2-бөлім **Раздел 2** **Section 2**

Қолданбалы математика **Прикладная математика** **Applied mathematics**

IRSTI 30.19.21 DOI: https://doi.org/10.26577/JMMCS20251262012

Askar K. Kudaibergenov, Askat K. Kudaibergenov, T.B. Tileuberdi*
Al-Farabi Kazakh National University, Almaty, Kazakhstan
*e-mail: togzhantileuberdi@gmail.com

PREDICTION OF DRILL STRING VIBRATIONS USING MACHINE LEARNING TOOLS

The objective of this work is to apply machine learning algorithms for the analysis of dynamic vibrations of drill strings. As part of the research, a mathematical model describing the vibrations of the drill string was developed; a finite difference scheme was implemented, and numerical modeling was carried out using the three-point sweep method. Based on the data obtained from the numerical solution, a machine learning model was created. The numerical modeling was implemented using C++, while data collection and the construction of the machine learning model were performed using the Python programming language. As a result, a predictive model was obtained, capable of accurately forecasting the dynamic vibrations of the drill string and determining optimal parameters, thereby improving the efficiency and safety of drilling operations.
Key words: drill string, vibrations, machine learning, linear regression, random forest.

Аскар К. Кудайбергенов, Аскат К. Кудайбергенов, Т.Б. Тілеуберді*
Әл-Фараби атындағы Қазақ ұлттық университеті, Алматы, Қазақстан
*e-mail: togzhantileuberdi@gmail.com

Бұрғылау бағандарының тербелістерін машиналық оқыту құралдары арқылы болжау

Бұл жұмыстың мақсаты бұрғылау бағандарының динамикалық тербелістерін талдау үшін машиналық оқыту алгоритмдерін қолдану. Зерттеу барысында бұрғылау бағанының тербелістерін сипаттайтын математикалық модель әзірленді; ақырлы-айырымдық сұлбасы құрылып, үш нүктелі қуалау әдісін қолдану арқылы сандық модельдеу жүргізілді. Сандық шешімнен алынған мәліметтерге сүйене отырып, машиналық оқыту моделі құрылды. Сандық модельдеу үшін C++ тілінде бағдарламалық код жасалды, ал деректерді жинау және машиналық оқыту моделін құру процестері Python бағдарламалау тілі арқылы жүзеге асырылды. Нәтижесінде бұрғылау бағанасыныў динамикалық тербелістерін жоғары дәлдікпен болжай алатын және оұтайлы параметрлерді анықтайтын болжамды модель алынды, бұл бұрғылау жұмыстарының тиімділігі мен қауіпсіздігін арттыруға ықпал етеді.
Түйін сөздер: бұрғылау бағаны, тербеліс, машиналық оқыту, сызықты регрессия, кездейсоқ орман.

Аскар К. Кудайбергенов, Аскат К. Кудайбергенов, Т.Б. Тілеуберді*
Казахский национальный университет имени аль-Фараби, Алматы, Казахстан
*e-mail: togzhantileuberdi@gmail.com

Прогнозирование колебаний бурильных колонн с применением инструментов машинного обучения

Целью данной работы является применение алгоритмов машинного обучения для анализа динамических колебаний бурильных колонн. В рамках проведённого исследования была разработана математическая модель, описывающая колебания бурильной колонны; реализована конечно-разностная схема и проведено численное моделирование с использованием метода трёхточечной прогонки. Основываясь на данных, полученных в результате численного решения, была создана модель машинного обучения. Для численного моделирования был разработан программный код на языке C++, а процессы сбора данных и построения модели машинного обучения были выполнены с использованием языка программирования Python.

В результате была получена прогнозная модель, способная с высокой точностью предсказывать динамические колебания бурильной колонны и определять оптимальные параметры, что способствует повышению эффективности и безопасности проводимых буровых операций.
Ключевые слова: бурильная колонна, колебания, машинное обучение, линейная регрессия, случайный лес.

1 Introduction

Oil and gas constitute a significant portion of the world's energy resources, and the drilling industry plays a key role in their extraction. Drilling is a complex technological process that involves penetrating rock formations to access hydrocarbon reservoirs. During the drilling process, various types of vibrations frequently occur in the drill string, which are an undesirable side effect. Excessive vibrations can damage drilling tools, cause premature equipment wear, reduce drilling efficiency, and increase overall costs. Therefore, monitoring, analyzing, and minimizing drill string vibrations have become a critical area of research in the drilling industry.

Machine learning, as part of modern technological advancements, offers extensive opportunities for addressing this challenge. These methods enable the analysis of large datasets, the identification of hidden patterns, and the prediction of complex system behavior in real time. In recent years, numerous studies have been conducted on the application of machine learning for vibration analysis and control in drill string operations.

