

UDC 62.5

CSCSTI 20.53.15

DOI: [10.54859/kjogi108887](https://doi.org/10.54859/kjogi108887)

Received: 30.06.2025.

Accepted: 07.04.2026.

Published: 30.06.2026.

## Review article

# Using Big Data and Analytics for Forecasting and Productivity Enhancement in the Oil and Gas Industry

**Aigerim B. Seitimbetova, Alevtina S. Shulgina-Tarashchuk, Aizhan S. Smailova**

*Karaganda Buketov University, Karaganda, Kazakhstan*

### ABSTRACT

Amid the digital transformation of the global economy, Big Data analytics and advanced analytics technologies are becoming key tools for enhancing business efficiency and sustainability. Their application is particularly relevant in the capital-intensive and high-risk oil and gas industry, where data-driven decision-making offers significant competitive advantages.

This study examines the opportunities and benefits of implementing Big Data and analytical solutions across various stages of the oil and gas production cycle – from geological exploration and drilling to processing and transportation. The study presents the main data sources and types characteristic of the industry, as well as modern analytical methods, including descriptive, predictive, prescriptive, and real-time analytics. Special attention is given to machine learning and artificial intelligence algorithms used for predicting equipment failures, optimizing drilling parameters, and modeling reservoir behavior.

Based on the analysis of case studies from leading international companies such as BP, Equinor, Gazprom Neft, and others, it is shown how digital tools can improve decision-making accuracy, reduce operational costs, and minimize technological risks. The study also examines the key challenges hindering the widespread adoption of Big Data in the sector, including a shortage of qualified personnel, integration difficulties between legacy and modern systems, cybersecurity concerns, and the high cost of digital transformation.

The analysis leads to the conclusion that data and analytics constitute strategic assets for the future development of the oil and gas industry. Digital technologies open up new horizons in forecasting, management, and sustainable production, paving the way for next-generation intelligent oil and gas enterprises.

**Keywords:** *Big Data; analytics; oil and gas industry; forecasting; productivity; digitalization.*

### To cite this article:

Seitimbetova AB, Shulgina-Tarashchuk AS, Smailova AS. Using Big Data and Analytics for Forecasting and Productivity Enhancement in the Oil and Gas Industry. *Kazakhstan journal for oil & gas industry*. 2026;8(2):48–58. DOI: [10.54859/kjogi108887](https://doi.org/10.54859/kjogi108887).

УДК 62.5

МРНТИ 20.53.15

DOI: [10.54859/kjogi108887](https://doi.org/10.54859/kjogi108887)

Получена: 30.06.2025.

Одобрена: 07.04.2026.

Опубликована: 30.06.2026.

**Научный обзор**

## Использование Big Data и аналитики для прогнозирования и повышения производительности в нефтегазовой отрасли

**А.Б. Сейтимбетова, А.С. Шульгина-Таращук, А.С. Смаилова**

*Карагандинский университет им. академика Е.А. Букетова, г. Караганда, Казахстан*

### АННОТАЦИЯ

В условиях цифровой трансформации мировой экономики технологии анализа больших данных (Big Data) и интеллектуальной аналитики становятся ключевыми инструментами для повышения эффективности и устойчивости бизнеса. Особенно актуальным их применение является в капиталоемкой и высокорисковой нефтегазовой отрасли, где решения, основанные на данных, способны обеспечить значительные конкурентные преимущества.

Настоящая статья посвящена исследованию возможностей и преимуществ внедрения Big Data и аналитических решений на различных этапах нефтегазового производственного цикла – от геологоразведки и бурения до переработки и транспортировки. Представлены основные источники и типы данных, характерные для отрасли, а также современные методы аналитики: описательный, прогностический, предписывающий и оперативный. Отдельное внимание уделено алгоритмам машинного обучения и искусственного интеллекта, применяемым для предсказания отказов оборудования, оптимизации параметров бурения и моделирования поведения залежей.

На основе анализа кейсов ведущих международных компаний, таких как BP, Equinor, Газпром нефть и др., показано, как цифровые инструменты позволяют повысить точность принятия решений, сократить издержки и минимизировать техногенные риски. Также рассматриваются ключевые вызовы, сдерживающие масштабное внедрение Big Data в отрасли: нехватка квалифицированных кадров, сложности интеграции старых и новых систем, вопросы кибербезопасности и высокая стоимость цифровизации.

В результате анализа формируется вывод о стратегической значимости данных и аналитики как об основном ресурсе будущего развития нефтегазовой промышленности. Цифровые технологии открывают новые горизонты в области прогнозирования, управления и устойчивого производства, формируя интеллектуальные нефтегазовые предприятия нового поколения.

**Ключевые слова:** Big Data, аналитика, нефтегазовая промышленность, прогнозирование, производительность, цифровизация.

### Как цитировать:

Сейтимбетова А.Б., Шульгина-Таращук А.С., Смаилова А.С. Использование Big Data и аналитики для прогнозирования и повышения производительности в нефтегазовой отрасли // Вестник нефтегазовой отрасли Казахстана. 2026. Том 8, №2. С. 48–58. DOI: [10.54859/kjogi108887](https://doi.org/10.54859/kjogi108887).

