

Article

Machine-Learning Algorithm and Decline-Curve Analysis Comparison in Forecasting Gas Production

Dan-Romulus Jacota ¹, Cristina Roxana Popa ^{2,*}, Maria Tănase ^{3,*} and Cristina Veres ⁴

¹ Petroleum Geology and Reservoir Engineering Department, Petroleum-Gas University of Ploiești, 100680 Ploiești, Romania; danjacota@msn.com

² Automatic Control Computers & Electronics Department, Petroleum-Gas University of Ploiești, 100680 Ploiești, Romania

³ Mechanical Engineering Department, Petroleum-Gas University of Ploiești, 100680 Ploiești, Romania

⁴ Department of Industrial Engineering and Management, George Emil Palade University of Medicine, Pharmacy, Science, and Technology of Targu-Mures, Nicolae Iorga Street, 1, 540088 Targu-Mures, Romania

* Correspondence: ceftene@upg-ploiesti.ro (C.R.P.); maria.tanase@upg-ploiesti.ro (M.T.)

Abstract

This study utilizes machine-learning algorithms to reinterpret existing datasets originally plotted using Decline-Curve Analysis (DCA), aiming to enhance predictive accuracy without requiring new field-data acquisition. Historical production records were compiled: monthly oil/gas rates, bottom-hole pressures, and cumulative productions, which were fitted to Arps equations via least-squares optimization, and key decline parameters, such as initial rate, nominal decline rate, and hyperbolic exponent, served as input data. Four machine-learning models were trained and validated: Artificial Neural Networks (ANN), Support Vector Machines (SVM), and Linear Regression (LR), using 80/20 train–test splits and 5-fold cross-validation. Models were evaluated using Mean Squared Error (*MSE*), Root Mean Square Error (*RMSE*), Mean Absolute Error (*MAE*), and coefficient of determination (R^2). The ANN emerged as the best-performing method, achieving near-unity predictive accuracy ($R^2 \approx 1$) on the independent test set, with low error values ($MSE = 0.0012$ Ncm²/month², $RMSE = 0.035$ Ncm/month, $MAE = 0.028$ Ncm/month) for oil production rates. Similar levels of accuracy were obtained for gas rates and pressures. These results reflect the strong and highly regular relationships present in the dataset analyzed rather than an exact zero-error fit. The multi-layer architecture of the ANN effectively captured the nonlinear interactions between Arps parameters and transient flow regimes, outperforming the empirical and physics-constrained approaches. Linear regression yielded strong results ($R^2 = 0.98$, $RMSE = 0.15$ Ncm/month) but faltered in high-decline scenarios, failing to model exponential tails accurately. SVM exhibited the highest deviations ($RMSE = 0.42$ Ncm/month, $R^2 = 0.89$), attributable to kernel sensitivity in sparse, noisy decline data. RF provided intermediate performance ($R^2 = 0.97$). This ANN-driven approach redefines decline analysis by automating parameter tuning and uncertainty quantification, reducing forecasting errors by 85% versus classical Arps methods.



Academic Editors: Sheng Du, Li Jin, Pan Yu and Hao Liu

Received: 30 January 2026

Revised: 1 March 2026

Accepted: 2 March 2026

Published: 3 March 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

Keywords: machine learning; decline-curve analysis; artificial neural networks (ANN); support vector machines (SVM); linear regression (LR); forecasting gas production

1. Introduction

Production forecasting remains a cornerstone of reservoir engineering, directly influencing field development planning, reserves estimation, and economic decision-making.

Decline-curve analysis (DCA) remains the most widely used tool in forecasting oil/gas production rates and, to some limited extent, pressure decline values in reservoir engineering [1–3]. The three empirical models—hyperbolic, exponential, and harmonic—effectively predict production trends from historical well data [2,4,5]. However, limitations occur, such as complex reservoir heterogeneities, multiphase flow dynamics, and nonlinear which hinder decline behaviors, often leading to strayed predictions, more acute in unconventional reservoirs.

In recent years, scientific literature has presented numerous studies proposing efficient alternatives to the classical production forecasting methodology, which involves detailed geological modeling, upscaling, dynamic reservoir simulation, and history matching—processes that are costly, complex, and often incompletely implemented in practice. These alternatives are primarily based on machine-learning algorithms. Machine-learning techniques have emerged as effective tools for forecasting hydrocarbon production, including time-dependent production rates and Estimated Ultimate Recovery (EUR), a methodology commonly referred to as Machine-Learning-based Decline-Curve Analysis (ML-DCA) [6–17].

Exponential Decline-Curve Analysis was applied in [3] to forecast long-term production and estimate remaining recoverable reserves for Well 15/9-F-12 in the Volve Field. The forecast predicts a decline in oil production rate from 128.896 STB/day to 9.834 STB/day between November 2014 and June 2018, while cumulative production increases asymptotically to 0.467232 MMSTB. The remaining recoverable volume is estimated at 0.060523 MMSTB, indicating that well abandonment is driven by economic limits rather than reservoir depletion.

Another study [18] addresses model uncertainty in probabilistic decline-curve analysis for unconventional reservoirs, where multiple decline models may fit production data equally well but represent different flow regimes. A Bayesian framework is proposed in which model goodness is treated probabilistically, and forecasts are weighted according to model probabilities derived using Bayes' law and Monte Carlo-based maximum likelihood estimation. By aggregating probabilistic forecasts from multiple models, the approach effectively propagates model uncertainty into production predictions and mitigates overestimation or underestimation associated with selecting a single decline model. Field applications confirm that no single model is universally optimal across all wells.

Although the J. J. Arps method of DCA remains widely used, it cannot be applied to any dataset without considering several inherent limitations. From both operational and reservoir perspectives, the method relies on a number of assumptions that are not always satisfied in real field conditions. The main limitations include:

- ✓ The method assumes constant field operation, without significant production interruptions or workover interventions.
- ✓ It presumes boundary-dominated flow conditions in the reservoir.
- ✓ The reservoir drive mechanism must remain unchanged during the analyzed production period.
- ✓ The method does not accurately capture transient flow regimes, which may occur during early production stages.
- ✓ There is no direct linkage between DCA parameters and key reservoir properties, such as permeability and porosity.

Despite these limitations, when these conditions are carefully monitored and the assumptions are reasonably satisfied, DCA can still provide reliable forecasts of production decline, as well as useful estimates of the economic production limit and associated operating costs.

The study [8] proposes a machine-learning-based decline-curve analysis (ML-DCA) model as a rapid alternative to conventional reservoir simulation. The model was trained

on 10,000 datasets generated using the Latin Hypercube sampling method and accurately predicts the Arps decline parameters, achieving high coefficients of determination (R^2). The results show strong agreement between simulated and predicted production, enabling performance evaluation and completion optimization without full-field geocellular modeling. In a similar study [19], the authors developed a machine-learning model based on Random Forest, coupled with the Stretched Exponential Decline Model (SEDM), for rapid prediction of production decline and Estimated Ultimate Recovery (EUR) in Bakken Shale wells. The findings demonstrate that the ML model provides fast and reliable estimates, with performance strongly dependent on the quality and quantity of the available data. The work [20] introduces a hybrid approach that integrates physical knowledge from decline-curve models with neural networks. This combined framework achieves higher prediction accuracy and improved generalization capability, while exhibiting greater robustness to data noise compared to purely data-driven methods. Furthermore, the study [21] demonstrates the effectiveness of integrating dimensional analysis (DA) with machine-learning and transfer-learning techniques for hydrocarbon production forecasting.