For instance, Saadeldin et al. [1] developed machine learning models for detecting drill string vibrations during horizontal drilling using surface sensor data. The study utilized radial basis functions (RBF), support vector machines (SVM), adaptive neuro-fuzzy inference systems (ANFIS), and functional networks (FN). These models successfully identified axial, torsional, and lateral vibrations with high accuracy, achieving correlation coefficients above 0.9 and a mean absolute percentage error (MAPE) of less than 7.5%. The results demonstrated that using surface data for vibration monitoring can significantly reduce costs by eliminating the need for expensive downhole sensors.

Another study by Saadeldin et al. [2] focused on predicting vibrations during the drilling of curve sections using real field data from multiple wells. The models were built using the same algorithms (RBF, SVM, ANFIS, and FN), but the emphasis was placed on curved well sections. The models showed impressive results, with ANFIS and SVM models achieving correlation coefficients of up to 0.99 and an error rate of less than 2.8%. The study confirmed that machine learning applications can significantly improve drilling performance in challenging conditions.

In the work [3], the authors investigated the application of different neural network architectures for predicting drill string vibrations. The tested models included fully connected networks, physics-informed neural networks, and long short-term memory (LSTM) networks. The results indicated that incorporating physical constraints into neural network structures enhanced prediction reliability, particularly in the context of nonlinear interactions between drilling equipment and rock formations.

Etaje [4] employed principal component analysis (PCA) and decision tree algorithms to identify optimal drilling zones with minimal vibrations. The proposed approach allowed for real-time adjustments of drilling parameters, thereby minimizing vibration-related risks. Notably, the study demonstrated high efficiency when using only surface data without requiring additional downhole measurements.

Hegde et al. [5] developed a model for classifying stick-slip vibrations using a random forest algorithm. The model was trained on historical data, including drill string rotation speed, bit load, and torque measurements. During testing, the model achieved 90% accuracy, highlighting its potential for integration into rate of penetration (ROP) optimization systems to improve the safety and efficiency of drilling operations. The authors in [6] investigated the application of machine learning for predicting drilling complications in oil and gas wells. Using historical data from 67 wells, they analyzed key drilling parameters such as standpipe pressure, hook load, rotary table torque, and rate of penetration to classify potential risks. Eight machine learning algorithms with gradient boosting (GB) demonstrating the highest accuracy in anomaly detection were tested. The study concluded that ML-based models can significantly enhance drilling efficiency by providing early warnings of complications, reducing non-productive time, and optimizing decision-making for drilling engineers. Future work suggests integrating geomechanical parameters to improve prediction accuracy. The authors in [7] developed a new machine learning-based model for predicting the rate of penetration (ROP) in vertical wells. The study compared physic-based models and data-driven methods, concluding that data-driven techniques provide better accuracy.

Despite significant progress in applying machine learning to analyze drill string vibrations, research in this area remains relevant. The complex dynamic processes that arise during interactions between drilling equipment and rock formations require the development of more accurate models capable of considering a wide range of influencing factors. An essential task is to create efficient algorithms capable of reliably predicting vibrations in real production environments and assisting operators in making informed drilling decisions.

The outcome of this work is a predictive model capable of accurately forecasting drill string vibrations and determining optimal drilling parameters. The implementation of this model contributes to improving the efficiency and safety of drilling operations by reducing downtime, minimizing accident risks, and optimizing operational parameters in real-time production conditions.

2 Materials and methods

2.1 Linear mathematical model

The linear model of the dynamics of a rotating drill string, compressed from both ends with a longitudinal load $N(x,t)$, is given by [8]:

$$\rho S\frac{\partial^2 u}{\partial t^2} + EI_{x_1}\frac{\partial^4 u}{\partial x^4} - \rho I_{x_1}\frac{\partial^4 u}{\partial x^2 \partial t^2} + \frac{\partial}{\partial x}\left(N(x,t)\frac{\partial u}{\partial x}\right) - \rho S\Omega^2 u = 0. \tag{1}$$

The parameters of the linear model are as follows: ρ is the density of the drill string material, which is typically made of steel or duralumin; S is the cross-sectional area of the transverse section; E is Young's modulus, a physical quantity that characterizes the elastic properties of the material, defining the relationship between stress and strain during tension or compression; I is the principal moment of inertia, describing the distribution of mass around the axis of rotation; $N(x,t)$ is the axial load that occurs during drilling, directed along the drill string and influenced by the string's weight, tensile forces, soil resistance, and dynamic loading; Ω is the angular speed of the drill string, representing its rotation around its own axis.

Taking into account the method of fixing the upper end of the drill string and its interaction with the rock at the bottom, as well as the fact that only rotational motion is possible while longitudinal and transverse movements are constrained, the boundary conditions are presented as follows:

$$u(x,t) = 0 \quad (x = 0, x = l), \tag{2}$$

$$EI_{x_1}\frac{\partial^2 u(x,t)}{\partial x^2} = 0 \quad (x = 0, x = l). \tag{3}$$

These conditions correspond to a hinged support.

