D. (2021). Machine learning for recovery factor estimation of an oil reservoir: A tool for de-risking at a hydrocarbon asset evaluation. Petroleum. doi: https://doi.org/10.1016/j.petlm.2021.11.005

9. Aliyuda, K., Howell, J. (2019). Machine-learning algorithm for estimating oil-recovery factor using a combination of engineering and stratigraphic dependent parameters. Interpretation, 7 (3), SE151–SE159. doi: https://doi.org/10.1190/int-2018-0211.1

10. Erofeev, A., Orlov, D., Ryzhov, A., Koroteev, D. (2019). Prediction of Porosity and Permeability Alteration Based on Machine Learning Algorithms. Transport in Porous Media, 128 (2), 677–700. doi: https://doi.org/10.1007/s11242-019-01265-3

11. Mahmoud, A., Elkatatny, S., Chen, W., Abdulraheem, A. (2019). Estimation of Oil Recovery Factor for Water Drive Sandy Reservoirs through Applications of Artificial Intelligence. Energies, 12 (19), 3671. doi: https://doi.org/10.3390/en12193671

12. Ristanto, T., Horne, R. (2018). Machine Learning Applied to Multiphase Production Problems. Tita Ristanto. Available at: https://www.researchgate.net/publication/327977359_Machine_Learning_Applied_to_Multiphase_Production_Problems

13. Liu, Y., Horne, R. N. (2013). Interpreting Pressure and Flow Rate Data from Permanent Downhole Gauges Using Convolution-Kernel-Based Data Mining Approaches. All Days. doi: https://doi.org/10.2118/165346-ms

14. Liu, Y., Horne, R. N. (2013). Interpreting Pressure and Flow Rate Data from Permanent Downhole Gauges with Convolution-Kernel-Based Data Mining Approaches. Day 2 Tue, October 01, 2013. doi: https://doi.org/10.2118/166440-ms

15. Tian, C., Horne, R. N. (2019). Applying Machine-Learning Techniques To Interpret Flow-Rate, Pressure, and Temperature Data From Permanent Downhole Gauges. SPE Reservoir Evaluation & Engineering, 22 (02), 386–401. doi: https://doi.org/10.2118/174034-pa

16. Cao, Q., Banerjee, R., Gupta, S., Li, J., Zhou, W., Jeyachandra, B. (2016). Data Driven Production Forecasting Using Machine Learning. Day 2 Thu, June 02, 2016. doi: https://doi.org/10.2118/180984-ms

17. Chen, S. (2019). Application of Machine Learning Methods to Predict Well Productivity in Montney and Duvernay. University of Calgary. Available at: https://higherlogicdownload.s3.amazonaws.com/SPE/5fc0079d-67a5-4dd9-a56f-190534ef5d3d/UploadedImages/2019_04_16_ML_SPE_Presentation_Revised.pdf