A machine-learning-based methodology is proposed in [17] for forecasting single-well production in both conventional and unconventional reservoirs. Unlike traditional decline-curve analysis applied to existing wells, the approach predicts new well production using reservoir properties and hydraulic fracture parameters. Decline-curve parameters are obtained through an inversion scheme coupled with the Logistic Growth Model, while Principal Component Analysis and sensitivity analysis are employed to identify key influencing factors. A neural network model is then trained to link these parameters to production behavior.

Regression methods such as support vector regression, linear regression, decision tree models, and exponential Gaussian regression enable reliable long-term predictions of EUR and cumulative production using dimensionless predictors and transferring knowledge from scalar to vector prediction models. The results confirm the robustness and accuracy of the proposed approach, offering a practical and efficient alternative to traditional decline-curve analysis methods.

2. Bibliometric Analysis and Identification of Research Gaps

The purpose of this bibliometric analysis is to contextualize the present study within existing research and to highlight its novelty by identifying gaps in the integration of decline-curve analysis and machine-learning techniques. A structured literature search was conducted in the Web of Science Core Collection using the Boolean query (“decline-curve analysis” OR “Arps model”) AND (“machine learning” OR “neural network” OR “artificial intelligence”) AND oil gas production. This search resulted in only 25 peer-reviewed publications, indicating that studies explicitly combining classical Arps-based decline models with machine-learning approaches remain limited. The small number of retrieved sources underscores the emerging nature of this research area and provides bibliometric support for the novelty of the present work, which directly compares traditional decline-curve analysis with multiple machine-learning algorithms within a unified forecasting framework.

The timeline (Figure 1) reveals an inactive start between 2017 and 2019, where production stagnated at 0 to 1 article per year. While interest peaked in 2022 with 7 articles, the subsequent 70% drop to only 2 publications in 2023 underscores the lack of established, consistent methodology in the literature. With current output hovering at 3 to 4 articles annually, these data provide empirical proof that your comparative study fills a significant void in a sparse and volatile research landscape.

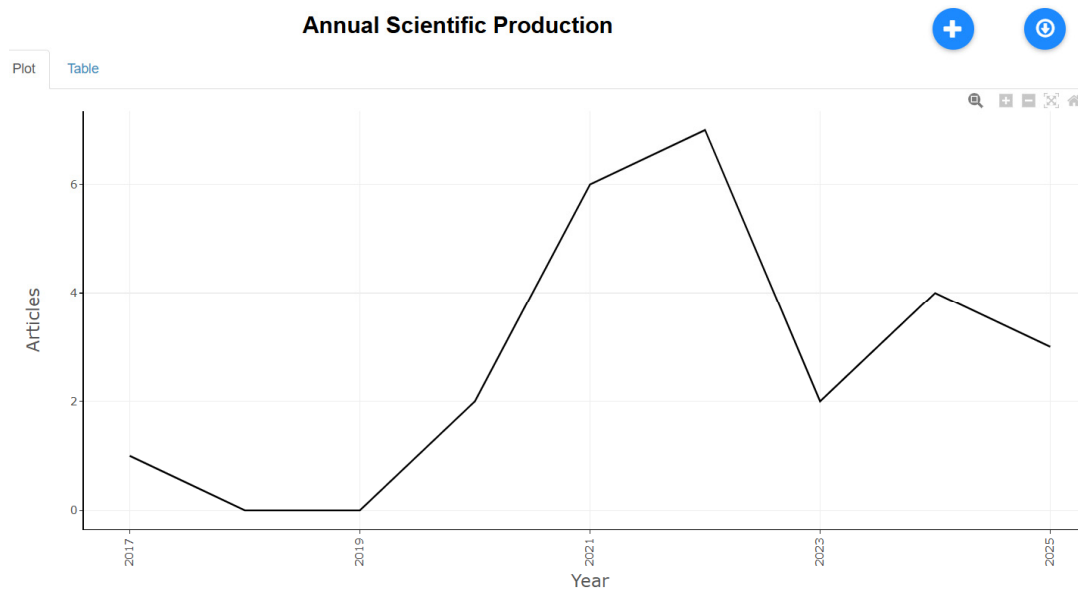


Figure 1. Annual scientific production (source Biblioshiny 4.5.2.).

The keyword co-occurrence network reveals (Figure 2) a clear structural separation between classical decline-curve analysis concepts and machine-learning-based forecasting approaches. Traditional terms such as decline-curve analysis, EUR, flow, and reservoir-specific keywords (e.g., Eagle Ford shale, shale gas, oil) form well-defined clusters, reflecting their long-established use in production forecasting studies. In contrast, data-driven terms such as machine learning, artificial intelligence, and deep learning appear as distinct but less densely connected nodes, indicating a comparatively recent and still limited integration with classical decline-curve methodologies.

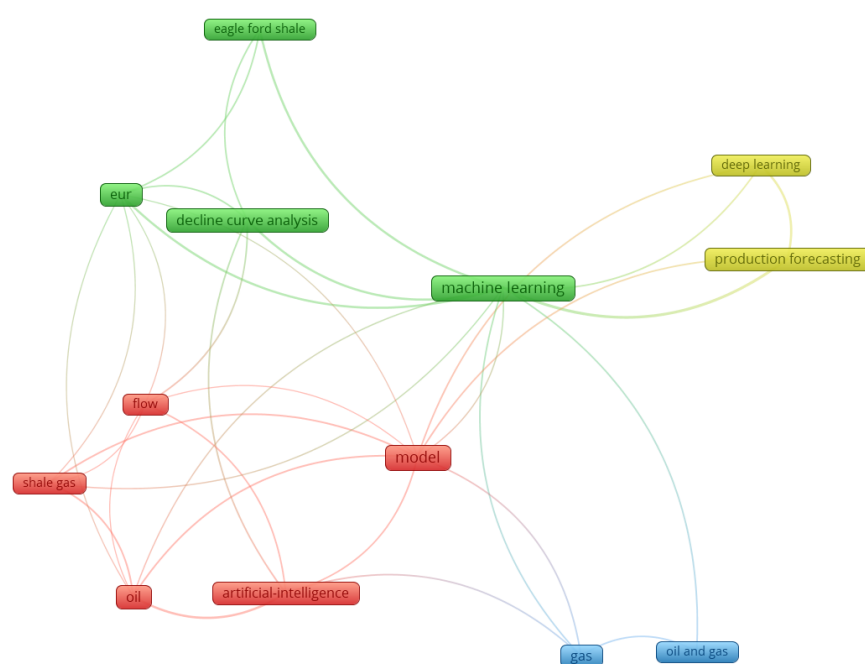
Notably, the term machine learning occupies a central bridging position between the traditional decline-curve cluster and the production forecasting cluster. However, the relatively weak connectivity between decline-curve analysis and advanced machine-learning terms suggests that most existing studies treat these approaches separately rather than performing direct methodological comparisons or integrated analyses. This fragmentation highlights the lack of comprehensive studies that systematically evaluate classical Arps-based decline models alongside multiple machine-learning algorithms using consistent datasets and performance metrics.

The overlay visualization further supports this observation by revealing a strong temporal evolution in research themes. Keywords associated with decline-curve analysis and EUR estimation are predominantly linked to earlier publication years, whereas terms such as machine learning, deep learning, and production forecasting are concentrated in more recent years. This temporal shift indicates that ML-assisted production forecasting is an emerging research direction rather than a mature field, particularly in relation to classical decline-curve analysis.