The initial conditions are defined as follows:

$$u(x,t) = 0 \quad (t = 0), \tag{4}$$

$$\frac{\partial u(x,0)}{\partial t} = C_1 \quad (t = 0), \tag{5}$$

where C_1 represents the displacement rate of the drill string in the Ox_1x_3 plane at the initial moment in time.

The numerical solution of equation (1) was obtained by discretizing it in a finite difference form. The finite difference scheme of equation (1) is given as follows:

$$\begin{aligned}
&\rho S\frac{u_i^{n+1} - 2u_i^n + u_i^{n-1}}{\Delta t^2} + EI_{x_1}\frac{u_{i-2}^n - 4u_{i-1}^n + 6u_i^n - 4u_{i+1}^n + u_{i+2}^n}{\Delta x^4} \\
&- \rho I_{x_1}\frac{u_{i+1}^{n+1} - 2u_i^{n+1} + u_{i-1}^{n+1} - 2(u_{i+1}^n - 2u_i^n + u_{i-1}^n) + u_{i+1}^{n-1} + 2u_i^{n-1} - u_{i-1}^{n-1}}{\Delta t^2\Delta x^2} \\
&+ \frac{1}{4\Delta x^2}\left(N_{i+1}^n(u_{i+1}^n - u_i^n) - N_i^n(u_i^n - u_{i-2}^n)\right) - \rho S\Omega^2 u_i^{n+1} = 0.
\end{aligned} \tag{6}$$

Since the scheme is semi-implicit, the three-point Thomas algorithm was applied.

2.2 Application of machine learning

2.2.1 Data collection and processing of vibration information

Machine learning data is a fundamental component of the training process for machine learning models and algorithms. The data was collected through numerical solutions of the mathematical model of drill string vibrations. Three parameters of the drilling process — initial velocity, longitudinal load, and angular velocity — were used as features, while the maximum vibration value at the exact center of the drill string was taken as the data point. The dataset consists of 300 data points and 3 features. The min-max normalization method was used to avoid difficulties when building the model. Data collection and processing were implemented using the Python programming language. The collected data was split into 75% training data and 25% test data using the train_test_split function from the Scikit-learn library. Figures 1 and 2 illustrate the training and test datasets.

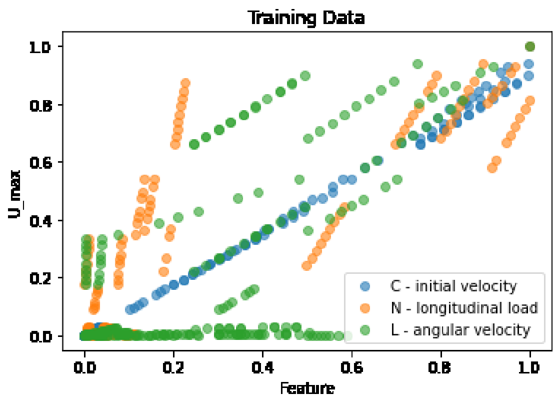


Figure 1: Training dataset

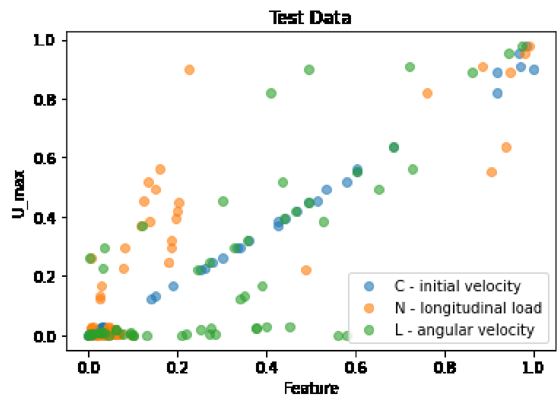


Figure 2: Test dataset

2.2.2 Random forest

The Random Forest algorithm is a machine learning method that uses an ensemble of decision trees for classification and regression tasks. The Random Forest algorithm operates as follows:

1. A random subset of objects (bootstrap sample) and a random subset of features (feature subset) are selected from the training dataset.

2. For each selected feature subset, a decision tree is built using an information criterion.

3. Steps 1 and 2 are repeated k times, where k is the number of trees in the forest.

4. For classifying a new object, each tree in the forest makes a prediction, and the final decision is determined by majority voting.

5. For regression tasks, each tree predicts a value, and the final prediction is calculated by averaging the results across all trees.