ӨОЖ 62.5

ҒТАХР 20.53.15

DOI: [10.54859/kjogi108887](https://doi.org/10.54859/kjogi108887)

Қабылданды: 30.06.2025.

Мақұлданды: 07.04.2026.

Жарияланды: 30.06.2026.

## Ғылыми шолу

# Мұнай-газ саласындағы өнімділікті болжау және жақсарту үшін Big Data пен аналитиканы пайдалану

**А.Б. Сейтімбетова, А.С. Шульгина-Тарашук, А.С. Смайылова**

*Академик Е.А. Бөкетов атындағы Қарағанды университеті, Қарағанды қаласы, Қазақстан*

## АҢДАТПА

Әлемдік экономиканың цифрлық трансформациясы жағдайында үлкен деректерді талдау (Big Data) және интеллектуалды аналитика технологиялары бизнестің тиімділігі мен тұрақтылығын арттырудың негізгі құралдарына айналууда. Бұл әсіресе капиталды көп қажет ететін және тәуекелі жоғары мұнай-газ саласында өзекті, өйткені деректерге негізделген шешімдер айтарлықтай бәсекелестік артықшылықтарды қамтамасыз ете алады.

Бұл мақала геологиялық барлау мен бұрғылаудан бастап өңдеу мен тасымалдауға дейінгі мұнай – газ өндірістік циклінің әртүрлі кезеңдерінде Big Data және аналитикалық шешімдерді енгізудің мүмкіндіктері мен артықшылықтарын зерттеуге арналған. Салаға тән деректердің негізгі көздері мен түрлері, сондай-ақ аналитиканың заманауи әдістері ұсынылған: сипаттамалық, болжамдық, нұсқамалық және жедел. Жабдықтың істен шығуын болжау, бұрғылау параметрлерін оңтайландыру және кен орындарының мінез-құлқын модельдеу үшін қолданылатын машиналық оқыту және жасанды интеллект алгоритмдеріне ерекше назар аударылады.

BP, Equinor, «Газпром нефть» және т. б. сияқты жетекші халықаралық компаниялардың жағдайларын талдау негізінде цифрлық құралдардың шешім қабылдау дәлдігін жақсартуға, шығындарды азайтуға және техногендік тәуекелдерді азайтуға қалай мүмкіндік беретіні көрсетілген. Сондай-ақ, салада Big Data-ның ауқымды енгізілуін тежейтін негізгі сын-қатерлер қарастырылууда: білікті кадрлардың жетіспеушілігі, ескі және жаңа жүйелерді интеграциялаудың күрделілігі, киберқауіпсіздік мәселелері және цифрландырудың қымбаттылығы.

Талдау нәтижесінде мұнай газ өнеркәсібінің болашақ дамуының негізгі ресурсы ретінде деректер мен талдаудың стратегиялық маңыздылығы туралы қорытынды жасалады. Цифрлық технологиялар болжау, басқару және тұрақты өндіріс саласында жаңа горизонттарды аша отырып, жаңа буынның интеллектуалды мұнай-газ кәсіпорындарын қалыптастыруға жол ашады.

**Негізгі сөздер:** *Big Data, талдау, мұнай-газ өнеркәсібі, болжау, өнімділік, цифрландыру.*

## Дәйексөз келтіру үшін:

*Сейтімбетова Ә.Б., Шульгина-Тарашук А.С., Смайылова А.С. Мұнай-газ саласындағы өнімділікті болжау және жақсарту үшін Big Data пен аналитиканы пайдалану // Қазақстанның мұнай-газ саласының хабаршысы. 2026. 8 том, №2, 48–58 б. DOI: [10.54859/kjogi108887](https://doi.org/10.54859/kjogi108887).*

## Introduction

In the context of the modern global economy, the oil and gas industry continues to play a key role in meeting the energy needs of industry, transportation, and the population. At the same time, it faces a number of serious challenges: the depletion of easily accessible hydrocarbon reserves, growing demands for environmental sustainability, the need to reduce operational costs, and the increasing pressure to improve extraction efficiency. To remain competitive and ensure sustainable development, oil and gas companies must seek new approaches to managing production processes, resources, and risks.

One of the most promising areas of industry transformation is the adoption of Big Data and data analytics technologies.

These tools enable companies to collect, process, and analyze massive volumes of information coming from various sources—drilling rigs, wells, sensors, satellites, production systems, and more. The data may include both structured information (e.g., sensor readings) and unstructured data (e.g., video footage, text-based reports, geophysical maps).

The use of advanced analytics, including machine learning and artificial intelligence (AI), allows companies not only to retrospectively analyze production processes but also to forecast reservoir behavior, determine optimal drilling parameters, detect equipment failure risks, and improve the accuracy of managerial decision-making. The implementation of Big Data in the oil and gas sector contributes to more precise reservoir modeling, reduced downtime, process automation, and improved technical and economic performance indicators.

Moreover, the development of the Industrial Internet of Things (IIoT), cloud computing, and digital twins is shaping a new digital environment in which data becomes a strategic asset. These technologies enable companies to shift from reactive to predictive and prescriptive management strategies, thereby achieving high productivity while reducing costs and environmental impact.