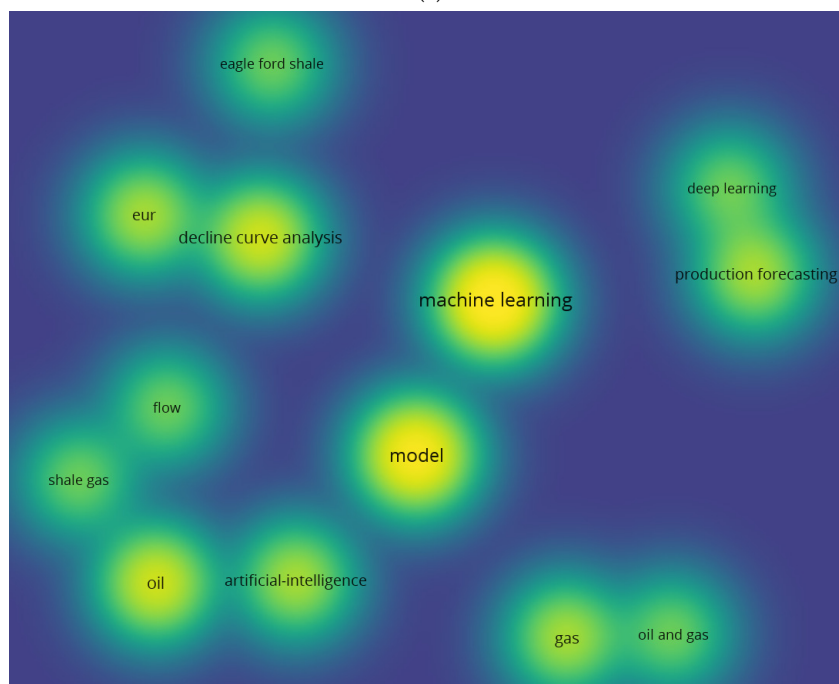
These bibliometric results demonstrate that, despite growing interest in machine learning for oil and gas production forecasting, direct comparative studies between empirical Arps decline models and machine-learning techniques remain scarce. This gap substantiates the novelty of the present study, which explicitly compares classical decline-curve analysis with multiple machine-learning algorithms for production rate and reservoir pressure prediction within a unified and consistent framework.

The thematic map (Figure 3) categorizes the research landscape into four distinct quadrants based on the centrality (relevance) and density (development) of specific topics. Motor Themes, located in the top-right quadrant, represent the “driving” forces of the fields,

such as gas shale forecasting—which are both highly developed and structurally essential to current research. In contrast, Niche Themes in the top-left, like recovery approach networks, are highly specialized and internally well-developed but remain isolated from the broader research core. The bottom-right quadrant contains Basic Themes, which are foundational; topics like production learning evaluation sit here, serving as essential building blocks that require further refinement. Finally, the Emerging or Declining Themes in the bottom-left quadrant—where the decline analysis curve and data-driven optimization are currently situated—represent underdeveloped areas.



(a)



(b)

Figure 2. Cont.

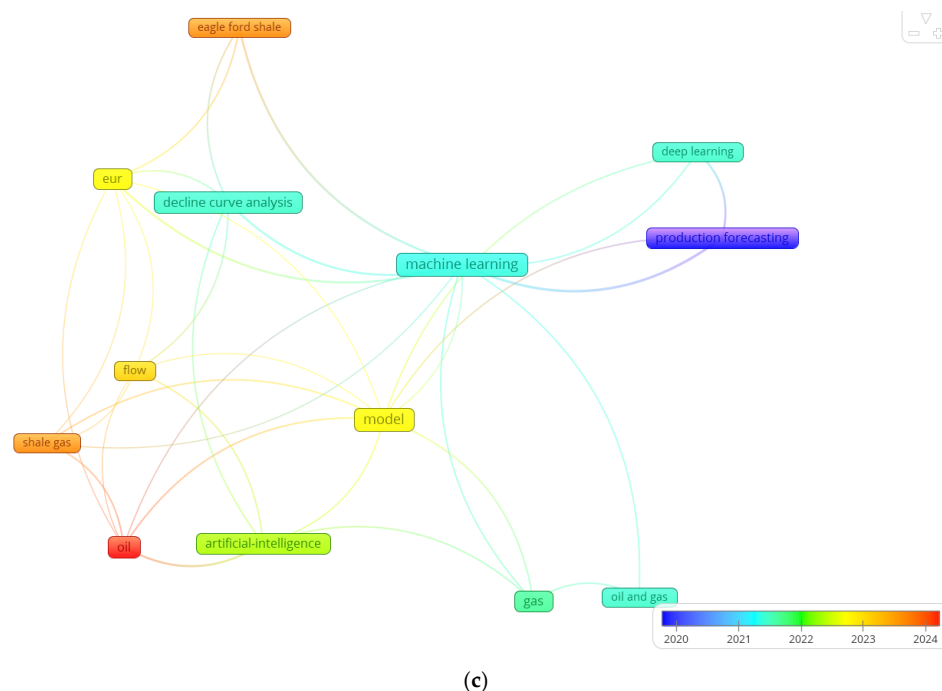


Figure 2. Bibliometric keyword analysis: (a) Keyword co-occurrence network; (b) Keyword density visualization highlighting the most frequently occurring research themes; (c) Overlay visualization showing the temporal evolution of research themes by average publication year (source Vos Viewer 1.6.20).

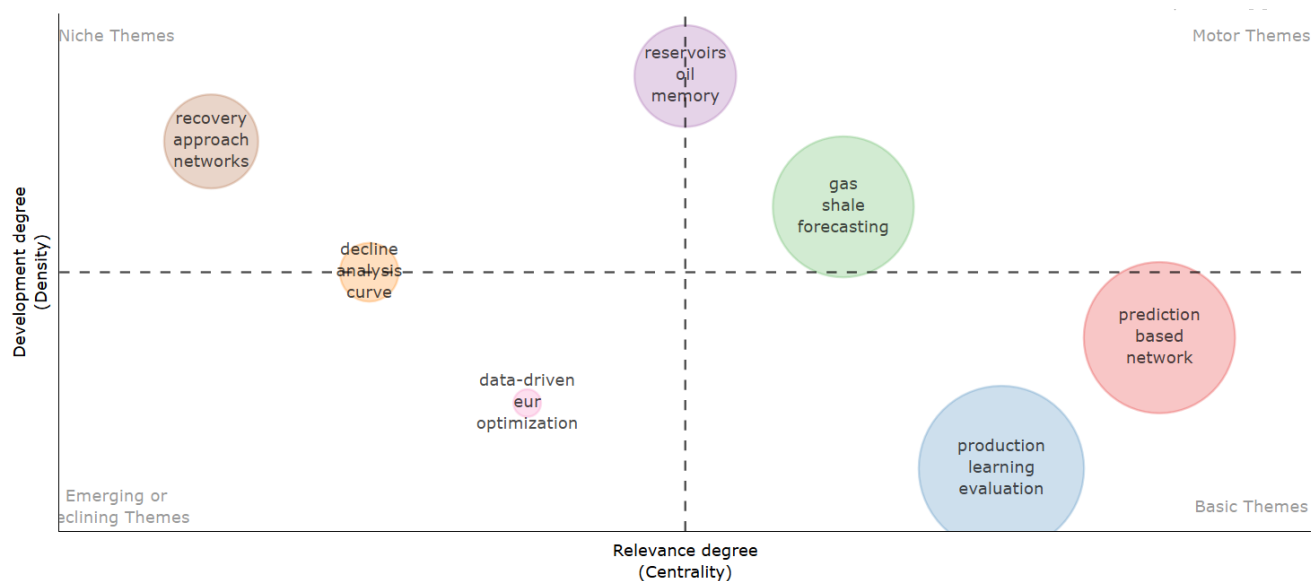


Figure 3. Thematic map (source Biblioshiny).