To build the Random Forest model, the number of trees (the `n_estimators` parameter) was set to 100. These trees are built independently of each other, and the algorithm selects a random subset of features for constructing each tree. To build a tree, a bootstrap sample is first created from the training dataset: from `n_samples` examples, `n_samples` examples are randomly selected with replacement. As a result, the sample has the same size as the original dataset, but some examples may be missing, while others may appear multiple times.

Next, a decision tree is built based on the generated bootstrap sample. The node-splitting algorithm selects a subset of features at each split and determines the best split using one of the selected features. The number of features to consider is controlled by the `max_features` parameter [9].

Let us visualize the decision boundaries of the first 10 trees and the aggregated prediction provided by the Random Forest model. Each plot in the graph corresponds to a decision tree trained on the training dataset. Including all three features results in a 3D graph (Figure 3).The model built using all decision trees is shown in Figure 4.

The "feature importance"function in the Random Forest algorithm was used to determine the importance of each feature. This metric indicates how significant each feature is in the decision-making process. The importance of the features in the dataset is shown in Figure 5.

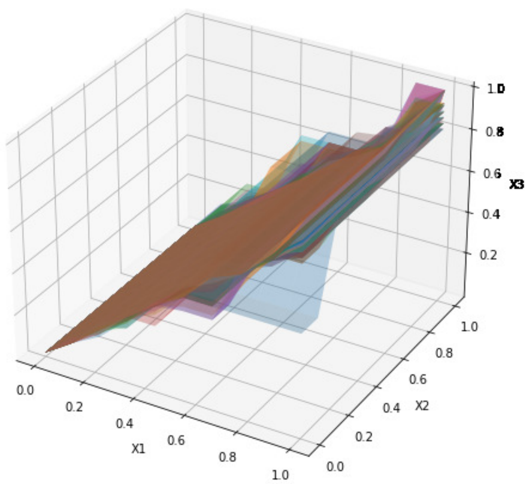


Figure 3: Decision trees model

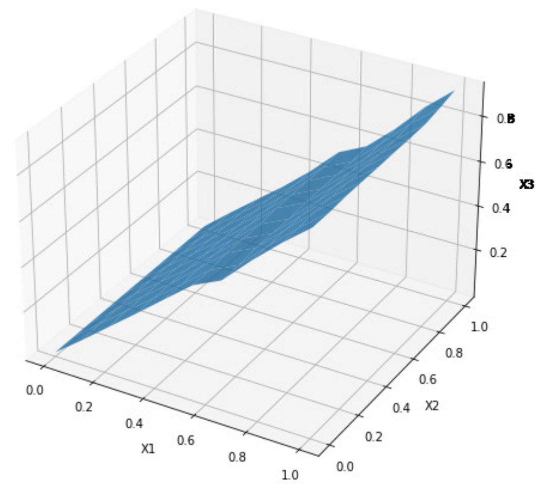


Figure 4: Random Forest model

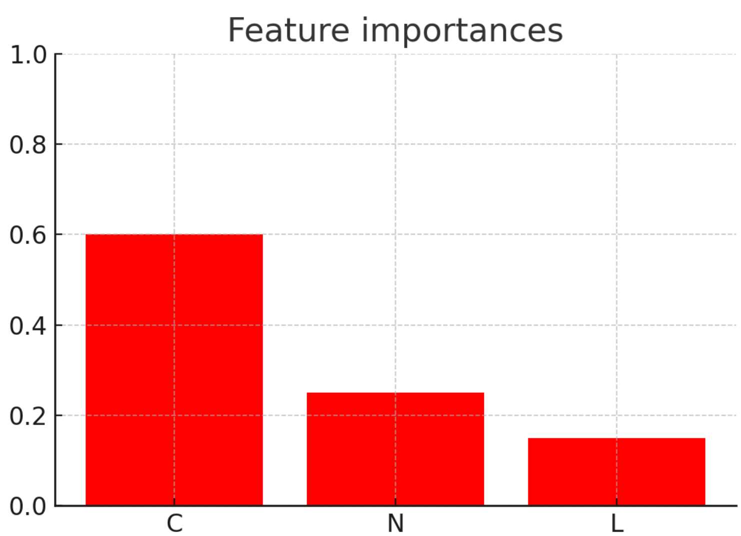


Figure 5: Visualization of feature importance

2.2.3 Linear regression

Linear regression is a machine learning method used to predict numerical values based on a linear relationship between features and the target variable. The general prediction formula for linear regression is as follows:

$$\hat{y} = w[0] \cdot x[0] + w[1] \cdot x[1] + w[p] \cdot x[p] + b \tag{7}$$

where $x[0]$ to $x[p]$ are the features. In our case, $p = 3$. w and b are the model parameters, while $\hat{y}$ is the prediction generated by the model [10].

The model is trained by finding the optimal coefficients w and b that minimize the mean squared error (MSE) between the predicted and actual target values. The parameters are adjusted to minimize the MSE, ensuring the best possible predictive performance. The 3D linear regression model is illustrated in Figure 6.