Thus, the use of Big Data and analytical technologies represents an integral component of the digital transformation of the oil and gas industry.

The aim of this article is to explore current capabilities and future prospects for applying Big Data and analytics to enhance forecasting and boost productivity. It will also analyze concrete examples of successful technology integration in practice.

The specific objective of this article is to examine the theoretical foundations and practical applications of Big Data and analytics in the oil and gas industry, with a focus on forecasting and productivity improvement tasks. In addition, a small-scale case study using data analysis and machine learning methods will be conducted to illustrate practical implementation [1].

## Literature review

The application of Big Data and analytics in the oil and gas industry is increasingly being explored both internationally and within Kazakhstan's scientific community. A number of publications in the Bulletin of the Oil and Gas Industry of Kazakhstan demonstrate progress in integrating digital solutions, machine learning algorithms, and intelligent analytics into key stages of extraction, monitoring, and geological exploration.

In the study by Zhenis, et al. [2], the authors examine the architecture of modern bottomhole pressure monitoring systems using machine learning algorithms. Particular attention is paid to real-time streaming processing of telemetry data from wells, the use of predictive analytics models, and the integration of Big Data infrastructures (Lambda and Kappa architectures). This research highlights the potential of intelligent well operation management in real time.

The article "Analysis of the well productivity decline in the Kashagan field" by Khassanov B.K. and Serniyazov Z.M. [3] analyzes the causes of declining well productivity, particularly scale formation. The authors link pressure and temperature data to model production dynamics, which aligns with the principles of predictive analytics.

Kolbikova E.S. [4] presents the application of machine learning methods in seismic and geophysical data analysis. Clustering of lithofacies types helps to model geological structures more accurately and predict reservoir properties more reliably.

The study by Bisikenov, et al. [5], demonstrates how tools for analyzing and interpreting PVT data and geochemical-geophysical measurements improve reserve estimation, which is essential for developing accurate mathematical models for production forecasting.

These studies show that Big Data analytics is used not only for monitoring extraction processes but also plays a critical role in exploration, planning, maintenance, and the development of digital models. Despite the diversity of topics, all these publications share a common idea: managing processes not after problems occur, but proactively, based on data. This approach helps reduce equipment downtime, increase production volumes, utilize resources more efficiently, and operate more sustainably.

## Materials and methods

Big Data refers to a set of technologies and methodologies for processing and analyzing extremely large, diverse, and rapidly growing datasets that traditional systems cannot efficiently handle. The key characteristics of Big Data are Volume, Velocity, Variety, Veracity, and Value.

Data analytics encompasses methods and tools used to extract meaningful insights from data, enabling informed and data-driven decision-making. In the oil and gas industry, analytics involves

statistical analysis, modeling, machine learning, and artificial intelligence (AI).

Core technologies include: Hadoop and Spark: frameworks for distributed processing of large-scale data, SQL and NoSQL database management systems: for storing and accessing data, Internet of Things (IoT): sensors and devices that collect real-time operational data, cloud platforms (AWS, Azure, Google Cloud): for scalable data infrastructure, machine Learning (ML) and Artificial Intelligence (AI): for forecasting, automation, and optimization.

**Data Collection and Integration.** In the oil and gas sector, data is gathered from a wide range of sources, including sensors on drilling rigs, wells, and pipelines, geolocation data, climate variables, operational records, and maintenance logs. IoT devices and SCADA systems enable real-time data acquisition, forming the foundation for timely analytics and operational control.

**Production Analysis and Forecasting.** The processing and analysis of Big Data enable the identification of patterns in extraction processes, optimization of equipment usage, forecasting of production output, and prediction of potential failures or accidents.

**Process Optimization.** Analytics contributes to cost reduction by forecasting equipment condition, optimizing maintenance schedules, improving planning quality, and minimizing downtime. The implementation of intelligent control systems enhances both productivity and safety.

**Data Description.** To illustrate practical applications, a simulated dataset representing oil production metrics from multiple wells over several years is used. The dataset includes parameters such as date, production volume, pressure, and temperature.

**Data Analysis Methods in the Oil and Gas Industry.** To leverage Big Data effectively, the oil and gas industry applies a wide array of analytical methods. These can be categorized based on their objectives and the types of data being analyzed, ranging from descriptive and diagnostic analytics to predictive and prescriptive approaches [6].

### Time Series Analysis

Oil and gas data often exhibit a temporal structure – parameters such as production volume, pressure, and temperature are measured periodically over time. Time series analysis allows companies to:

- Identify trends, such as increases or decreases in oil production over time;
  - Detect seasonal fluctuations and cycles, for example, the impact of climatic conditions on production;
  - Assess the effects of events, such as equipment maintenance or failures, on production metrics.
- Examples of time series methods include:
- moving Average – used to smooth data and reveal underlying trends;
  - autocorrelation – evaluates the relationship between data points at different time intervals;
  - Arima models (AutoRegressive Integrated Moving Average) – a statistical method for modeling and forecasting time series data.

```

main.py
1 import pandas as pd
2 from scipy.stats import pearsonr
3 import matplotlib.pyplot as plt
4 #Example data:** month, volume of water injection (m³), volume of oil production (tons)
5 data = {
6     "Month": ["Jan", "Feb", "Mar", "Apr", "May", "Jun"],
7     "Water_Injection": [12000, 13500, 15000, 14000, 16000, 15500],
8     "Oil_Production": [5000, 5400, 5800, 5600, 6000, 5900]
9 }
10 # Create DataFrame
11 df = pd.DataFrame(data)
12 # Calculating the Pearson correlation coefficient
13 corr_coef, p_value = pearsonr(df["Water_Injection"], df["Oil_Production"])
14 print("Pearson correlation coefficient:", round(corr_coef, 4))
15 print("P-value:", round(p_value, 4))
16 # Visualization
17 plt.scatter(df["Water_Injection"], df["Oil_Production"], color='green')
18 plt.title("Correlation between Water Injection and Oil Production")
19 plt.xlabel("Water Injection (m³)")
20 plt.ylabel("Oil Production (tons)")
21 plt.grid(True)
22 plt.tight_layout()
23 plt.show()

```

Figure 1. Program

Correlation Analysis. Correlation analysis helps identify relationships between various production and operational parameters, such as:

- the relationship between pressure and production volume;
- the impact of temperature on raw material quality;
- the link between equipment condition and the frequency of unplanned shutdowns.

Common methods include:

- Pearson correlation coefficient – measures linear relationships between two variables;
- spearman's rank correlation coefficient – assesses monotonic relationships that may not be linear.

**Results**

The results of correlation analysis help determine which factors most significantly affect productivity and operational efficiency [7].

Next, we consider a Python-based example for calculating the Pearson correlation coefficient in the oil production industry. The program analyzes the re-

lationship between the volume of water injection and oil production at a field (Fig. 1):

This program loads monthly data, calculates the Pearson correlation coefficient (using Pearson from scipy.stats module), displays the result (correlation value and p-value), and also generates a dependency plot (Fig. 2):

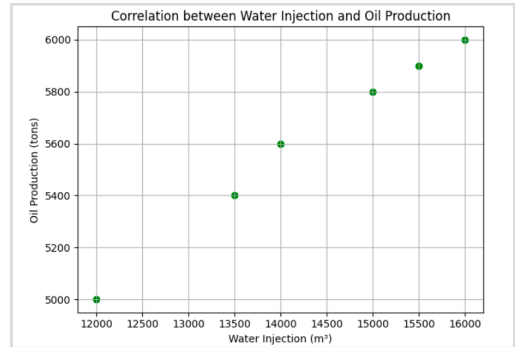


Figure 2. Plot

```

main.py
1 import pandas as pd
2 from sklearn.linear_model import LinearRegression
3 import matplotlib.pyplot as plt
4 from mpl_toolkits.mplot3d import Axes3D
5 # Sample data: water injection volume (m³), reservoir pressure (bar), oil production (tons)
6 data = {
7     "Water_Injection": [12000, 13000, 15000, 14000, 16000, 15500],
8     "Reservoir_Pressure": [230, 240, 250, 245, 260, 255],
9     "Oil_Production": [5000, 5300, 5800, 5600, 6100, 6000]
10 }
11 # Create a DataFrame
12 df = pd.DataFrame(data)
13 # Define independent variables (features) and dependent variable (target)
14 X = df[["Water_Injection", "Reservoir_Pressure"]]
15 y = df["Oil_Production"]
16 # Create and fit the linear regression model
17 model = LinearRegression()
18 model.fit(X, y)
19 # Predict oil production based on input features
20 df["Predicted_Production"] = model.predict(X)
21 # Output regression coefficients
22 print("Intercept:", model.intercept_)
23 print("Coefficients:", dict(zip(X.columns, model.coef_)))
24 # 3D visualization of actual vs. predicted oil production
25 fig = plt.figure()
26 ax = fig.add_subplot(111, projection='3d')
27 ax.scatter(df["Water_Injection"], df["Reservoir_Pressure"], df["Oil_Production"],
28           c='blue', label="Actual Production")
29 ax.scatter(df["Water_Injection"], df["Reservoir_Pressure"], df["Predicted_Production"],
30           c='red', label="Predicted Production")
31 # Set axis labels and title
32 ax.set_xlabel("Water Injection (m³)")
33 ax.set_ylabel("Reservoir Pressure (bar)")
34 ax.set_zlabel("Oil Production (tons)")
    
```

Figure 3. Multiple Linear Regression in Python

The graph presents a scatter plot where the horizontal axis (X) represents the volume of water injection, and the vertical axis (Y) represents the volume of oil production. Each point on the chart corresponds to a specific month from the dataset. The points lie very close to a straight line that slopes upward as water injection increases. This provides a visual confirmation of a strong positive linear correlation, as indicated by the Pearson correlation coefficient. The chart title, "Correlation between Water Injection and Oil Production", along with the axis labels, helps to clearly interpret what is being displayed.

Thus, the output of the code demonstrates a very strong and statistically significant positive linear relationship between water injection volume and oil production volume in the provided data, which is clearly illustrated in the scatter plot.