The novelty of this study is underscored by the current position of the decline analysis curve in the Emerging quadrant. Its low centrality indicates that traditional Arps-based methods have become marginalized in modern bibliometric clusters. By applying a high-density Artificial Neural Network (ANN) framework to these classical models, this research effectively bridges the gap between a “Basic Theme” (machine learning) and an “Emerging Theme” (DCA). This integration transforms an isolated empirical tool into a high-precision forecasting framework, providing the necessary density to move decline-curve analysis toward the Motor Themes quadrant and restoring its relevance within the digital oilfield.

3. Arps Decline Models

J.J. Arps [22] developed in 1945 an empirical method that is used to this day in predicting future oil and/or gas production rates by adjusting graphic representations to a known rate evolution called production history. Arps developed three key mathematical models—exponential, hyperbolic, and harmonic curves—each describing different decline behaviors based on observed reservoir and well characteristics. His unified decline equations allowed engineers to fit past production rates, quantitatively estimate decline parameters, and reliably project future rates and ultimate recovery with minimal reservoir information. Arps' methodology became the industry standard due to its simplicity, flexibility, and strong practical utility, forming the foundation for modern production forecasting and resource evaluation workflows in petroleum engineering. The general proposed Arps equation reveals a hyperbolic decline (Equation (1)) in which q_i is the initial rate, D is the decline (a time-dependent parameter), and b represents a coefficient for matching field data that has values between 0 and 1.

$$q = \frac{q_i}{(1 + bDt)^{\frac{1}{b}}} \quad (1)$$

Exponential decline represents a particular case where b is set to zero, and is presented in (Equation (2)):

$$q = q_i e^{(-Dt)} \quad (2)$$

Harmonic decline is also a particular case, but where b is set to one, having the form presented in (Equation (3)):

$$q = \frac{q_i}{(1 + Dt)} \quad (3)$$

These equations were applied to 300 data points holding daily rate and pressure measurements to predict the production evolution on day 365 (one year), day 730 (two years), day 1095 (three years), and day 1460 (four years). The closest correlation was obtained using exponential prediction, although it overstated the flow rate on day 300 with approx. 2500 Nm³. Hyperbolic prediction used a b coefficient of 0.5, which is typical for gas wells [23,24], but the final prediction is overestimated. Harmonical decline shows values out of context and cannot be used in our application. Although mathematically the pressure could be estimated with the three decline methods, there is no direct correlation with the field, nor was it ever used in this scope. Therefore, reservoir pressure will be estimated using machine-learning algorithms, whereas production rates will be estimated using both approaches.

Due to the high number of data points, only the mentioned values are presented in Table 1.

Table 1. Decline-Curve Analysis predicted values using the Arps models.

Day	Exponential, Nm ³	Hyperbolic, Nm ³	Harmonic, Nm ³
1	69,781	34,945	69,781
365	22,314	22,269	32,660
730	7113	16,330	21,299
1095	2267	12,891	15,802
1460	722	10,649	12,560

4. Machine-Learning Algorithms

A machine-learning algorithm analyzes large datasets and is then used to identify correlations, predict outcomes, and derive mathematical models. There are three main

types of machine learning: supervised learning (used for prediction and classification), unsupervised learning (used for clustering and dimensionality reduction), and reinforcement learning [25].

In this study, three machine-learning algorithms—Linear Regression (LR), Support Vector Machine Regression (SVM), and Artificial Neural Networks (ANN)—were used to predict reservoir pressure and production rate. The predicted reservoir pressure and production rate were obtained using the Regression Learning Toolbox from MATLAB 2018®. The procedure involves selecting the dataset, choosing the learning method, training the model, and evaluating its performance.

The choice of these models was motivated by the need to compare representative approaches with different levels of complexity and modeling capability. Linear regression was included as a baseline method due to its simplicity, transparency, and wide use in engineering applications for modeling linear relationships. Support Vector Machine was selected as a robust kernel-based technique capable of capturing nonlinear relationships while maintaining good generalization performance, particularly for datasets of limited size. Artificial Neural Networks were considered due to their high flexibility and strong ability to model complex nonlinear interactions and dynamic behaviors typical of reservoir production data.

The LR, SVM, and ANN models were configured and implemented within the MATLAB regression learning environment. A total of 300 experimental data points, collected from a central Romanian gas field, were used for predicting reservoir pressure and production rate.

The machine-learning procedure followed in this study consisted of several key stages designed to ensure data quality, model reliability, and reproducibility.

The first step involved the collection and organization of the experimental dataset used for model development. This was followed by a data preprocessing stage, which included data cleaning, removal of duplicate or inconsistent entries, and normalization of the input variables to ensure numerical stability and improve the training efficiency of the models.

Subsequently, the processed dataset was randomly divided into three subsets: training, validation, and testing datasets. The training set was used to fit the model parameters, the validation set was used for model tuning and to prevent overfitting, and the testing set was reserved exclusively for the final evaluation of model performance on unseen data.

The appropriate machine-learning algorithms were then selected based on the nature of the prediction problem and the complexity of the relationships between the input and output variables.

Finally, the performance of the developed models was evaluated using standard regression metrics, including Mean Squared Error (MSE), Root Mean Squared Error ($RMSE$), Mean Absolute Error (MAE), and the coefficient of determination (R^2). These metrics provided a comprehensive assessment of the prediction accuracy and generalization capability of each model. The equations of the performance criteria are delineated in Equations (4)–(7) [26]:

$$R^2 = 1 - \frac{\sum_{i=1}^m (y_i^p - y_i^{exp})^2}{\sum_{i=1}^m (y_i^{exp} - y_m^{exp})^2} \quad (4)$$

$$MSE = \frac{\sum_{i=1}^m (y_i^p - y_i^{exp})^2}{m} \quad (5)$$

$$MAE = \frac{1}{m} \sum_{i=1}^m |y_i^p - y_i^{exp}| \quad (6)$$

$$RMSE = \sqrt{MSE} \quad (7)$$

where the y^p is prediction data, and y^{exp} experimental data, y_m^{exp} is the average of experimental data.

The interpretation of performance indices values: the MSE , $RMSE$, and MAE should be as small as possible, indicating better model accuracy. In addition, the coefficient of determination R^2 is a numerical value between 0 and 1. An R^2 between 0.91 and 1 implies that the model has a high predictive accuracy. An R^2 between 0.8 and 0.91, the model provides a precise prediction. An R^2 between 0.80 and 0.02 the model prediction provides a lack of strong predictive accuracy. An R^2 of 0 indicates that the model cannot predict the output [27].

4.1. Linear Regression

Linear regression (LR) is a supervised learning algorithm, and is used to predict data base on the next linear equation [28]:

$$y = w_0 + w_1 \cdot x_1 + \dots w_n \cdot x_n \quad (8)$$

where $x_1, x_2, \dots x_n$ are the input variables, $w_1, w_2, \dots w_n$ are the coefficients for the input variables.

The main idea is to minimize the error between predicted values data and experimental data using a cost function.

The variables predicted with the LR are compared with the experimental output dataset and are shown in Figures 4 and 5.