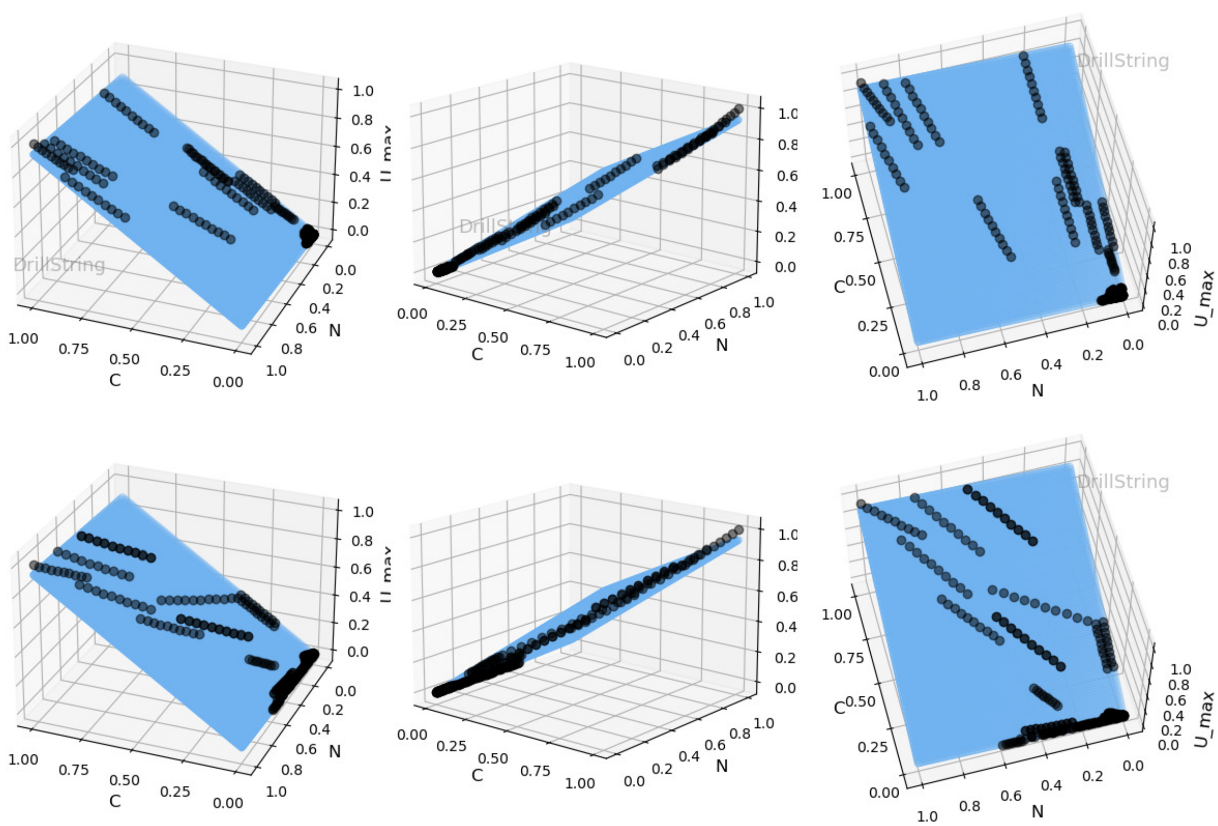


Figure 6: The 3D linear regression model

Figure 7 shows the visualization of the linear model predicting the initial velocity based on vibration values.