**Regression Analysis**

Regression is a method used to model the relationship between a dependent variable and one or more independent variables. In the oil and gas sector, regression analysis is often used to:

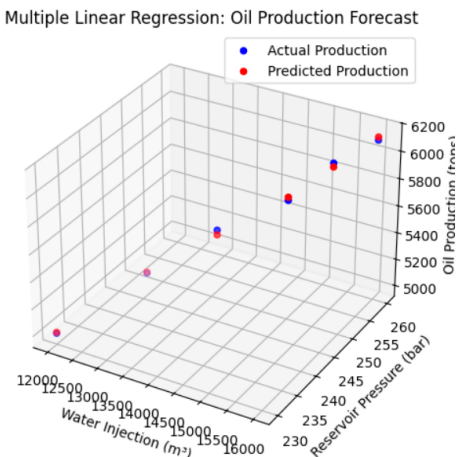
- forecast production volumes based on technical and operational parameters;
- evaluate the impact of drilling process changes on oil quality.

Types of regression include:

- linear regression - for modeling simple linear relationships;
- multiple regression - incorporates several influencing factors simultaneously;
- polynomial regression - used to capture more complex, nonlinear relationships.

Here is a Python program for regression analysis in the oil and gas industry. In this case, the model captures the relationship between oil production volume and two factors: water injection volume and reservoir pressure. The program demonstrates the use of multiple linear regression (Fig. 3).

The program also generates a 3D plot showing both the actual data points and the predicted values, and prints the regression coefficients, which help evaluate the contribution of each parameter (Fig. 4):



**Figure 4. Plot of Multiple Linear Regression**

**Machine Learning.** Machine learning methods are widely used for analyzing large datasets and uncovering complex patterns. These algorithms can be trained on historical data to make predictions or perform classifications.

Key approaches include:

- supervised learning: linear models, decision trees, random forests, and gradient boosting algorithms-used for both regression and classification tasks;
  - neural networks: applied for modeling complex relationships and time series data;
  - unsupervised learning: clustering methods (e.g., k-means) for grouping wells or processes with similar characteristics;
  - principal Component Analysis (PCA) is used for dimensionality reduction, enabling data visualization and the identification of key influencing factors.
- Machine learning enables the automatic detection of patterns, prediction of equipment failures, optimization of operational modes, and overall cost reduction [8].

**Failure Analysis and Predictive Maintenance**

By leveraging operational data such as vibration, temperature, and pressure, data analytics and machine learning models can identify early indicators of potential equipment failures. This allows for proactive maintenance planning, preventing unplanned downtime and accidents.

**Data Visualization**

Visualization is an essential component of Big Data analysis. Tools such as graphs, heat maps, and interactive dashboards support the following:

- intuitive understanding of complex data;
- quick identification of anomalies and trends;
- decision-making support across all levels of management [9].

**Application Example.** The following section (Fig. 5) presents a program that uses a linear regression model to analyze production dynamics and forecast oil output for the coming months. The program visualizes both the original data and the generated forecast. This solution can be adapted to real-world datasets from oil companies, such as daily/monthly production volumes, well parameters, and more.

The program outputs the forecasted production volumes for each of the upcoming months to the console, with a precision of two decimal places.

Thus, the program builds a simple linear regression model that captures the relationship between monthly oil production volume and time, and uses this model to forecast future values.

- Blue dots: These represent the actual oil production volumes for each of the 12 months. They illustrate how production levels varied from month to month;
- Green line: This is the fitted linear regression model. It represents the best-fit line that describes the relationship between production volume

```

main.py
1 import numpy as np
2 import pandas as pd
3 import matplotlib.pyplot as plt
4 from sklearn.linear_model import LinearRegression
5 # Пример данных: месяц и объем добычи (в тыс. тонн)
6 data = {
7     "Месяц": np.arange(1, 13), # 12 месяцев
8     "Добыча": [350, 360, 355, 370, 375, 380, 385, 395, 400, 405, 410, 415]
9 }
10 df = pd.DataFrame(data)
11 # Подготовка данных для модели
12 X = df["Месяц"].values.reshape(-1, 1)
13 y = df["Добыча"].values
14 # Обучение модели
15 model = LinearRegression()
16 model.fit(X, y)
17 # Прогноз на следующие 3 месяца
18 future_months = np.array([13, 14, 15]).reshape(-1, 1)
19 future_prediction = model.predict(future_months)
20 # Визуализация
21 plt.scatter(X, y, color='blue', label='Фактические данные')
22 plt.plot(X, model.predict(X), color='green', label='Модель')
23 plt.plot(future_months, future_prediction, 'r--', label='Прогноз')
24 plt.title("Прогноз добычи нефти")
25 plt.xlabel("Месяц")
26 plt.ylabel("Объем добычи (тыс. тонн)")
27 plt.legend()
28 plt.grid(True)
29 plt.show()
30
31 # Вывод прогнозов
32 for month, value in zip(future_months.flatten(), future_prediction):
33     print(f"Прогноз на месяц {month}: {value:.2f} тыс. тонн")
    
```

Figure 5. Oil production forecasting program

and month based on the given data. The line clearly reflects an overall upward trend in the dataset;

- Red dashed line with circles: This line displays the forecasted oil production volumes for the next three months (the 13th, 14th, and 15th months). The red circles mark the predicted values for each of these months, and the dashed line connects them, showing the projected trend in production according to the model;
- Title: “Oil Production Forecast” – Indicates the subject of the graph;
- Axis labels: “Month” and “Production Volume (thousand tons)” – Clarify the content of the horizontal and vertical axes, respectively. The horizontal axis represents the month number, while the vertical axis shows the production volume in thousands of tons;
- Legend – Explains the visual elements on the graph, specifying which components correspond to “Actual Data”, “Model”, and “Forecast”;

- Grid – Added to enhance readability and make it easier to interpret values along the axes (Fig. 6).