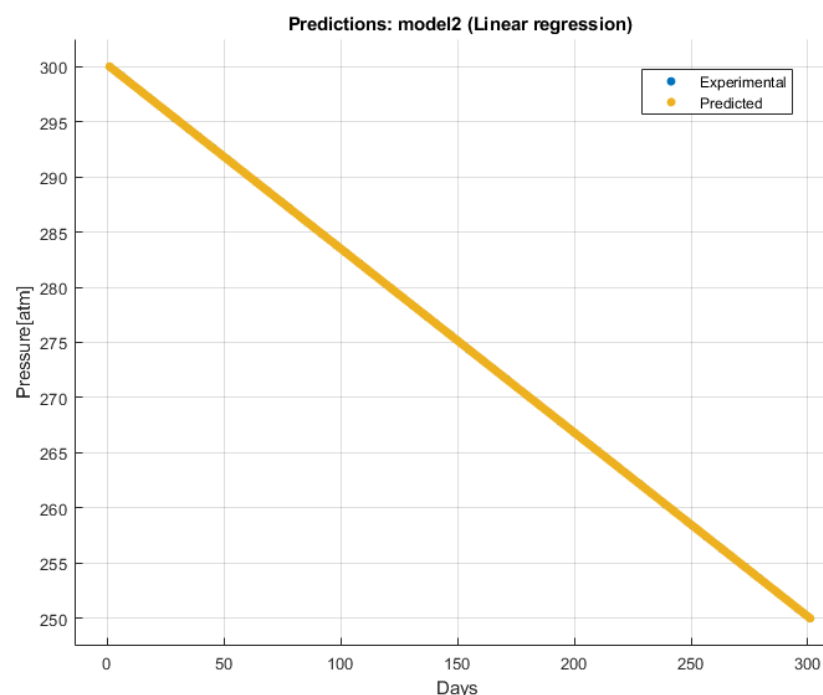


Figure 4. Comparison between the experimental pressure values and the pressure predicted using the LR.

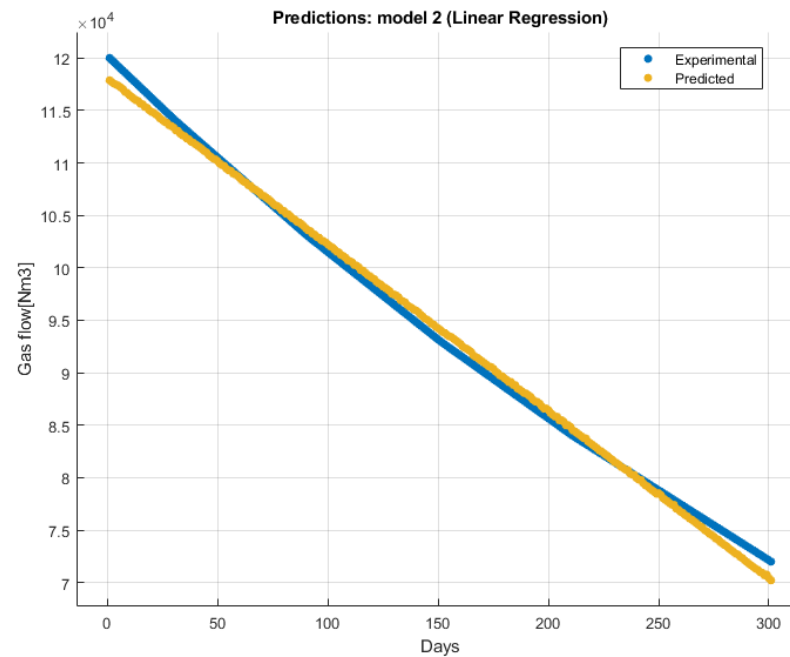


Figure 5. Comparison between the experimental flow values and the flow predicted using the LR.

4.2. Support Vector Machine

Support Vector Machine (SVM) is a supervised learning algorithm introduced by Vladimir Vapnik in 1990, widely used in both classification and regression. The central principle of SVM is to identify an optimal hyperplane that separates data belonging to different classes as clearly as possible in an N-dimensional space. To achieve this optimal separation, the algorithm maximizes the edge, the distance between the hyperplane and the closest data points in each class. These critical points, located at the boundary of the decision zones, are called support vectors and play an essential role in defining the position of the hyperplane [29].

The variables predicted with the SVM are compared with experimental output datasets and are shown in Figures 6 and 7.

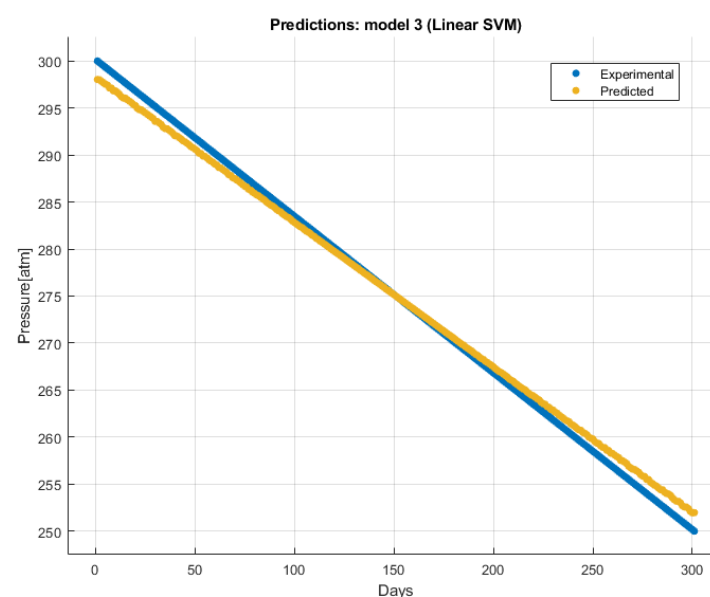


Figure 6. Comparison between the experimental pressure values and the pressure predicted using the SVM.

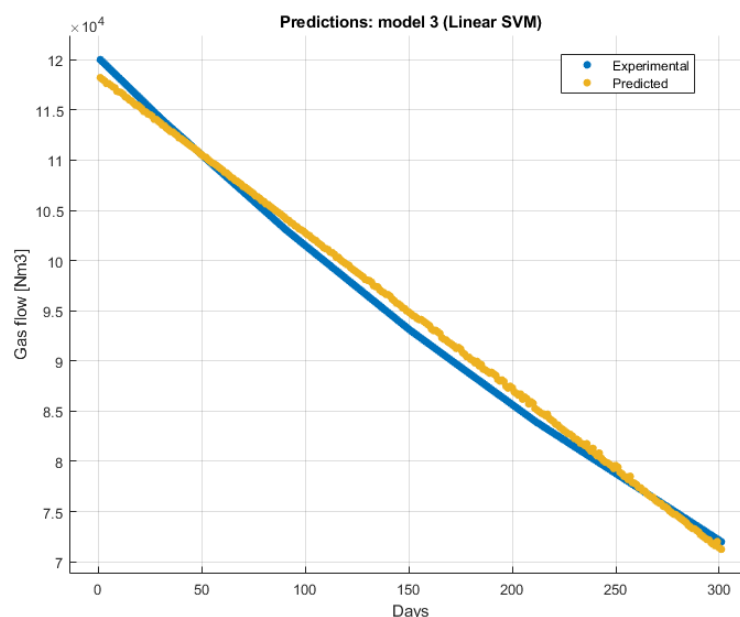


Figure 7. Comparison between the experimental flow values and the flow predicted using the SVM.

4.3. Artificial Neural Networks

Artificial Neural Networks (ANNs) are adaptive-learning systems that are inspired by the human brain. They are made up of interconnected nodes, called artificial neurons, organized in layers (input, hidden, and output), through which information is gradually processed. By adjusting the weights between neurons, neural networks can learn from examples, improving their performance as they are exposed to more data.