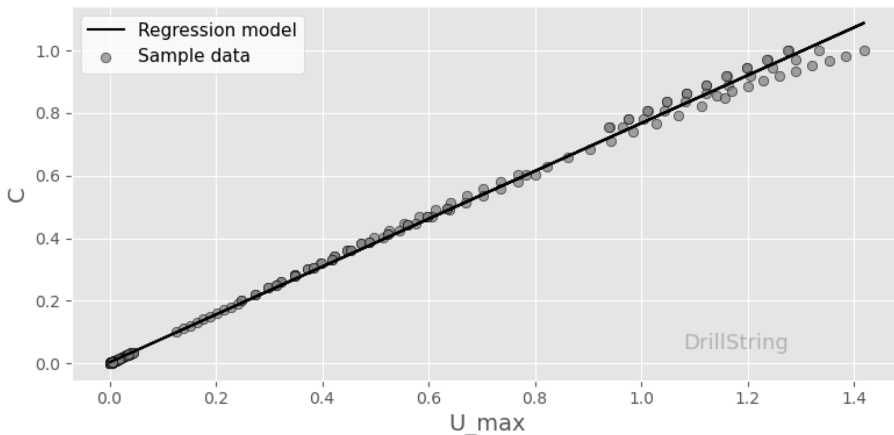


Figure 7: Visualization of the linear model

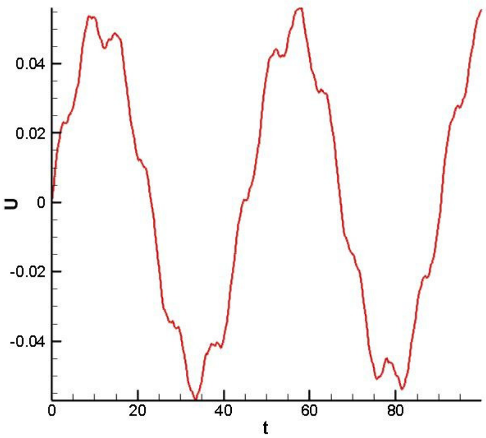
3 Results and discussions

The parameters obtained from [8] in Table 1 were used to obtain the results of the numerical solution of the linear mathematical model.

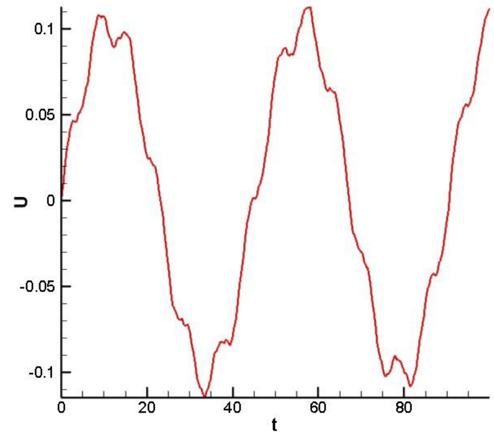
Table 1: Parameter values of the drilling system

Young's modulus	$2.1 \times 10^{11}\ Pa$
Drill string density	$7800\ kg/m^3$
Drill string length	$200\ m$
Inner diameter of the drill string	$0.12\ m$
Outer diameter of the drill string	$0.20\ m$
Cross-sectional area of the string	$2.1 \times 10^{-2}\ m^2$
Angular velocity of the string	$0.083\ rad/s$
Longitudinal load	$1200\ N$
Initial velocity	$0.01\ m/s$
Spatial step	$0.1\ m$
Time step	2×10^{-5}

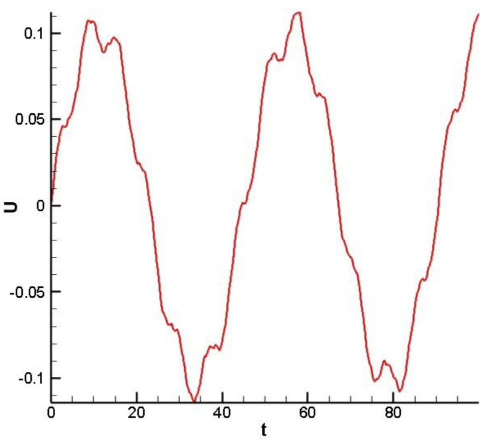
Figure 8 shows the vibration graphs at different parts of the drill string. As a result, we observe that the largest vibrations occur in the middle of the drill string.This is due to the resonance vibrations caused by the interaction of the elastic properties of the drill string material, gyroscopic forces, and external loads, including friction against the wellbore walls. The maximum vibration amplitudes occur in the middle of the drill string due to the distribution of mass, length, and boundary fixation conditions [11].



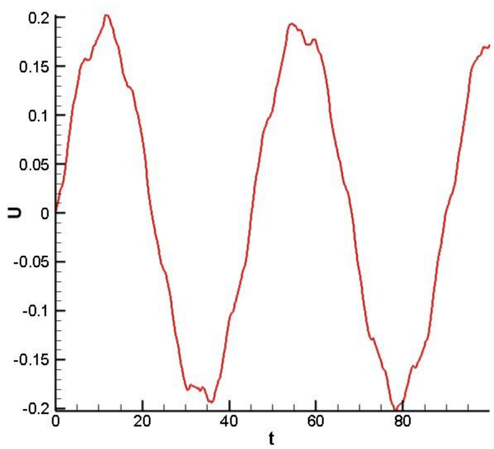
a.The time variation of $U\left[\frac{l}{2}\right]$



b.The time variation of $U\left[\frac{l}{4}\right]$



c.The time variation of $U\left[\frac{3l}{4}\right]$



d.The time variation of $U\left[\frac{l}{3}\right]$

Figure 8: The time variation of displacement U at different cross-sections

3.1 Evaluation of machine learning model performance

To evaluate the performance of the developed models, 25% of the dataset was set aside as test data before model training. This test dataset will be used to assess the model's performance by generating predictions for each test sample based on the provided features. The predicted values will then be compared with the actual observed values. For this purpose, the `score` method of the trained model object will be applied. Figures 9, 10 illustrate the comparison of the Random Forest model and the Linear Regression model with the test dataset.