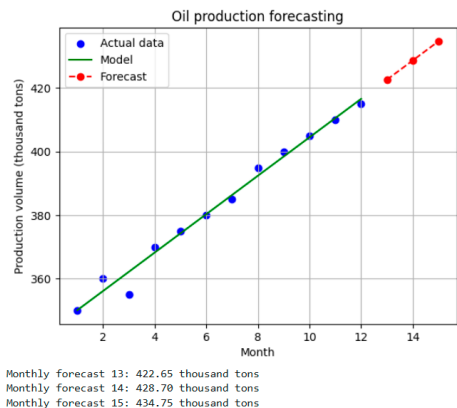


Figure 6. Oil production forecast

This 3D plot visualizes data based on month and production volume. To create a three-dimensional representation, the Z-axis corresponds to the index of each data point.

It is important to note that the Z-axis in this context does not represent any physical or statistically meaningful variable related to production or time.

The plot illustrates the spatial distribution of the data points in a three-dimensional space; however, the relationship between month and production volume is still best observed on the XY-plane (Fig. 7).

3D Визуализация данных (Месяц, Добыча, Индекс)

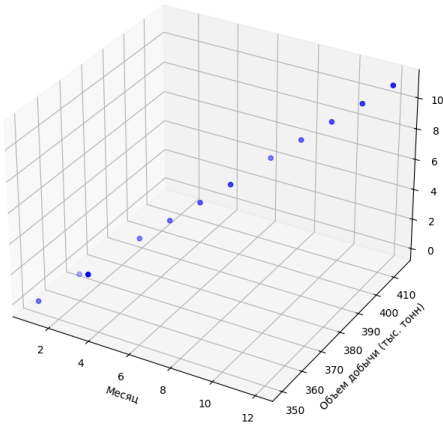


Figure 7. 3D data visualization (Month, Production, Index)

Overall, the chart clearly demonstrates how linear regression is used to model trends in the data and extrapolate them to generate forecasts for future periods.

## Conclusion

In the context of increasingly complex extraction processes, stricter environmental regulations, and the growing need for cost optimization, the oil and gas industry is undergoing a large-scale digital transformation. One of its key components is the implementation of Big Data and analytics technologies, which enable the extraction of valuable insights

from massive volumes of heterogeneous and high-frequency data.

An analysis of existing scientific and practical developments reveals that the use of intelligent analytics in the oil and gas sector covers several critically important areas:

- Forecasting oil and gas production using machine learning models and time series analysis;
  - Optimization of equipment performance and predictive maintenance, aimed at reducing the risk of failures and unplanned downtime;
  - Interpretation of geophysical and seismic data through clustering methods, neural networks, and regression analysis;
  - Enhancing field development efficiency through EOR (Enhanced Oil Recovery) technology screening, supported by databases and proxy models;
  - Digital modeling and the creation of digital twins, enabling more accurate simulation of development scenarios and more informed decision-making [10].
- Research by Kazakhstani scientists highlights the active development of solutions tailored to the local characteristics of oil fields. For example, the application of Lambda and Kappa architectures for bottomhole pressure monitoring, neural network methods for core sample analysis, and the deployment of platforms such as ABAI demonstrate the intellectual potential of national science and industry.

It is important to emphasize that the successful implementation of Big Data analytics requires not only technical solutions but also a cultural shift in decision-making practices within oil and gas companies. This includes building competencies in data analysis, fostering integration between IT and operational departments, and ensuring a reliable digital infrastructure.

In conclusion, Big Data and analytics are becoming indispensable tools for improving efficiency, safety, and sustainable development in the oil and gas industry. Companies that invest today in intelligent systems and digital platforms will gain a strategic advantage tomorrow—through cost reduction, increased productivity, and more agile business models. In the long term, such approaches contribute not only to commercial success but also to the creation of a smarter, greener, and more adaptive energy future.

## ADDITIONAL INFORMATION

**Funding source.** This study was not supported by any external sources of funding.

**Competing interests.** The authors declare that they have no competing interests.

**Authors' contribution.** All authors made a substantial contribution to the conception of the work, acquisition, analysis, interpretation of data for the work, drafting and revising the work, final approval of the version to be published and agree to be accountable for all aspects of the work. The greatest contribution is distributed as follows: Aigerim B. Seitimbetova – manuscript prepara-

## ДОПОЛНИТЕЛЬНО

**Источник финансирования.** Авторы заявляют об отсутствии внешнего финансирования при проведении исследования.

**Конфликт интересов.** Авторы декларируют отсутствие явных и потенциальных конфликтов интересов, связанных с публикацией настоящей статьи.