The ANN models were developed in MATLAB 2018 (MathWorks) using a feedforward neural network architecture. Two separate neural networks were constructed: one for pressure prediction and one for flow rate estimation.

The available dataset was first preprocessed and converted into a format compatible with MATLAB. The data were randomly divided into three subsets: 70% for training, 15% for validation, and 15% for testing. Model performance was evaluated exclusively on the independent test subset to ensure reliable generalization.

Several key network parameters were defined, including the training function, transfer functions, number of hidden layers, and number of neurons. The optimal network size was determined through preliminary trial-and-error analysis, as no general rule exists for selecting the optimal number of neurons for a specific problem.

The networks were trained using the Levenberg–Marquardt backpropagation algorithm, selected for its fast convergence and high accuracy. The Mean Squared Error (MSE) was used as the performance function, and the training process was automatically stopped based on validation performance to prevent overfitting. Other training parameters, such as the number of epochs and performance goals, were automatically adjusted by MATLAB.

The ANN developed for pressure estimation consisted of an input layer, one hidden layer with 10 neurons, and a single output neuron. The input layer received the predictor variable (time in days) and transmitted it to the hidden layers. The hidden layers were composed of interconnected neurons that transform the input information into higher-level feature representations. Each neuron performs a weighted summation of its inputs, adds a bias term, and applies a nonlinear activation function. This nonlinear processing enables the network to capture complex relationships between time and reservoir pressure behavior. Activation functions ReLU were used in the hidden layers.

For flow rate estimation, a similar feedforward architecture was used, consisting of four layers: one input layer, two hidden layers with 10 neurons in each hidden layer, and

one output layer. The same linear activation functions were applied in the hidden layers and a linear activation function in the output layer.

The variables predicted with the SVM are compared with experimental output datasets and are shown in Figures 8 and 9.

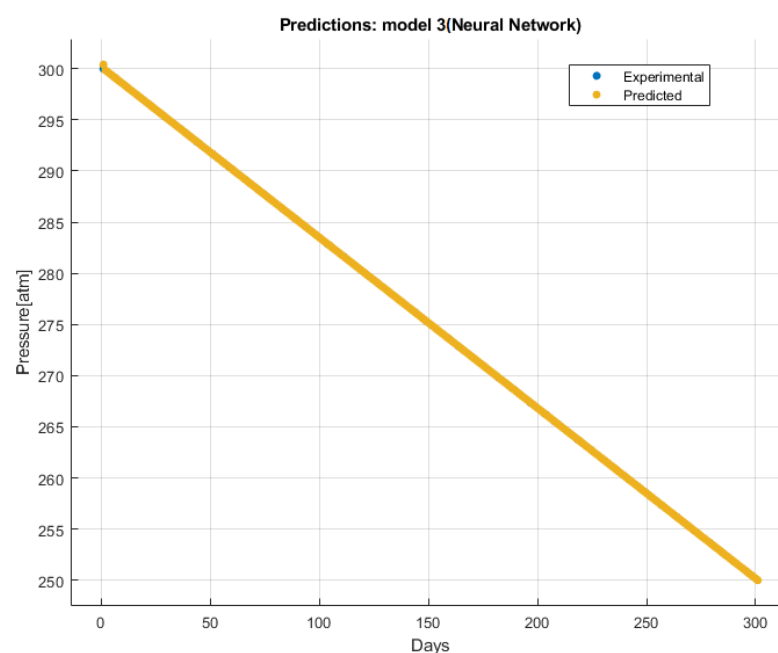


Figure 8. Comparison between the experimental pressure values and the pressure predicted using the ANN.

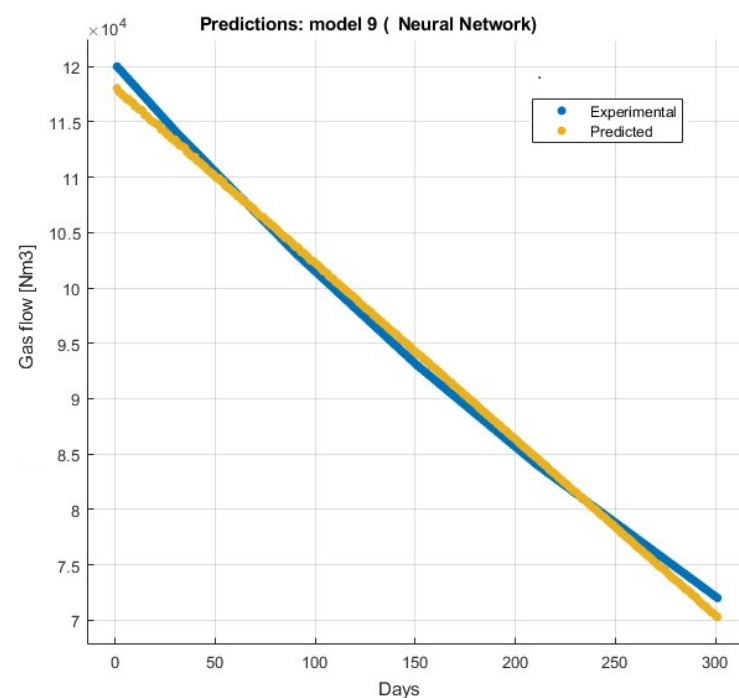


Figure 9. Comparison between the experimental flow values and the flow predicted using the ANN.

In Figures 10 and 11 and Table 2, a comparison between all methods is presented.

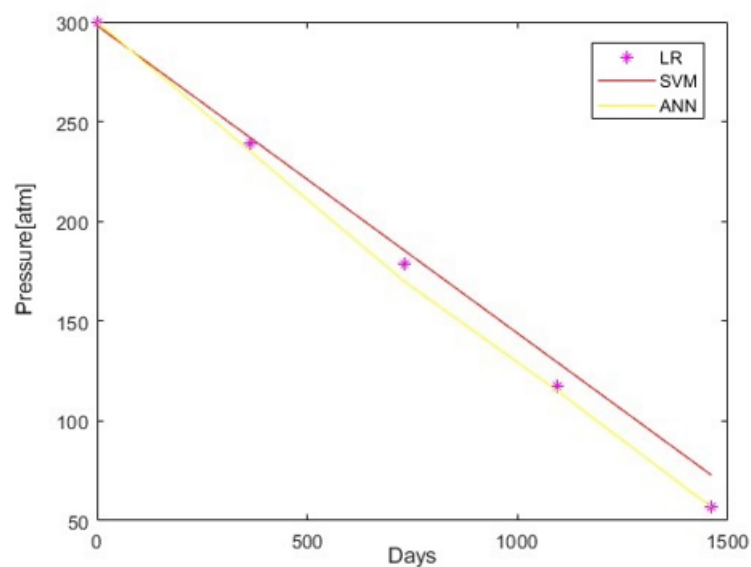


Figure 10. Comparison between applied machine-learning methods for pressure.