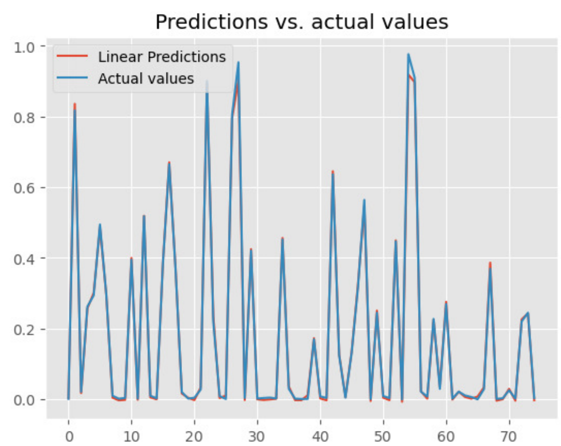


Figure 9: Comparison of the linear regression model with actual data

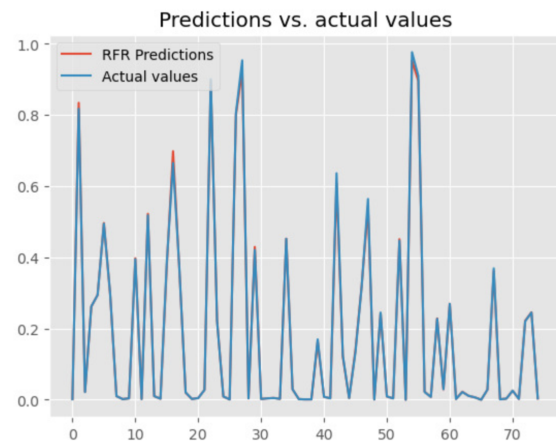


Figure 10: Comparison of the random forest model with actual data

As observed, the predicted values closely align with the actual values, with only minor discrepancies in certain areas. The x-axis represents the 75 test samples, while the y-axis corresponds to their respective values. The results indicate that the model demonstrates high accuracy in predicting vibration characteristics.

Table 2 presents a comparison of the results obtained from the Random Forest model, the Linear Regression model, and the numerical solution of the mathematical model under various conditions.

Table 2: Comparison of model predictions with actual values

C	N	L	Random Forest U_{max}	Linear Regression U_{max}	Actual Value U_{max}
0.46	3000	0.15	0.28344	0.286592	0.28610
0.30	2800	0.05	0.18170	0.18549	0.18639
0.15	3800	0.33	0.09693	0.09369	0.09355
0.70	2500	0.25	0.44384	0.438285	0.435873

Table 3 presents the results of the second linear regression analysis, where the initial velocity serves as the independent variable and the vibration amplitude as the dependent variable. Additionally, the table includes the corresponding results obtained from the mathematical model.

Table 3: Comparison of the model's predicted initial velocity with the actual value

U_max	Linear Regression, C	Actual Value, U_max
0.33	0.5	0.31
0.0025	0.0093	0.005
0.085	0.13	0.084
0.22	0.34	0.21
0.75	1.14	0.70

The linear regression model demonstrated better performance than the random forest model, making it the preferred choice for this task. Random forest offers high accuracy, robustness to large datasets, and feature importance evaluation but suffers from complex interpretability and high computational cost. Linear regression, on the other hand, provides simplicity, efficiency, and clear result interpretation while being sensitive to outliers and limited in capturing nonlinear relationships. Given its reliable performance on the given dataset, linear regression is recommended for predicting drill string vibrations.

4 Conclusion

The results of this study demonstrate that the application of machine learning techniques significantly improves the accuracy of drill string vibration predictions. Three models were developed and tested. Two models were designed to predict the maximum vibration amplitude based on input parameters: one using the random forest algorithm and the other using linear regression. Comparative analysis showed that linear regression performed better on new data due to the linear dependency between initial velocity and maximum vibration amplitude. However, with larger datasets, the performance of linear regression may decline, making the random forest algorithm more suitable for such cases. The third model predicts the initial velocity required to keep the vibration amplitude below a specified threshold, aiding in optimizing operational parameters and ensuring wellbore stability. Linear regression was chosen for this model due to its simplicity and accuracy with the given dataset. The use of machine learning techniques in this context offers several advantages, including faster and more accurate predictions, process optimization through reduced preparation time, and automated data analysis for real-time decision-making. However, the models require high-quality data and sufficient sample size for reliable performance, which can be challenging when dealing with complex drilling conditions. Future work could involve expanding the dataset with more diverse parameters and exploring advanced models to capture nonlinear relationships more effectively. Overall, the study confirms the potential of machine learning techniques for predicting drill string vibrations and optimizing drilling operations in the oil and gas industry.