**Вклад авторов.** Все авторы подтверждают соответствие своего авторства международным критериям ICMJE (все авторы внесли существенный вклад в разработку концепции, проведение исследования и подготовку статьи, прочли

tion and editing; Alevtina S. Shulgina – analysis of the results; Aizhan S. Smailova – data collection and preparation.

и одобрили финальную версию перед публикацией). Наибольший вклад распределён следующим образом: Сейтимбетова А.Б. – формирование и редактирование рукописи; Шулгина-Тарашук А.С. – анализ полученных результатов; Смаилова А.С. – сбор и подготовка данных.

## REFERENCES

1. Nguyen TN, Gosine RG, Warriar P. A Systematic Review of Big Data Analytics for Oil and Gas Industry 4.0. *IEEE Access*. 2020;8:61183–61201. doi: [10.1109/ACCESS.2020.2979678](https://doi.org/10.1109/ACCESS.2020.2979678).
2. Zhenis DK, Kasenov AK, Ibrayev AY, Shayakhmet KN. A review of machine learning techniques for bottomhole pressure monitoring in production wells. *Kazakhstan journal for oil & gas industry*. 2025;7(2):61–72. doi: [10.54859/kjogi108797](https://doi.org/10.54859/kjogi108797).
3. Khassanov BK, Serniyazov ZM. Analysis of the well productivity decline in the Kashagan field. *Kazakhstan journal for oil & gas industry*. 2020;2(2):28–33. doi: [10.54859/kjogi95647](https://doi.org/10.54859/kjogi95647).
4. Kolbikova ES. Lithofacial analysis and possibilities for prediction of properties on geophysical research and seismic exploration data by methods of machine learning. *Kazakhstan journal for oil & gas industry*. 2021;3(4):32–37. doi: [10.54859/kjogi99690](https://doi.org/10.54859/kjogi99690).
5. Dukessova NK, Kunzharikova KM, Bissikenova LM, Bektas GZ. PVT Data Evaluation and Geochemical Fingerprinting: Approaches and Results. *Kazakhstan journal for oil & gas industry*. 2025;7(1):79–89. doi: [10.54859/kjogi108768](https://doi.org/10.54859/kjogi108768).
6. Alrabeh M, Abuzaid A. New Artificial Intelligence and Big Data Analytics Process to Enhance Non Metallic Pipe Deployments in Digital Oil Fields Using Workflows for Disparate Data Sets. Abu Dhabi International Petroleum Exhibition & Conference; 2020 Nov 9–12; Abu Dhabi, UAE. Available from: [onepetro.org/SPEADIP/proceedings-abstract/20ADIP/20ADIP/D041S114R003/452660](https://onepetro.org/SPEADIP/proceedings-abstract/20ADIP/20ADIP/D041S114R003/452660).
7. Giunta G, Bernasconi G, Giro RA, Cesari S. Digital Transformation of Historical Data for Advanced Predictive Maintenance // Abu Dhabi International Petroleum Exhibition & Conference; 2020 Nov 9–12; Abu Dhabi, UAE. Available from: [onepetro.org/SPEADIP/proceedings-abstract/20ADIP/20ADIP/D011S019R002/452657](https://onepetro.org/SPEADIP/proceedings-abstract/20ADIP/20ADIP/D011S019R002/452657).
8. Sletcha B, Vivas C, Saleh FK, et al. Digital Oilfield: Review of Real time Data flow Architecture for Upstream Oil and Gas Rigs // SPE International Conf. and Exhibition on Formation Damage Control; 2020 Feb 19–21; Lafayette, USA. Available from: [onepetro.org/SPEFD/proceedings-abstract/20FD/20FD/D021S013R005/446223](https://onepetro.org/SPEFD/proceedings-abstract/20FD/20FD/D021S013R005/446223).
9. Mahzari P, Emambakhsh M, Temizel C, Jones AP. Oil production forecasting using deep learning for shale oil wells under variable gas oil and water oil ratios. *Petroleum Science and Technology*. 2021;39(3):445–468. doi: [10.1080/10916466.2021.2001526](https://doi.org/10.1080/10916466.2021.2001526).
10. Gupta I, Samandarli O, Burks A, et al. Autoregressive and Machine Learning Driven Production Forecasting – Midland Basin Case Study // SPE/AAPG/SEG Unconventional Resources Technology Conference; 2021 July; Houston, USA. Available from: [chooser.crossref.org/?doi=10.15530%2Furtec-2021-5184](https://chooser.crossref.org/?doi=10.15530%2Furtec-2021-5184).

## СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ

1. *Nguyen T.N., Gosine R.G., Warriar P.* A Systematic Review of Big Data Analytics for Oil and Gas Industry 4.0 // *IEEE Access*. 2020. Vol. 8. P. 61183–61201. doi: [10.1109/ACCESS.2020.2979678](https://doi.org/10.1109/ACCESS.2020.2979678).
2. *Женис Д.К., Касенов А.К., Ибраев А.Е., Шаяхмет КН.* Машинное обучение в системах мониторинга забойного давления в эксплуатационных скважинах: обзор // *Вестник нефтегазовой отрасли Казахстана*. 2025. Т. 7, №2. С. 61–72. doi: [10.54859/kjogi108797](https://doi.org/10.54859/kjogi108797).
3. *Хасанов Б.К., Серниязов Ж.М.* Анализ снижения продуктивности скважин месторождения Кашаган // *Вестник нефтегазовой отрасли Казахстана*. 2020. Т. 2, №2. С. 28–33. doi: [10.54859/kjogi95647](https://doi.org/10.54859/kjogi95647).
4. *Колбикова Е.С.* Литофациальный анализ и возможности прогнозирования свойств по данным геофизических исследований и сейсморазведки методами машинного обучения // *Вестник нефтегазовой отрасли Казахстана*. 2021. Т. 3, №4. С. 32–37. doi: [10.54859/kjogi99690](https://doi.org/10.54859/kjogi99690).
5. *Дукессова Н.К., Кунзхарикова К.М., Биссикенова Л.М., Бектас Г.Ж.* Оценка данных PVT и геохимический фингерпринтинг: подходы и результаты // *Вестник нефтегазовой отрасли Казахстана*. 2025. Т. 7, №1. С. 79–89. doi: [10.54859/kjogi108768](https://doi.org/10.54859/kjogi108768).
6. *Alrabeh M., Abuzaid A.* New Artificial Intelligence and Big Data Analytics Process to Enhance Non Metallic Pipe Deployments in Digital Oil Fields Using Workflows for Disparate Data Sets // Abu Dhabi International Petroleum Exhibition & Conference; November 9–12, 2020; Abu Dhabi, UAE. Available from: [onepetro.org/SPEADIP/proceedings-abstract/20ADIP/20ADIP/D041S114R003/452660](https://onepetro.org/SPEADIP/proceedings-abstract/20ADIP/20ADIP/D041S114R003/452660).
7. *Giunta G., Bernasconi G., Giro R.A., Cesari S.* Digital Transformation of Historical Data for Advanced Predictive Maintenance // Abu Dhabi International Petroleum Exhibition & Conference; November 9–12, 2020; Abu Dhabi, UAE. Available from: [onepetro.org/SPEADIP/proceedings-abstract/20ADIP/20ADIP/D011S019R002/452657](https://onepetro.org/SPEADIP/proceedings-abstract/20ADIP/20ADIP/D011S019R002/452657).
8. *Sletcha B., Vivas C., Saleh F.K., et al.* Digital Oilfield: Review of Real time Data flow Architecture for Upstream Oil and Gas Rigs // SPE International Conf. and Exhibition on Formation Damage Control; February 19–21, 2020; Lafayette, USA. Available from: [onepetro.org/SPEFD/proceedings-abstract/20FD/20FD/D021S013R005/446223](https://onepetro.org/SPEFD/proceedings-abstract/20FD/20FD/D021S013R005/446223).
9. *Mahzari P., Emambakhsh M., Temizel C., Jones A.P.* Oil production forecasting using deep learning for shale oil wells under variable gas oil and water oil ratios // *Petroleum Science and Technology*. 2021. Vol. 39, Issue 3. P. 445–468. doi: [10.1080/10916466.2021.2001526](https://doi.org/10.1080/10916466.2021.2001526).
10. *Gupta I., Samandarli O., Burks A., et al.* Autoregressive and Machine Learning Driven Production Forecasting – Midland Basin Case Study // SPE/AAPG/SEG Unconventional Resources Technology Conference; July 2021; Houston, USA. Available from: [chooser.crossref.org/?doi=10.15530%2Furtec-2021-5184](https://chooser.crossref.org/?doi=10.15530%2Furtec-2021-5184).

**AUTHORS' INFO**

**\*Aigerim B. Seitimbetova**  
ORCID [0009-0000-8755-7992](https://orcid.org/0009-0000-8755-7992)  
e-mail: [sab.buketov.2022@gmail.com](mailto:sab.buketov.2022@gmail.com).  
**Alevtina S. Shulgina-Tarashchuk**  
ORCID [0009-0000-4759-9389](https://orcid.org/0009-0000-4759-9389)  
e-mail: [alevtinash79@mail.ru](mailto:alevtinash79@mail.ru).  
**Aizhan S. Smailova**  
ORCID [0000-0003-2936-0336](https://orcid.org/0000-0003-2936-0336)  
e-mail: [smailova.buketov@gmail.com](mailto:smailova.buketov@gmail.com).

**ИНФОРМАЦИЯ ОБ АВТОРАХ**

**\*Сейтимбетова Айгерим Байдуллаевна**  
ORCID [0009-0000-8755-7992](https://orcid.org/0009-0000-8755-7992)  
e-mail: [sab.buketov.2022@gmail.com](mailto:sab.buketov.2022@gmail.com).  
**Шульгина-Таращук Алевтина Сергеевна**  
ORCID [0009-0000-4759-9389](https://orcid.org/0009-0000-4759-9389)  
e-mail: [alevtinash79@mail.ru](mailto:alevtinash79@mail.ru).  
**Смаилова Айжан Сагиндыковна**  
ORCID [0000-0003-2936-0336](https://orcid.org/0000-0003-2936-0336)  
e-mail: [smailova.buketov@gmail.com](mailto:smailova.buketov@gmail.com).

\*Автор, ответственный за переписку / Corresponding Author