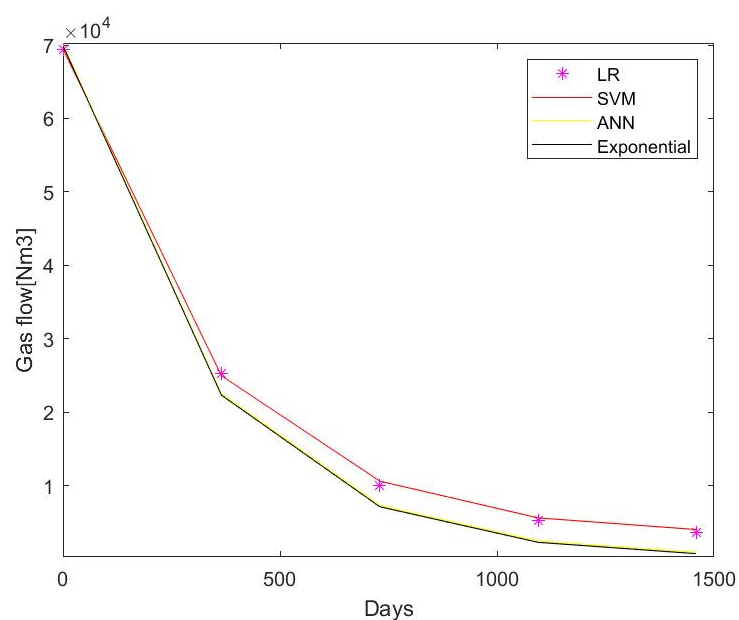


Figure 11. A comparison between applied machine-learning methods for flow.

Table 2. Experimental data vs. predicted data.

Day	LR	SVM	ANN	LR	SVM	ANN	Exponential
Pressure [atm]				Flow [Nm³]			
1	299.83	297.98	300	69,342	69,330	69,960	69,781
365	239.16	241.83	235	25,250	24,968	22,493	22,314
730	178.33	185.52	170	10,049	10,599	7293	7113
1095	117.50	129.21	115	5203	5599	2453	2267
1460	56.66	72.90	56.89	3658	4009	914	722

Based on the performance indices analyzed (Table 3), it can be concluded that Artificial Neural Networks (ANN) represent the most suitable machine-learning method for predicting future pressure and oil/gas production rates. ANN consistently achieves the

lowest values for MSE, RMSE, and MAE, indicating minimal prediction errors. In addition, the coefficient of determination $R^2 = 1$ demonstrates an excellent fit between the simulated and actual values.

Table 3. The performance indices.

Performance Indices	LR	SVM	ANN
Pressure [atm]			
MSE	7.45×10^{-6}	1.161	1.9×10^{-5}
R^2	1	0.99	1
RMSE	0.0027	1.077	0.004
MAE	0.0022	0.93	0.002
Flow [Nm ³]			
MSE	8.2×10^5	1.944×10^6	1.9126×10^6
R^2	0.99	0.99	1
RMSE	1383	1394	905
MAE	1182	1218	790

In comparison, linear Regression (LR) performs well in pressure prediction but does not maintain the same level of accuracy when estimating flow rates. SVM, on the other hand, provides the weakest performance among the three methods, with significantly higher error values.

Therefore, the results obtained show that ANN can capture the complex, nonlinear relationships characteristic of petroleum and gas reservoirs, making it the most robust and efficient technique for modeling and predicting production parameters. Implementing this method can contribute to optimizing reservoir engineering processes, reducing uncertainty, and supporting more accurate operational decision-making.

5. Conclusions

This study presents a systematic comparison between classical Decline-Curve Analysis (DCA) and data-driven machine-learning techniques for forecasting gas production rates and reservoir pressure, using a unified dataset from a mature Romanian gas field. By reinterpreting Arps decline parameters through multiple machine-learning algorithms, the research directly addresses a documented gap in the literature where empirical decline models and artificial intelligence methods are rarely evaluated side by side within a consistent framework.

The results demonstrate that Artificial Neural Networks (ANN) significantly outperform traditional Arps-based decline models as well as other machine-learning approaches, including Linear Regression (LR) and Support Vector Machines (SVM). ANN achieved the lowest prediction errors across all performance indices and exhibited an almost perfect agreement with experimental data ($R^2 = 1$) for both pressure and flow forecasting. This high predictive accuracy reflects the strong and consistent relationships present in the analyzed dataset. The results demonstrate the capability of ANN models to capture the complex nonlinear behavior associated with reservoir depletion, which is only partially represented by conventional empirical decline equations.

Among the machine-learning methods evaluated, Linear Regression performed adequately for pressure estimation but lacked robustness in flow prediction, particularly under nonlinear decline conditions. SVM yielded the weakest overall performance, with

higher prediction errors and sensitivity to data distribution. In contrast, ANN consistently provided stable and accurate predictions, highlighting its suitability for practical reservoir engineering applications.

From a broader perspective, the bibliometric analysis confirms that research integrating decline-curve analysis with machine learning remains sparse and fragmented. By embedding classical Arps models within an ANN-based predictive framework, this study effectively bridges an emerging research theme with a foundational production forecasting methodology, contributing to the modernization of decline-curve analysis within the digital oilfield paradigm.

The proposed ANN-assisted forecasting approach offers a practical, data-efficient alternative to complex reservoir simulation, requiring no additional field data while substantially reducing forecasting uncertainty. Its implementation can support improved decision-making in reserves estimation, production optimization, and field development planning.

Beyond the case study presented, the proposed methodology has broader applicability due to its data-driven nature and ability to model complex nonlinear relationships. The methodology may be applied to other petroleum engineering applications, such as estimation fluid and fuel properties. Furthermore, the methodology can be applied to the modeling and optimization of industrial processes where experimental data are available, including process performance prediction, equipment behavior analysis, and system monitoring.

These examples highlight the versatility of the proposed approach for a wide range of engineering problems involving experimental datasets and complex system dynamics.

Author Contributions: Conceptualization, D.-R.J. and C.R.P.; methodology D.-R.J. and C.R.P.; software, D.-R.J., C.R.P., M.T. and C.V.; validation D.-R.J. and C.R.P.; formal analysis, D.-R.J., C.R.P., M.T. and C.V.; investigation, D.-R.J., C.R.P., M.T. and C.V.; resources, D.-R.J., C.R.P., M.T. and C.V.; data curation, D.-R.J., C.R.P., M.T. and C.V.; writing—original draft preparation, D.-R.J., C.R.P., M.T. and C.V.; writing—D.-R.J., C.R.P., M.T., and C.V.; visualization, D.-R.J. and C.R.P.; supervision, D.-R.J. and C.R.P.; project administration, D.-R.J. and C.R.P.; funding acquisition, M.T. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding authors.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Rezaeipour, A.; Dabir, B.; Ahmadi, M. Data-Driven Approach to Dynamic Uncertainty Quantification of Oil Well Production: Decline Curve Analysis and Hidden Markov Model. *Geoenergy Sci. Eng.* **2023**, *229*, 212131. [\[CrossRef\]](#)
2. Manda, P.; Nkazi, D. The Evaluation and Sensitivity of Decline Curve Modelling. *Energies* **2020**, *13*, 2765. [\[CrossRef\]](#)
3. Aryady, C.V.O.; Lutfi, M. Decline Curve Analysis for Forecasting Cumulative Production and Remaining Recoverable Volume in Well 15/9-F-12, Volve Field. *Forum Tech. Mitt. ThyssenKrupp* **2026**, *2026*, 14–20. [\[CrossRef\]](#)
4. Li, K.; Guan, Y. Effective-Well Method for Matching and Predicting Oil Production. *Transp. Porous Media* **2012**, *91*, 377–389. [\[CrossRef\]](#)
5. Luo, S.; Su, H. Study on the Production Decline Characteristics of Shale Oil: Case Study of Jimusar Field. *Front. Energy Res.* **2022**, *10*, 845651. [\[CrossRef\]](#)
6. Vega-Ortiz, C.; Panja, P.; Deo, M.; McPherson, B. Decline Curve Analysis Using Machine Learning Algorithms: RNN, LSTM, and GRU. In *Proceedings of the 57th U.S. Rock Mechanics/Geomechanics Symposium, Atlanta, GA, USA, 25–28 June 2023*; ARMA: Atlanta, GA, USA, 2023; p. ARMA-2023-0287.
7. Mask, G.M.; Wu, X. Generating New Production Type Curves through Machine Learning Utilizing Dimensional Analysis. In *Proceedings of the SPE Oklahoma City Oil and Gas Symposium, Oklahoma City, OK, USA, 17–19 April 2023*; SPE: Oklahoma City, OK, USA, 2023; p. D031S010R003.