5 Acknowledgement

This research is funded by the Science Committee of the Ministry of Science and Higher Education of the Republic of Kazakhstan (Grant No. AP23490543).

References

[1] Saadeldin R., Gamal H., Elkatatny S., "Detecting downhole vibrations through drilling horizontal sections: machine learning study," *Scientific Reports*, 13(6204), 2023.

[2] Saadeldin R., Gamal H., Elkatatny S., "Machine Learning Solution for Predicting Vibrations while Drilling the Curve Section," *ACS Omega*, 8(39), 35822-35836, 2023.

[3] Chen J., Yang L., Huang Q., Zhang W., Qiao J., Zhang X., Jiang D., "Dynamic analysis of drill string vibration enhanced by machine learning models," *Preprint on SSRN*, 2023.

[4] Etaje D. C., "Identifying the Optimum Zone for Reducing Drill String Vibrations," Master's thesis, University of Calgary, Calgary, Canada, 2018.

[5] Hegde C., Millwater H., Gray K., "Classification of drilling stick slip severity using machine learning," *Journal of Petroleum Science and Engineering*, 179, 1023-1036, 2019.

[6] Islamov A., et al., "Research Risk Factors in Monitoring Well Drilling—A Case Study Using Machine Learning Methods," *Journal of Petroleum Science and Engineering*, 192, 107338, 2021.

[7] Mehrad M., Bajolvand M., Ramezanzadeh A., Ghavidel Neycharan J., "Developing a new rigorous drilling rate prediction model using a machine learning technique," *Journal of Petroleum Science and Engineering*, 192, 107338, 2020.

[8] Kudaibergenov A. K., Kudaibergenov A. K., Khajiyeva L. A., "Modelling of resonance regimes of drill string nonlinear dynamics," *International Journal of Mechanics*, 11, 92-100, 2017.

[9] Hastie T., Tibshirani R., Friedman J., *The Elements of Statistical Learning*, 2nd ed., Springer, 2009.

[10] Muller A., Guido S., *Introduction to Machine Learning with Python: A Guide for Data Scientists*, 1st ed., O'Reilly Media, 2016.

[11] Khajiyeva L., Umbetkulova A., "Resonance Vibrations and Stability of Drill Strings Complicated by Frictional Forces," *International Journal of Mathematics and Computers in Simulation*, 11, 147-158, 2017.

Авторлар туралы мәлімет:

Кудайбергенов Аскар Канатович - PhD, Математикалық және компьютерлік модельдеу кафедрасының аға оқытушысы, әл-Фараби атындағы Қазақ ұлттық университеті (Алматы, Қазақстан, пошта: askarkud@gmail.com);

Кудайбергенов Аскат Канатович - PhD, Математикалық және компьютерлік модельдеу кафедрасының аға оқытушысы, әл-Фараби атындағы Қазақ ұлттық университеті (Алматы, Қазақстан, пошта: askhatkud92@gmail.com);

Тілеуберді Тогжан Бақытқызы (корреспонденция авторы) - Математикалық және компьютерлік модельдеу кафедрасының магистранты, әл-Фараби атындағы Қазақ ұлттық университеті (Алматы, Қазақстан, пошта: togzhantileuberdi@gmail.com).

Информация об авторах:

Кудайбергенов Аскар Канатович - PhD, старший преподаватель кафедры Математического и компьютерного моделирования, Казахский национальный университет имени аль-Фараби (Алматы, Казахстан, почта: askarkud@gmail.com);

Кудайбергенов Аскат Канатович - PhD, старший преподаватель кафедры Математического и компьютерного моделирования, Казахский национальный университет имени аль-Фараби (Алматы, Казахстан, почта: askhatkud92@gmail.com);

Тілеуберді Тоғжан Бақытқызы (автор для корреспонденции) - магистрант кафедры Математического и компьютерного моделирования, Казахский национальный университет имени аль-Фараби (Алматы, Казахстан, почта: togzhantileuberdi@gmail.com).

Information about authors:

Kudaibergenov Askar Kanatovich - PhD, Senior Lecturer at the Department of Mathematical and Computer Modeling, Al-Farabi Kazakh National University (Almaty, Kazakhstan, e-mail: askarkud@gmail.com);

Kudaibergenov Askat Kanatovich - PhD, Senior Lecturer at the Department of Mathematical and Computer Modeling, Al-Farabi Kazakh National University (Almaty, Kazakhstan, e-mail: askhatkud92@gmail.com);

Tileuberdi Togzhan Bakytkyzy (corresponding author) - Master student at the Department of Mathematical and Computer Modeling, Al-Farabi Kazakh National University (Almaty, Kazakhstan, e-mail: togzhantileuberdi@gmail.com).

Received: February 26, 2025
Accepted: April 23, 2025