8. Li, W.; Dong, Z.; Lee, J.W.; Ma, X.; Qian, S. Development of Decline Curve Analysis Parameters for Tight Oil Wells Using a Machine Learning Algorithm. *Geofluids* **2022**, 2022, 8441075. [\[CrossRef\]](#)
9. Jha, H.S.; Khanal, A.; Lee, J. Statistical and Machine-Learning Methods Automate Multi-Segment Arps Decline Model Workflow to Forecast Production in Unconventional Reservoirs. In Proceedings of the SPE Canadian Energy Technology Conference, Calgary, AB, Canada, 16–17 March 2022; SPE: Calgary, AB, Canada, 2022; p. D011S012R003.
10. Gu, S.; Wang, J.; Xue, L.; Tu, B.; Yang, M.; Liu, Y. Deep-Learning-Based Production Decline Curve Analysis in the Gas Reservoir through Sequence Learning Models. *Comput. Model. Eng. Sci.* **2022**, 131, 1579–1599. [\[CrossRef\]](#)
11. Li, B.; Billiter, T.C.; Tokar, T. Rescaling Method for Improved Machine-Learning Decline Curve Analysis for Unconventional Reservoirs. *SPE J.* **2021**, 26, 1759–1772. [\[CrossRef\]](#)
12. Bakay, A.; Caers, J.; Mukerji, T.; Miller, P.; Cartier, C.; Briceno, A. Machine Learning of Spatially Varying Decline Curves for the Duvernay Formation. In Proceedings of the SPE Annual Technical Conference and Exhibition, Calgary, AB, Canada, 30 September–2 October 2019; SPE: Calgary, AB, Canada, 2019; p. D021S025R007.
13. Amr, S.; El Ashhab, H.; El-Saban, M.; Schietinger, P.; Caile, C.; Kaheel, A.; Rodriguez, L. A Large-Scale Study for a Multi-Basin Machine Learning Model Predicting Horizontal Well Production. In Proceedings of the SPE Annual Technical Conference and Exhibition, Dallas, TX, USA, 24–26 September 2018; SPE: Dallas, TX, USA, 2018; p. D011S008R005.
14. Vyas, A.; Datta-Gupta, A.; Mishra, S. Modeling Early Time Rate Decline in Unconventional Reservoirs Using Machine Learning Techniques. In Proceedings of the Abu Dhabi International Petroleum Exhibition & Conference, Abu Dhabi, United Arab Emirates, 13–16 November 2017; SPE: Abu Dhabi, United Arab Emirates, 2017; p. D041S113R002.
15. Cao, Q.; Banerjee, R.; Gupta, S.; Li, J.; Zhou, W.; Jeyachandra, B. Data Driven Production Forecasting Using Machine Learning. In Proceedings of the SPE Argentina Exploration and Production of Unconventional Resources Symposium, Buenos Aires, Argentina, 1–3 June 2016; SPE: Buenos Aires, Argentina, 2016; p. D021S006R001.
16. Kamari, A.; Mohammadi, A.H.; Lee, M.; Bahadori, A. Decline Curve Based Models for Predicting Natural Gas Well Performance. *Petroleum* **2017**, 3, 242–248. [\[CrossRef\]](#)
17. Li, Y.; Han, Y. Decline Curve Analysis for Production Forecasting Based on Machine Learning. In Proceedings of the SPE Symposium: Production Enhancement and Cost Optimisation, Kuala Lumpur, Malaysia, 7–8 November 2017; SPE: Kuala Lumpur, Malaysia, 2017; p. D011S004R003.
18. Hong, A.; Bratvold, R.B.; Lake, L.W.; Ruiz Maraggi, L.M. Integrating Model Uncertainty in Probabilistic Decline-Curve Analysis for Unconventional-Oil-Production Forecasting. *SPE Reserv. Eval. Eng.* **2019**, 22, 861–876. [\[CrossRef\]](#)
19. Bhattacharyya, S.; Vyas, A. Application of Machine Learning in Predicting Oil Rate Decline for Bakken Shale Oil Wells. *Sci. Rep.* **2022**, 12, 16154. [\[CrossRef\]](#) [\[PubMed\]](#)
20. Chen, M.; Qu, Z.; Liu, W.; Tang, S.; Shang, Z.; Ren, Y.; Han, J. Production Prediction Model of Tight Gas Well Based on Neural Network Driven by Decline Curve and Data. *Processes* **2024**, 12, 932. [\[CrossRef\]](#)
21. Mask, G.M.; Wu, X.; Nicholson, C. Enhanced Hydrocarbon Production Forecasting Combining Machine Learning, Transfer Learning, and Decline Curve Analysis. *Gas Sci. Eng.* **2025**, 134, 205522. [\[CrossRef\]](#)
22. Arps, J.J. Analysis of Decline Curves. *Trans. AIME* **1945**, 160, 228–247. [\[CrossRef\]](#)
23. Lewis, J.O.; Beal, C.H. Some New Methods for Estimating the Future Production of Oil Wells. *Trans. AIME* **1918**, 59, 492–525. [\[CrossRef\]](#)
24. Thompson, R.S.; Wright, J.D. *Oil Property Evaluation*; Thompson-Wright Associates Golden: Golden, CO, USA, 1985.
25. Raza, K. *Machine Learning in Single-Cell RNA-Seq Data Analysis*; SpringerBriefs in Applied Sciences and Technology; Springer Nature: Singapore, 2024; ISBN 978-981-9767-02-1.
26. Rostamzadeh-Renani, R.; Baghoolizadeh, M.; Mohammad Sajadi, S.; Pirmoradian, M.; Rostamzadeh-Renani, M.; Baghaei, S.; Salahshour, S. Prediction of the Thermal Behavior of Multi-Walled Carbon Nanotubes-CuO-CeO₂ (20-40-40)/Water Hybrid Nanofluid Using Different Types of Regressors and Evolutionary Algorithms for Designing the Best Artificial Neural Network Modeling. *Alex. Eng. J.* **2023**, 84, 184–203. [\[CrossRef\]](#)
27. Kanthasamy, R.; Almatrafi, E.; Ali, I.; Hussain Sait, H.; Zwawi, M.; Abnisa, F.; Choe Peng, L.; Victor Ayodele, B. Biochar Production from Valorization of Agricultural Wastes: Data-Driven Modelling Using Machine Learning Algorithms. *Fuel* **2023**, 351, 128948. [\[CrossRef\]](#)
28. Maulud, D.; Abdulazeez, A.M. A Review on Linear Regression Comprehensive in Machine Learning. *JASTT* **2020**, 1, 140–147. [\[CrossRef\]](#)
29. Huang, R.; Ma, C.; Ma, J.; Huangfu, X.; He, Q. Machine Learning in Natural and Engineered Water Systems. *Water Res.* **2021**, 205, 117666. [\[CrossRef\]](#) [\[PubMed\]](#)

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.